Humanities and Big Data in Ibero-America

Humanidades Digitales y Big Data en Iberoamérica

Digital Humanities and Big Data in Ibero-America

Editado por / Edited by
Ana Gallego Cuiñas y / and Azucena González Blanco

Volumen / Volume 2

Humanities and Big Data in Ibero-America

Theory, methodology and practical applications

Edited by
Ana Gallego Cuiñas and Daniel Torres-Salinas

DE GRUYTER

Esta publicación es resultado de la Unidad Científica de Excelencia “Iber-Lab. Crítica, Lenguas y Culturas en Iberoamérica” (Ref. UCE2018-04) de la Universidad de Granada

ISBN 978-3-11-075351-6
e-ISBN (PDF) 978-3-11-075352-3
e-ISBN (EPUB) 978-3-11-075361-5
DOI https://doi.org/10.1515/9783110753523

Library of Congress Control Number: 2023941375

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the internet at http://dnb.dnb.de.

Cover image: as creative atelier/DigitalVision Vectors/Getty Images
Typesetting: Integra Software Services Pvt. Ltd.
Printing and binding: CPI books GmbH, Leck

www.degruyter.com

Contents

Ana Gallego Cuiñas & Daniel Torres Salinas

Introduction. Towards Expanded Humanities: Review and Agenda

The *expanded humanities,*[1] making use of Rosalind Krauss's celebrated concept, allude not only to other forms of humanist knowledge, marked by the combined use of sociological and technological methods and tools, but also to beyond the human into the digital space. In the third decade of the twenty-first century, this category, along with those of "Digital Humanities",[2] "Big Humanities" (Lane 2016), and "Augmented Humanities" (Mendoza 2016), is beginning to shake off suspicion[3] and is becoming the object of growing interest in the academic world. On the one hand, within the humanist and cultural field, the development of new research techniques has been strengthened in areas such as literary studies, linguistics, philosophy, criticism, history, and the cultural industry over the last decade, through the use of large databases, text corpora and algorithms that open up renewed pathways to knowledge of our past, present and future. On the other hand, in the social, political and economic sphere, the humanist and ethical tradition is called upon to tackle the datafication of the world and the problems brought about by: the growing lack of privacy (Véliz 2021);[4] the new dialectical relationship between the virtual and the real (Fisher 2016); the control and commercialization of the data we generate by companies and platforms (Srnicek 2018); the internet of things[5] (Han 2021); the over-representation of the world and its

1 The notion of 'expanded humanities' is a theoretical proposal by Ana Gallego Cuiñas.

2 There are multiple definitions of the Digital Humanities, although the large majority refer to digital collections and archives, databases, online biographies, et cetera.

3 We cannot deny that until relatively recently, the relationship between the Humanities and Big Data seemed almost oxymoronic.

4 Rivers of ink have flowed in the last decade over the issue of privacy and the (bad) use of our data: from the consideration that the subject who uses the internet is the product, to think about new forms of social control. Extremely interesting exhibitions have also been held, such as *Big Bang Data* at the Centro de Cultura Contemporánea de Barcelona in 2014, whose catalogue *Anonimízate. Manual de Autodefensa Electrónica* was a huge success. See: https://www.cccb.org/rcs_gene/Anonimitza_t_def_CAT_ENG.pdf. Here there is an overview of the main electronic self-defence resources that have been created: Surveillance Camera Players; iSee from Institute of Applied Autonomy, iSee; Life: A User's Manual (2003–2006) by Michelle Teran; CV Dazzle and Off Pocket, by Adam Harvey; Invisible by Biogenfutur, or Blackphone by Silent Circle, among others.

5 Our "intelligent" devices (in the domestic and labour spheres, in the street, etc.) are hyperconnected and extract our personal data (habits of consumption, sociability, movement, fiscal or banking data, medical files, etc.) that can be sold for financial and political gain.

https://doi.org/10.1515/9783110753523-001

excessive abstraction (Berardi 2019); information overload (Tello 2018);[6] and *Data Mining*, *Machine Learning* and the use of intelligent tools such as ChatGPT and algorithmic governmentality (Sadin 2017).

With this starting point, the members of the Excellence Cluster "Iber-Lab. Crítica, Lenguas y Culturas en Iberoamérica" ("Iber-Lab: Criticism, Languages and Cultures in Ibero-America") of the University of Granada, bring together in this book papers from specialists in Literature, Spanish Language, Linguistics, Philosophy, Theory, Cultural Studies, Economics, and Data Science, in order to discuss the epistemic nature of Big Data, its theoretical, diachronic and synchronic problems, as well as the variety of its methods and applications in the Humanities. Generally speaking, there are three objectives – and sections – that make up the backbone of this volume:

1) *Theoretical*, in which we explore, debate and outline a critical framework of humanist thought for computational techniques and big data.
2) *Methodological*, which shows different computational methodologies and tools for analysing big data in the Humanities.
3) *Practical*, which presents some practical applications and their field of validity in various humanist disciplines.

The pairing of art and technology has helped assemble our societies since prehistory, and crystallizes the symbolic and material root that constitutes our culture. Plato was the first to think about a humanist criticism of technology, in *Phaedrus*, thus initiating the opposition between culture and technology. Marx also highlighted the threat to culture that came from the relation of capital with machines, but today the interaction between humans and machines is already a fact, and such are the advances in Artificial Intelligence (AI) that they are now directly talking about technological genesis (Hayles 2012). In reality, algorithms have not only learned and do learn from the past of humans, but they also have autonomy and interact with us and between each other, capable even of arousing affection. Indeed, in 2014, MIT created Story Telling Robot, which told stories to children, who ended up developing an emotional relationship with the machine. Furthermore, the last decade has witnessed a surge in post- and trans-humanist positions.

We have therefore passed from the Anthropocene to the Capitalocene, and as a consequence of this system, the Technocene is emerging, or more accurately, the *Technocapitalocene*, based on the capitalism of things and data. Big Data have

6 "From 2014 down to today, 2017, we have created as much information as the period from prehistory to 2014. And the most impressive, for me, is that digital information is going to surpass in quantity all the biological information that exists on the planet." (in Costa 2022: 10).

opened the door to the digitalization of (almost all) our world,[7] which means operating with the processes of fragmentation, multiplication, abstraction and globalization of information, which give rise to infinite possibilities of economic, political, social, cultural, artistic and also academic use. This new paradigm, which is not necessarily either good or bad, is an "experimental laboratory", as Flavia Costa (2021: 10) calls it, to trial the new epistemologies and methodologies that the humanities of the future will have to define, and these cannot be less than a kind of "expanded humanities". We must therefore confront the challenge of "stopping to think" (Ibáñez 2014: 131) about the design of the agenda that defines the Humanities' relationship with technology, particularly with big data and AI. Moreover, we cannot evade the fact that certain themes and problems, such as abstraction, materiality, reproducibility, and the dangers of the introduction and perpetuation of gender and colonial biases in theory and technological praxis, make the intervention of a humanist, inclusive and decolonial gaze absolutely vital in academic and cultural studies that use computational tools.

We advocate, therefore, for the trans-epistemic coming-together of the humanities, culture and technology, to transcend the methodological emphasis of the digital humanities, which is fundamental but not sufficient, to promote theoretical, philosophical and "situated" (Haraway 1995) research areas in "los Sures" (literally, "the Souths", meaning the Global South) (Boaventura & Meneses 2012), and specifically, in the Ibero-American world. In other words, we argue for a humanities that, as well as dataistic techniques due to their useful or critical *expansion*, include a reflection on the place of enunciation of big data, which includes the mandatory incorporation of gender and decolonial studies. The way in which we use – and search for – data is conditioned by our ideology and its biases.[8] The reading of the past, like the reading of the future, depends on the (situated) gaze of the present, because the data do not speak for themselves, but need a watchful and humanist interpretation.

Nevertheless, we cannot not lose sight of the difficulty that this line of research entails in our field, because all the humanistic disciplines self-legitimize (cf. Gallego Cuiñas 2022) in their fight for:

7 For example, on the platform https://www.internetlivestats.com/, you can see the information that is uploaded daily to the internet, since its creation, throughout the world. The growth is truly dizzying and hair-raising.

8 The datacritical platform serves as an example, "an organization that strengthens critical narratives through the use of data", which works on issues of "gender, climate crisis and structural inequalities in Latin America". See: https://datacritica.org/.

- the *objects of study*, given that either they study texts or the authors, spaces, times, societies, theories, ideologies, practices, et cetera. Different variables do not tend to be combined, nor do they tend to do criticism of criticism.
- the *methods*, which correspond to qualitative paradigms (close, theoretical, exegetic or hermeneutic reading) or quantitative paradigms (the corpora, the statistics, the sociological or digital tools).
- The *frameworks of readability*, which are crystallized in the prevalence of particular approaches: positivist, aesthetic, social, cultural, economic, political, diachronic, synchronic, depending on the perspective taken.

Additionally, we need to recognize that the limits and tools of the *expanded humanities* are not fully clear, although that is one of its greatest virtues, a sign of its potential and futurability. The shift toward other epistemes and methodologies entails a notable critical and technological effort, which moreover stirs up the debate on what is authentic or valid for each discipline – that is, on the legitimacy of methods. This is in spite of the fact that it is clear that the well-oiled, correct and pure methods do not truly advance knowledge; rather this is done by those that face up to ontological and epistemic challenges, such as those proposed here. The studies by Daniel Torres Salinas, Sara Mariottini, Wenceslao Arroyo-Machado, Ana Gallego Cuiñas, Azucena González Blanco and José Antonio Pérez Tapias tackle the theoretical challenge that the inclusion of big data in the Humanities entails, above all for the inclusion of new object of study and frameworks of readability in literary studies and philosophy. In the following section, devoted to methodologies that combine the humanistic with data science, Wenceslao Arroyo, Nicolás Robinson, Francisco Benítez, Esteban Romero, Miguel Calderón and Carolina Gainza address how scientometrics, blockchain, linguistic corpora, and algorithms expand the possibilities of the humanities, language and literature in Spanish. And in the third section, focused on practical applications that can serve as an example to other researchers, Carolina Ferrer opts for the possibilities of criticometrics, Diana Roig-Sanz, Alessio Cardillo and Ventsislav Ikoff examine network science, Pedro Ruiz the combination of qualitative and quantitative analyses in poetry, and Ana Gallego Cuiñas and Daniel Torres Salinas look at the study of writer figures and the reception of literature on social networks.

All of them refer to object of study based in the Ibero-American area to drive the idea that there is a need not only for humanist but also decolonial and inclusive Big Data. Mass data tend to make invisible both the ideology and the situated

materiality[9] of the information and of the media, which by themselves do not light up the world, as Byung-Chul Han (2021: 18) would say. It is as important to reveal the correlations between data and the establishment of patterns as it is to deconstruct their ideological and geopolitical bias (Habermas 1986), the task of the humanist researcher, trained in *close reading*, alert against false neutrality. And at the same time, this researcher must be open both to the challenges imposed by the digital society of their time and to the use of the new, technological "toolbox", as Benjamin understood it, that big data and AI make available to them.

This book is planned to pave the way into this field of opportunities for the humanities (language, literature, philosophy, cultural studies) and the social sciences (data science and economics) and is aimed to function as a kind of introductory manual, theoretical and practical, to the Humanities and Big Data in Ibero-America, or better still, to the *expanded humanities of the Global South*. This should be of use for researchers interested in an emerging subject area that has and will have an indisputable epistemic impact on humanistic studies for the rest of the twenty-first century.

Bibliography

Berardi, Franco (2019): *Futurabilidad. La era de la impotencia y el horizonte de posibilidad*. Buenos Aires: Caja Negra.

Costa, Flavia (2021). *Tecnoceno: Algoritmos, biohackers y nuevas formas de vida*. Buenos Aires: Taurus.

Fisher, Mark (2016): *Realismo capitalista. ¿No hay alterantiva?* Buenos Aires: Caja Negra.

Gallego Cuiñas, Ana (2022): *Cultura literaria y políticas de mercado. Editoriales, ferias y festivales*. Berlin: De Gruyter.

Habermas, Jürgen (1986): *Ciencia y técnica como ideología*. Madrid: Tecnos.

Han, Byung-Chul (2021): *No-cosas. Quiebras del mundo de hoy*. Barcelona: Taurus.

Haraway, Donna (1995): *Ciencia, cyborgs y mujeres. La reinvención de la naturaleza*. Madrid: Cátedra.

Ibáñez, Jesús (2014): *El delirio del capitalismo*. Madrid: Los libros de la catarata.

Lane, Richard J (2016): *The Big Humanities: Digital Humanities/Digital Laboratories*. New York: Routledge.

Mendoza, Juan José (2016): "Humanidades aumentadas: Por una historia de las practicas cruzadas", in *Luthor*, 28, pp. 1–13.

9 The expanded humanities, as well as being decolonial and inclusive, must be sustainable – committed to what has been called "digital sobriety" – taking into account the carbon footprint of websites and the high energy consumption of high resolution, which have a pronounced environmental impact.

Sadin, Éric (2017): *La humanidad aumentada. La administración digital del mundo*. Buenos Aires: Caja Negra.

Sousa Santos, Boaventura de y Mª Paula Meneses (2014): *Epistemologías del Sur (Perspectivas)*. Madrid: Akal.

Srnicek, Nick (2018): *Capitalismo de plataformas*. Buenos Aires: 2018.

Tello, Andrés Maximialiano (2018): *Anarchivismo. Tecnologías políticas del archivo*. Buenos Aires: La Cebra.

Véliz, Carissa (2021): *Privacidad es poder. Datos, vigilancia y libertad en la era digital*. Barcelona: Debate.

1 Theoretical Framework

Sara Mariottini, Wenceslao Arroyo-Machado
and Daniel Torres-Salinas

A Brief Introduction to Big Data for Humanists

1 Brief Introduction

We usually associate big data with its cruder, more conventional and perhaps more obscure applications; those of the interconnected, data-driven world, where every interaction and every 'like' leaves a trace, every click is recorded and privacy is supervised by third parties as millions of data records are gathered from millions of users 24 hours a day. Here are some of the figures regarding this phenomenon: according to Eric Schmidt, every day we generate as much data as all the data produced by the whole of humanity in 2003 (Siegler 2010). These data are generated by the 4.66 billion active internet users who, to give an example, can publish 3.3 million posts on Facebook or perform 3.3 million searches on Google every minute (Alonso 2020). According to the predictions for 2025, there will be 163 zettabytes[1] of data in the world (Zgurovsky & Zaychenko 2020). In this context, one of the most common uses of big data is in digital marketing, although we can find it everywhere, whether politics (Pascual & Peinado 2018; Rands, 2018), finance, with its algorithms for surveillance and decision-making (Hasan, Popp & Oláh, 2020), health monitoring (Sun et al. 2020), the recommendation systems of entertainment platforms (Fayyaz et al, 2020) or sports (Torgler 2020).

There is also talk of a new research paradigm in the academic realm. Big data is changing the way we generate and analyze scientific results due to the massive generation of data, heavy reliance on technology and the widespread use of mathematical models, algorithms and artificial intelligence (AI). The era of big data is here to stay and will accelerate learning in all scientific fields. Universities and research institutes already promote interdisciplinary collaboration and stimulate "cross-fertilization" between different fields which have data science as a common axis (Galeano & Peña 2019). The increased capacity of acquisition, processing and analysis of data with the potential to reveal patterns has contributed to the connection of different scientific disciplines. Some of the most outstanding examples include the Large Hadron Collider (Ortíz 2019), radio telescopes such as

1 Various current estimates indicate that the volume of data in 2021 stands at 44 zettabytes (van der Aalst 2016; Kugler 2018). A zettabyte is equivalent to one billion terabytes.

https://doi.org/10.1515/9783110753523-002

the Square Kilometer Array (Scaife 2020) and the NASA Center for Climate Simulation (NCCS) (Schnase et al. 2011) and the application of big data in education to analyze students (Fischer et al. 2020).

However, this sudden intrusion in many areas has caused some bewilderment. The term 'big data' is still somewhat confusing for researchers, as most associate it with its most basic objectives such as data collection and processing of operations and do not have a clear overview of its scope and implications (Favaretto et al. 2020). Moreover, there is a certain sense of uneasiness towards big data as it is a cultural phenomenon in a state of constant change and evolution and the use of this concept as a buzzword further aggravates its conceptual vagueness. Therefore, the aim of this chapter is to offer a synthetic vision of what is understood as big data to serve as a starting point for researchers in the field of humanities.

2 Characteristics and Definition of Big Data

The raw material of big data is obviously the data, which is understood as a symbolic representation of an attribute which may be qualitative or quantitative. In the case of big data they have been translated into a digital format allowing their use and processing and are catalogued to facilitate their processing and analysis using multiple techniques. The magnitudes of big data require the use of significant computational resources. Another fundamental aspect of big data is that it may be collected effortlessly through all kinds of devices such as smartphones, social media, sensors, etc. These gadgets determine the essential aspects of big data, namely its exaggerated volume, the speed of its collection and its variety (Laney 2001; Ward & Barker 2013). The EU[2] defines big data as:

> large amounts of different types of data produced from various types of sources, such as people, machines or sensors. This data includes climate information, satellite imagery, digital pictures and videos, transition records or GPS signals. Big Data may involve personal data: that is, any information relating to an individual, and can be anything from a name, a photo, an email address, bank details, posts on social networking websites, medical information, or a computer IP address.

However, although the characteristics of data are clear to some authors, there is no univocal definition of big data. As a result, new characteristics are added such

2 European Commission, Directorate-General for Justice and Consumers, The EU Data Protection Reform and Big Data, Publications Office, 2018, https://data.europa.eu/doi/10.2838/190200.

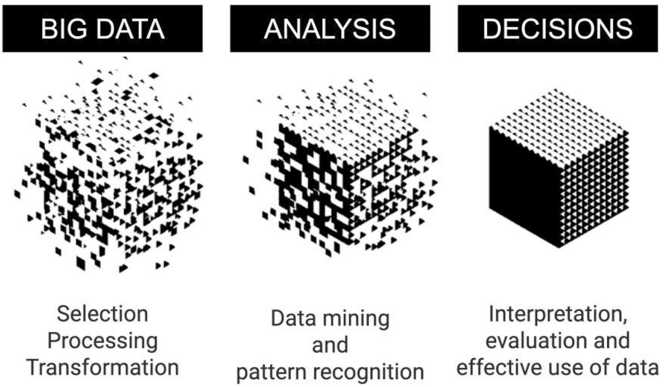


Figure 1: Metaphor of the essence of big data and its main processes.

as its capacity not only to be captured but also to be stored for permanent updating and continuous exploitation. The latter analysis processes are related to data visualization and prediction and involve the use of methods that extract value and meaning from the data (Figure 1). These analysis techniques are oriented towards three main objectives: the search for patterns, the identification of associations and the development of models that allow us to make forecasts. As can be seen, big data is a complex discipline. A simple definition that captures the above concepts is provided by the Gartner IT Glossary,[3] which defines big data as:

> high-volume, high-velocity and/or high-variety information assets that demand cost effective, innovative forms of information processing that enable enhanced insight, decision-making, and process automation.

This definition offers a framework consisting of three facets: volume, velocity and variety (physical characteristics of the data), to which we can also add veracity and value, i.e., the data must be of good quality, relevant and reliable and must allow us to achieve our objectives, and the data must provide added value to help us decide or understand a phenomenon holistically. These five characteristics make up what in big data literature has come to be known as the 5 Vs (Favaretto et al, 2020). Some authors go even further and talk about the 7 Vs, adding volatility and validity to the above (Khan, Uddin & Gupta 2014). These latter two attributes refer to the need to consider the feasibility of a big data project and the form of data presentation. While the view of the 7 Vs is somewhat Manichean and synthetic, it effectively introduces the attributes, processes and actions needed in any

3 https://www.gartner.com/en/information-technology/glossary.

big data project. Nonetheless, certain sectors of the social sciences consider this definition to be too technological, with a certain utopian character (Kitchin & McArdle 2016; Gandomi & Haider 2015).

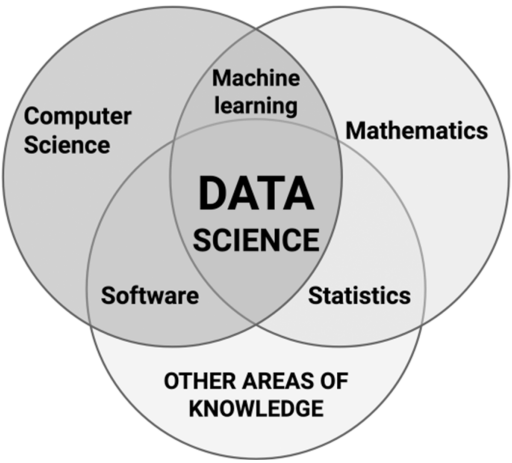


Figure 2: Disciplinary relationships of data or data science.

Precisely because not all big data share the same characteristics, it makes sense to use a purely 'technological' definition. However, from a humanistic viewpoint this definition can be improved by emphasizing the human side of data and promoting re-humanization of the digitized social product that we have become for big data. A very significant percentage of big data is devoted to studying the hyper-connected population of the so-called turbo-capitalism (Luttwak 2000), excluding from its discourse all individuals alien to big data flows. At the same time, algorithms are social products and can also reflect the prejudices, social stigmas and ineptitudes of the developer (Mac 2021; Jiménez de Luis 2021). Therefore, a merely technological definition of big data provides us with a framework, but at the same time it obviates an ethical and humanized approach, overlooking the fact that data are generated by people and algorithms are sometimes simply a mere aggregate of emotions.

As we can see, when we talk about big data we are faced with a complex phenomenon that has given rise to a new multidisciplinary field called data science (Figure 2). Data science has its origins in computer science and maintains a close relationship with AI and the internet of things, that increasingly palpable world where every activity, every click and every step is recorded and stored and even the most unassuming and irrelevant gadget (a light bulb, a refrigerator, etc.) can generate data and be connected to the internet. Data science is therefore an intimate combination of technology and mathematics aimed at understanding human

behavior and making it increasingly predictable. From our standpoint, with the advent of big data we are faced with an epistemological and ontological problem that opens up a world of opportunities for social sciences and humanities in terms of definition, methodology, deconstruction and new integrations. The following basic introduction outlines some of the basic elements of working with big data.

3 Methodological Elements of Big Data

3.1 Main Types of Data and Formats

Until the first definition of big data appeared in the 1990s, all data was, in effect, small data and therefore it did not need to be labeled as such (Faraway & Augustin, 2018). Due to the difficulties of generating, processing, analyzing and storing data, it was produced in a very controlled manner using samples that limited its life cycle and size. Today, big data is generated continuously and is intended to be flexible in scope and scalable in its production. Although big data may claim to be exhaustive, it is nevertheless a representation and a sample of the social reality limited to a specific moment in time (Mayer-Schönberger & Cukier 2013). For this very reason, the data captured are conditioned by the following aspects (Li 2015):

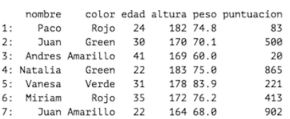
- The data collection framework (data collection devices and/or sensors, the parameters used, etc.)
- The technology/platform used (which can produce variations and biases in the data generated)
- The context in which the data are generated (data are always considered in relation to the circumstances)
- The data ontology used (how they are calibrated and classified)
- The regulatory environment governing privacy, data protection and security

Once we know what conditions the data, we can move on to consider the different types of data. Big data can also be classified into three classes according to the structure (Table 1):

- Structured data: data stored in tables with a well-defined length and format which can be easily sorted and processed by any data management tool. Examples of structured data include dates, data sheets and databases.
- Semi-structured data: information that is not regular and therefore cannot be managed in a standard way. Examples of semi-structured data include HTML, JSON and XML.

– Unstructured data: binary data that has no identifiable internal structure. This is a massive, disorganized conglomerate of data that has no value until it is organized and stored. Examples of unstructured data include images, videos, audio files and PDFs.

Table 1: Classification of big data.

Unstructured data	Semi-structured data	Structured data
		

Another classification may be made based on the format of the data. Below are examples of the main data formats and their description (Table 2). The following section offers an explanation of a selection of the main formats that allow data analysis.

Table 2: Typical data and file formats.

Format	Description
.xlsx/xls Microsoft Excel spreadsheet	Proprietary file format for the storage of structured data in tables. Microsoft Excel allows data display and analysis, although it is of limited use with large volumes of data due to its inefficiency.
.txt Plain text	Plain text files are the universal free format for storing information. Their content may be structured in different formats.
.csv/tsv Comma/Tab separated values	Text file format made up of structured data in tables with comma-separated (csv) or tab-separated (tsv) fields. This is the most basic and efficient format for storing structured data.
.xml Extensible Markup Language	Text file format for semi-structured data storage and data exchange between applications.
.json JavaScript Object Notation	Standard text file format for semi-structured data storage and data exchange between applications, which is lighter and more legible than XML format.

3.2 Main Types of Data and Formats

3.2.1 Basic Big Data Techniques: Basic Classification of Methods

This section will deal with machine learning techniques, a branch of AI for big data processing which essentially aims to identify patterns in the data in order to make inferences. Machine learning algorithms may be classified into three main paradigms:

- Supervised learning (SL): the algorithm learns from several examples how given inputs generate specific outputs (they are labeled) and is thus able to make inferences for new cases. Classic examples include linear regressions and decision trees (DT).
- Unsupervised learning (UL): unlike supervised learning, the algorithm does not have labeled outputs and instead of learning which combination of attributes generates them it searches for patterns in the input data. A classic example is clustering algorithms such as k-Means.
- Reinforcement learning (RL): the algorithm learns from the experience developed in a dynamic environment where it receives rewards. It also does not need to know the labeled output. One example is deep neural networks (DNN).

3.2.2 Examples of Popular Techniques

Some of the most popular machine learning techniques are outlined below. These should be understood as mere examples as there is a host of different techniques that can be used. Firstly, a decision tree (DT) is a hierarchical supervised learning model. It can be seen as a flowchart starting from a root and branching out along different nodes until it reaches a leaf. Each node tests the data, and the branches represent the concrete result of the test. Ultimately rules are generated indicating each of the paths from the root to the leaf. Decision tree models are one of the most common machine learning models because of their recursive 'divide and conquer' nature and the fact they are descriptive and easy to understand (Flach 2012).

Other techniques are concerned with deep learning, a subfield of machine learning that bases its high-level learning process on artificial neural networks. Generally speaking, a simple neural network is composed of an input layer, a hidden layer and an output layer. Inspired by the architectural depth of the brain, neural network researchers have for decades sought to develop and train deep multi-layer neural networks so that the model can learn increasingly complex levels of abstraction (Bengio 2009). The goal of each layer is to extract relevant features from the incoming data and after training all the layers one by one, they

are all put together and the whole network is refined (Alpaydin 2014). A good example is the generative adversarial networks (GAN) that are known to be commonly used for the generation of 'deepfake' images (Figure 3). Its applications are endless in fields such as advertising and arts and crafts. For example, it can help to create new shoe designs or generate a painting inspired by a great artist from the past using a photo.

Figure 3: Deepfake of a building.

Thirdly, we should mention clustering techniques. These are machine learning approaches that attempt to find similar patterns and relationships between data points in order to group them. Each cluster is composed of data points which due to their attributes are similar to each other rather than those of another cluster (Sarkar et al. 2018). For example, this technique is commonly used in social network analysis to detect communities of users based on their social relationships and/or interests (Arroyo-Machado, Torres-Salinas & Robinson-Garcia 2021).

3.3 Big Data Tools

We will cap off this methodological section by describing some useful tools for data analysis and processing at different levels (Table 3). The most powerful and most directly used tools in the analysis of big datasets are data processing frameworks, of which Apache Hadoop and Apache Spark are practically the standard. However, other tools are very popular due to their versatility and power, allowing their application for anything from small data through to large volumes of data, such as the programming languages Python and R. Data mining software also exists that allows the development of models in a visual environment without requiring use of a programming language, such as KNIME and Weka. Finally,

also worth mentioning are cloud computing applications, which allow the possibility of contracting a complete customized and scalable work environment that can be accessed via the internet, thus dispensing with the purchase, installation and configuration of equipment and so reducing both costs and time. One of the most popular options is Google Cloud.

Table 3: Main tools for big data analysis.

BIG DATA FRAMEWORKS	
Apache Hadoop	Framework for fast data processing
Apache Spark	Framework for data storage and querying
Apache Hive	Framework for data storage and querying
VERSATILE PROGRAMMING LANGUAGES	
Python	Programming language widely used for data science
R	Programming language widely used for statistical analysis
Scala	Programming language useful for big data processing
INTERACTIVE ENVIRONMENT TOOLS	
KNIME	Tool focused on data mining processes
Weka	Data mining tool that includes a collection of machine learning algorithms
RapidMiner	Tool that includes data mining and machine learning processes
CLOUD COMPUTING	
Google Cloud	Cloud computing services by Google
AWS	Cloud computing services by Amazon
Azure	Cloud computing services by Microsoft

4 Big Data Applied to Humanities

This section outlines some of the recent interactions between data science and humanities and social sciences. Given its multidisciplinary nature, we will see how this epistemological hybridization is taking place in several specialty areas. The art world was one of the first to pay attention to this phenomenon, giving rise to what has come to be known as art data. Within this field, one of the most frequently cited projects is the Wind Map (Viégas & Wattenberg 2012), which has several unique features; first of all, the work is exhibited in the MoMa and was created by a computer scientist and a scientific journalist (Figure 4). The project consists of a living map of the winds that sweep across the United States based on data from the National Digital Forecast Database, represented with trails reminiscent of the brushstrokes of Renaissance painters which endow the meteorological data with beauty.

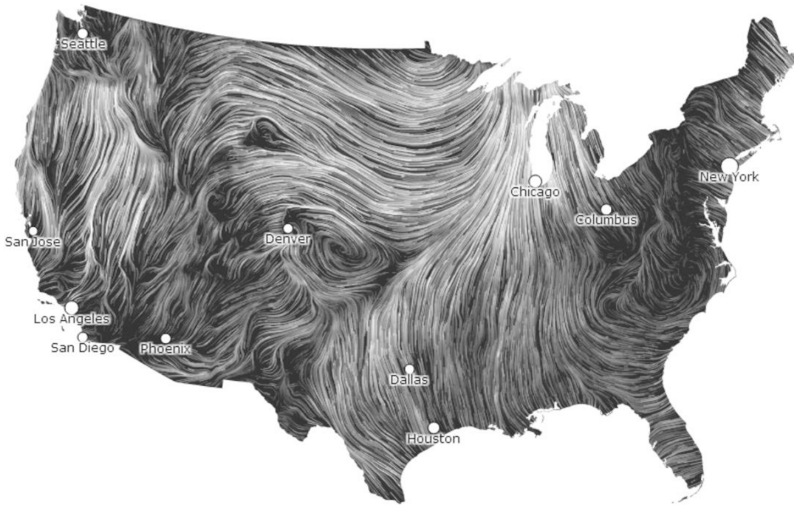


Figure 4: Image from the Wind Map project that combines art with U.S. meteorological data.

Another area where data science has proven effective is heritage and archaeology, where simple information systems are being replaced by systems that integrate multiple sources (sensors, digital libraries, social networks, etc.) (Amato et al. 2017). Projects in this area are often complex, but we will begin with a small example to illustrate its possibilities. In Bogota (De Urbina 2021), digital photographs of urban scenes from Panoramio were characterized based on the collective perception of the population using semi-structured data (photographer, date, coordinates, event or tags). In a European context, the European project ATHENA (Nisantzi et al. 2018) integrates remote sensing technologies applied to cultural heritage and centralizes the data in a single point.[4] ATHENA collects data using active and passive remote sensing systems which are mainly used in archaeological contexts. Meanwhile, the SCRABS project is a combination of the two previous proposals. Described by the authors as a "Smart Context-awaRe Browsing assistant for cultural EnvironmentS", it is a paradigmatic example of the collaboration between computer scientists, archaeologists, architects and cultural managers (Amato et al. 2017).

The researchers Zgurovsky and Zaychenko (2020) sought to identify the regularity of systemic global conflicts based on analysis of historical big data. So far,

4 https://cordis.europa.eu/project/id/691936/es.

an analysis of the complete list of global conflicts occurring since 2500 BC shows that up until the 7th century BC these conflicts did not follow any regular pattern. However, a periodic pattern was revealed in the series of global conflicts following the emergence of higher forms of organization, with the authors relying on analysis of historical data relating to global conflicts that have taken place from 705 BC through to the present day. Using a range of primary sources, they attempted to foresee the next global conflict which they called "the conflict of the 21st century".

Some of Google's projects could also be seen as examples of big data applied to humanities. For example, in 2004 it began the ambitious mass digitization of more than 100 million books through Google Books, generating one of the largest masses of unstructured data. Some of its applications can be found in Google Books Ngram Viewer, an online search engine that charts the frequencies of any set of search strings using a yearly count of n-grams found in printed sources published between 1500 and 2019. Figure 5 shows the frequency of searches for two eighteenth-century poets in the English corpus of Google Books, revealing the interest in their work at different chronological points in time. These techniques fall under what has come to be called Text Corpus Visualizations (Hai-Jew 2015).

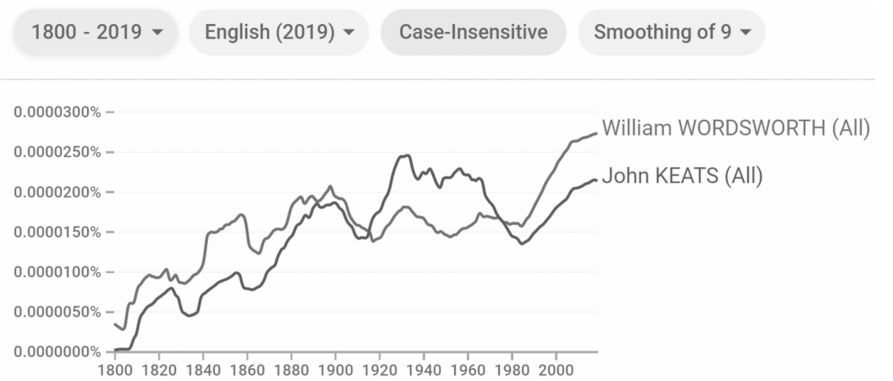


Figure 5: Text Corpus Visualizations using Ngram.

5 The Magister Ludi of Data

To conclude, we will briefly discuss the risks of the misuse of big data. These risks have originated in the current information society due to its dependence on ICT, which has given rise to a context of vulnerability driven by the most accelerated form of capitalism, also known as turbo-capitalism (Luttwak 2000). The dangers

of this new environment are evident in the case of entertainment applications and services offered as a free service but which turn the user into the product by accessing, processing and making an economic profit from the data they generate. This is a risk that often goes unnoticed, with the need to access and consume information inevitably overcoming the privacy and rights of the consumer. All the interactions produced on the internet end up feeding algorithms, which use them to filter and catch our attention with whatever the companies that program them want. However, these tools overlook many relevant issues by converting human beings into numbers (Dodson 2008), a risky simplification that could potentially have catastrophic consequences. An example of this is the Black-Scholes equation and other similar models, which some authors point to as being complicit in the culture of excessive risk and unbridled speculation that eventually led to the 2008 financial crisis (Stewart 2012; Harford 2012; O'Donnell 2015).

Other scandals also stand out in this context, such as the company Cambridge Analytica which made improper use of Facebook data in 2016 to influence voters during the Brexit referendum (Hern 2019) and the elections of Donald Trump (Rosenberg, Confessore & Cadwalladr 2018). In both cases, personal data were unlawfully collected by creating political profiles of users in order to send personalized information (García Fernández 2018). The consequences were incalculable and it triggered a legal storm that caused Facebook to lose billions in stock market value, as well as suffering social rejection (Hindman 2018). Apart from the influence these algorithmic models have on our daily activity, there is also the added risk of learning biased or prejudiced behaviors. Social media are precisely one of the main vehicles for tracking and monitoring activity, but it is precisely in these same spaces where we are witnessing an increase in hate speech (Müller & Schwarz 2020) and sexist discourses (Rodriguez-Sanchez, Carrillo-de-Albornoz & Plaza 2020).

Finding a way to avoid falling into bias traps or negative behaviors learned from humans is but one aspect of an even greater challenge: codification of the innumerable differences and nuances of humanity in areas such as culture, politics, religion, sexuality and morality (Webb 2021). This is a major problem because AI as it is currently conceived cannot be attributed intelligence because it is closed to the world in which it has been programmed and cannot see beyond it, being insensitive to and ignoring the dynamics of a world in constant change (Masís 2009). Ultimately, AI learns patterns from the past and provides us with an approximation of the reality of that moment in a specific context.

Therefore, it does not seem entirely clear that the solution to this problem lies in the indiscriminate increase of data. Indeed, the level of knowledge is often confused with the volume of data, when in many cases it is the smaller and better curated collections that allow us to find useful solutions in an efficient way (Olson, Wyner & Berk 2018). Smart data is thus proposed as the transformation of

big data into quality data after its cleansing (Triguero et al. 2019). In relation to all this, in the same way that human beings can see their critical capacity being limited in the face of information overload (Marta-Lazo 2018), big data algorithms can also end up leading to other kinds of problems when data are processed without paying any prior attention to them. That is why it is risky to directly point to the data with the highest number of instances and/or properties as being more relevant. In fact, there are already visible signs of this limited view in the academic realm, where the existence of a gap between the so-called 'data-rich' and 'data-poor' research fields has been identified (Sawyer 2008).

The last novel by Herman Hesse tells the story of Joseph Knecht, the Magister Ludi of Castalia or highest authority of the Glass Bead Game, a kind of high-level humanistic entertainment which is essentially an abstract synthesis of all arts and sciences (Hesse 2012). Players aim to establish relationships between all knowledge based on a given topic. What Hesse seemed to anticipate here is a metaphor for the fate of knowledge, encoded in data and highly connected, although in Castalia the game is controlled not by technocrats but by humanists. As Byung-Chul Han points out, big data is a rudimentary source of knowledge and AI is incapable of thinking (Han 2021), so now the role of humanists as Magister Ludi in this game of data becomes essential and immediate. As in Castalia, someone must oversee data science and the universe of non-things to establish their associations; in short, to forge a more human interpretation of the data we generate in our world.

Bibliography

Aalst, Wil van der (2016): "Data Science in Action", en Wil van der Aalst (Hrsg.), *Process Mining: Data Science in Action*. Berlin: Springer, pp. 3–23.

Alonso, Rodrigo (2020): "Qué es Big Data o Macrodatos y qué tiene que ver con el hardware", en *HardZone* https://hardzone.es/reportajes/que-es/big-data-macrodatos/ (letzter Zugriff 01.10.2021).

Alpaydin, Ethem (2014): *Introduction to machine learning (Adaptive computation and machine learning)*. Cambridge: The MIT Press.

Amato, Flora, Vincenzo Moscato, Antonio Picariello, Francesco Colace, Massimo De Santo, Fabio A. Schreiber & Letizia Tanca (2017): "Big Data Meets Digital Cultural Heritage: Design and Implementation of SCRABS, A Smart Context-awaRe Browsing Assistant for Cultural Environments", in *Journal on Computing and Cultural Heritage*, 10, pp. 1–23.

Arroyo-Machado, Wenceslao, Daniel Torres-Salinas & Nicolas Robinson-Garcia (2021): "Identifying and characterizing social media communities: a socio-semantic network approach to altmetrics", in *Scientometrics*, 126, pp. 9267–9289.

Bengio, Y. (2009): “Learning Deep Architectures for AI”, in *Foundations and Trends® in Machine Learning*, 2, pp. 1–127.

De Urbina, Amparo (2021): “Una propuesta de valoración de patrimonio urbano a la luz de Panoramio, una fuente Big-Data. El caso del centro de Bogotá”, en *Territorios* (45).

Dodson, Sean (2008): “Was software responsible for the financial crisis?”, in *The Guardian*, Abschn. Technology.

Faraway, Julian J. & Nicole H. Augustin (2018): “When small data beats big data”, in *Statistics & Probability Letters*, 136, pp. 142–145.

Favaretto, Maddalena, Eva De Clercq, Christophe Olivier Schneble & Bernice Simone Elger (2020): “What is your definition of Big Data? Researchers’ understanding of the phenomenon of the decade”, in Florian Fischer (ed.) *PLOS ONE*, 15.

Fayyaz, Zeshan, Mahsa Ebrahimian, Dina Nawara, Ahmed Ibrahim & Rasha Kashef (2020): “Recommendation Systems: Algorithms, Challenges, Metrics, and Business Opportunities”, in *Applied Sciences*, 10.

Fischer, Christian, Zachary A. Pardos, Ryan Shaun Baker, Joseph Jay Williams, Padhraic Smyth, Renzhe Yu, Stefan Slater, Rachel Baker & Mark Warschauer (2020): “Mining Big Data in Education: Affordances and Challenges”, in *Review of Research in Education*, 44, pp. 130–160.

Flach, Peter (2012): *Machine learning: the art and science of algorithms that make sense of data*. Cambridge: Cambridge university press.

Galeano, Pedro & Daniel Peña (2019): “Data science, big data and statistics”, in *TEST*, 28, pp. 289–329.

Gandomi, Amir & Murtaza Haider (2015): “Beyond the hype: Big data concepts, methods, and analytics”, in *International Journal of Information Management*, 35, pp. 137–144.

García Fernández, Aníbal (2018): “Cambridge Analytica, el big data y su influencia en las elecciones”, in *CELAG* https://www.celag.org/cambridge-analytica-el-big-data-y-su-influencia-en-las-elecciones/ (letzter Zugriff 15.12.2021).

Hai-Jew, Shalin (Hrsg.) (2015): *Enhancing Qualitative and Mixed Methods Research with Technology*: (Advances in Knowledge Acquisition, Transfer, and Management). IGI Global. doi:10.4018/978-1-4666-6493-7.

Han, Byung-Chul (2021): *No-cosas: quiebras del mundo de hoy*. Barcelona: Taurus.

Harford, Tim (2012): “Black-Scholes: The maths formula linked to the financial crash”, in *BBC News*, Abschn. Magazine.

Hasan, Md. Morshadul, József Popp & Judit Oláh (2020): “Current landscape and influence of big data on finance”, in *Journal of Big Data* 7, p. 21.

Hern, Alex (2019): “Cambridge Analytica did work for Leave.EU, emails confirm”. in *The Guardian*, Abschn. UK news.

Hesse, Hermann (2012): *El juego de los abalorios*. Madrid: Alianza Editorial.

Hindman, Matthew (2018): “Cómo funcionaba el modelo de Cambridge Analytica, según la persona que lo construyó”, in *El País*.

Jiménez de Luis, Ángel (2021): “Racista por defecto: la discriminación de los algoritmos que Silicon Valley no soluciona | Tecnología”, in *El Mundo*.

Khan, M. Ali-ud-din, Muhammad Fahim Uddin & Navarun Gupta (2014): “Seven V’s of Big Data understanding Big Data to extract value”, in *Proceedings of the 2014 Zone 1 Conference of the American Society for Engineering Education*, pp. 1–5.

Kitchin, Rob & Gavin McArdle (2016): “What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets”, in *Big Data & Society*, 3.

Kugler, Logan (2018): “The war over the value of personal data”, in *Communications of the ACM*, 61, pp. 17–19.

Laney, Doug (2001): “3D data management: Controlling data volume, velocity and variety”, in *META group research note*, 70, p. 1.

Li, Gang (2015): “Big data related technologies, challenges and future prospects”, in *Information Technology & Tourism*, 15, pp. 283–285.

Luttwak, Edward (2000): *Turbocapitalismo. Quiénes ganan y quiénes pierden en la globalización*. Barcelona: Crítica.

Mac, Ryan (2021): “Facebook Apologizes After A.I. Puts ‘Primates’ Label on Video of Black Men”, in *The New York Times*.

Marta-Lazo, Carmen (2018): “Calidad informativa en la era de la digitalización: Fundamentos profesionales vs. Infopolución”, in *Calidad informativa en la era de la digitalización*. Madrid: Dykinson, pp. 1–208.

Masís, Jethro. “Fenomenología Hermenéutica e Inteligencia Artificial: Otra Urbanización de La Provincia Heideggeriana.” *Actas de Las Primeras Jornadas Internacionales de Hermenéutica*, 2009.

Mayer-Schönberger, Viktor & Kenneth Cukier (2013): *Big data: A revolution that will transform how we live, work, and think*. New York: Houghton Mifflin Harcourt.

Müller, Karsten & Carlo Schwarz (2020): From Hashtag to Hate Crime: Twitter and Anti-Minority Sentiment. New York: SSRN Scholarly Paper Rochester.

Nisantzi, Argyro, Diofantos G. Hadjimitsis, Athos Agapiou, Vasiliki Lysandrou, Andreas Christofe, Marios Tzouvaras, Christiana Papoutsa, et al. (2018): “Remote sensing archaeology knowledge transfer: examples from the ATHENA twinning project”, in Nektarios Chrysoulakis, Thilo Erbertseder & Ying Zhang (Hrsg.), *Remote Sensing Technologies and Applications in Urban Environments III*. Berlin: SPIE.

O’Donnell, Benedict (2015): “Cracking the maths behind the economy”, in *Horizon: the EU Research & Innovation magazine* https://ec.europa.eu/research-and-innovation/en/horizon-magazine/cracking-maths-behind-economy (letzter Zugriff 09.12.2021).

Olson, Matthew, Abraham Wyner & Richard Berk (2018): “Modern Neural Networks Generalize on Small Data Sets”, in S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, & R. Garnett (Hrsg.), *Advances in Neural Information Processing Systems*, *31*.

Ortíz, Georgina (2019): “Toshiba ayuda en los experimentos del Gran Colisionador de Hadrones”, in *Big Data Magazine* https://bigdatamagazine.es/toshiba-ayuda-en-los-experimentos-del-gran-colisionador-de-hadrones-lhc (letzter Zugriff 01.07.2021).

Pascual, Manuel G. & Fernando Peinado (2018):” Cambridge Analytica ofreció sus servicios en España”, in *El País*.

Rands, Kevin. “How Big Data Has Changed Politics.” *CIO*, 2018, https://www.cio.com/article/221882/how-big-data-has-changed-politics.html.

Rodriguez-Sanchez, Francisco, Jorge Carrillo-de-Albornoz & Laura Plaza (2020): “Automatic Classification of Sexism in Social Networks: An Empirical Study on Twitter Data”, in *IEEE Access* 8. 219563–219576. doi:10.1109/ACCESS.2020.3042604.

Rosenberg, Matthew, Nicholas Confessore & Carole Cadwalladr (2018): “How Trump Consultants Exploited the Facebook Data of Millions”, in *The New York Times*.

Sarkar, Dipanjan, et al. *Practical Machine Learning with Python*. Apress, 2018, https://doi.org/10.1007/978-1-4842-3207-1.

Sawyer, Steve (2008): “Data Wealth, Data Poverty, Science and Cyberinfrastructure”, in *Prometheus*. Routledge, 26, pp. 355–371.

Scaife, A. M. M. (2020): “Big telescope, big data: towards exascale with the Square Kilometre Array”, in *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 378, p. 2166.

Schnase, John L., William P. Webster, Lynn A. Parnell & Daniel Q. Duffy (2011): “The NASA Center for Climate Simulation Data Management System”, in *2011 IEEE 27th Symposium on Mass Storage Systems and Technologies (MSST)*. Denver: CO, pp. 1–6.

Siegler, MG (2010):” Eric Schmidt: Every 2 Days We Create As Much Information As We Did Up To 2003”, in *TechCrunch* https://techcrunch.com/2010/08/04/schmidt-data/ (letzter Zugriff 25.03.2023).

Stewart, Ian (2012): “The mathematical equation that caused the banks to crash”, in *The Observer*.

Sun, Limin, Zhiqiang Shang, Ye Xia, Sutanu Bhowmick & Satish Nagarajaiah (2020): “Review of Bridge Structural Health Monitoring Aided by Big Data and Artificial Intelligence: From Condition Assessment to Damage Detection”, in *Journal of Structural Engineering*, 146, 04020073. doi:10.1061/(ASCE)ST.1943-541X.0002535.

Torgler, Benno (2020): “Big Data, Artificial Intelligence, and Quantum Computing in Sports”, in Sascha L. Schmidt (Hrsg.), *21st Century Sports (Future of Business and Finance)*. Berlin: Springer, pp. 153–173.

Triguero, Isaac, Diego García-Gil, Jesús Maillo, Julián Luengo, Salvador García & Francisco Herrera (2019): “Transforming big data into smart data: An insight on the use of the k-nearest neighbors algorithm to obtain quality data”, in *WIREs Data Mining and Knowledge Discovery*, 9, e1289. doi:10.1002/widm.1289.

Viégas, Fernanda Bertini, and Martin Wattenberg. *Wind Map*. Interactive software, 2012, https://www.moma.org/collection/works/163892.

Ward, Jonathan Stuart & Adam Barker (2013): *Undefined By Data: A Survey of Big Data Definitions. arXiv*. doi:10.48550/arXiv.1309.5821.

Webb, Amy (2021): “Los nueve gigantes”, in *Cómo las grandes tecnológicas amenazan el futuro de la humanidad*. Barcelona: Península.

Zgurovsky, Michael Z. & Yuriy P. Zaychenko (2020): *Big Data: Conceptual Analysis and Applications* (Studies in Big Data). Berlin: Springer.

Ana Gallego Cuiñas

Literature Seen Through Big Data and Artificial Intelligence: Key Concepts and Critical Challenges

> In the clinic of the art of reading,
> the one with the best vision is not
> always the one who reads best
> Ricardo Piglia

These lines from *El último lector* (*The Last Reader*) (2005) by Ricardo Piglia clearly reflect the fact that the literary is a matter of perspective or scale, a radically historical ideological and aesthetic positioning that constructs a truth. For Piglia, the ideal reader is one who literally cannot read well – and in this statement there are indubitable echoes of Harold Bloom's *The Anxiety of Influence* (1973) – because their vision (or point of view) (Kittler 2010) compels them to read *up close*, like a short-sighted person who needs a magnifying glass to make out anything tiny and particular – the hidden structure of text that becomes a system of secret correspondences that have to be uncovered in every era. This metaphor perfectly illustrates the critical approach of *close reading* (Empson 1966; Richard 2004), based on hermeneutics and/or narratology, which has predominated in literary studies since the beginning of the last century. The critical obverse of this would be *distant reading* (Moretti 2016), sociological and/or quantitative in nature, which was developed in the second half of the twentieth century and would fit the metaphor of the *far-sighted* reader, who with the perspective of distance can access the general context of the texts to establish formal and material relations of a discursive, social, cultural and economic nature.

In this paper, I take this dual perspective as a basis for the following proposal: literary criticism of the twenty-first century needs to overcome this counterposition in approaches, misleadingly understood as opposites, to practise a combined mode of reading in which textual interpretation, materialism and datalism complement each other, for the sake of a more thorough and organic intellection of the literary fact and of its social function. The hypothesis I begin with is that, on the one hand, the distant reading that the social sciences and computational techniques adopt is ever more necessary to analyse the aesthetic and material function of literature in society, both in diachrony and synchrony. On the other hand, all data analysis requires careful – *close* – attention to the structures, qualitative and quantitative, that appear as a result of research. Thus the epistemic crossover between literature, sociology and big data is not only possible but

https://doi.org/10.1515/9783110753523-003

desirable, since through the intertwining of these scales – or methodological strategies (English and Underwood 2016) – we can achieve better findings with greater breadth and concision, which benefits both theory and literary criticism, and computational science. We could call this approach *cross-reading*,[1] which in turn suggests another ocular metaphor, *cross-eyed reading*, displaced and interposed – a form championed by Ricardo Piglia himself (2005) as a highly productive way of reading for the Argentine cultural field.

This essay therefore presents a *crisscrossed* and *situated* reflection on Literature and Big Data centred on two main themes that overlap and supplement one another: first, the use of Big Data and Artificial Intelligence (AI) in literary culture, based both on the mechanisms of production, circulation and consumption of literature in the market, and on the impact and utility of computational methods in the field of criticism; second, the use of certain literary and philosophical categories that could be advantageous for the – "situated" (Haraway 1995) – epistemic and political intellection of the functions of Big Data and AI. This proposal is undoubtedly only a starting point, which has the ultimate aim of contributing to the much-needed design of an agenda for the *literary criticism of data*[2] that contemplates the multiple possibilities of collaboration and dialogue between the Humanities, Sociology, Data Science and AI.

1 Use of Big Data in Culture and Literary Criticism

The first thing that needs to be stated in this introduction is that to think about the integration of big-data techniques or, what amounts to the same, about quantitative methods of measuring data, in the sphere of literature is to base our thought, first of all, in the social question (Halavais 2015) – that is to say, in the sociology of literature. This discipline examines the literary object as social fact or product, common and collective, which crystallizes the eternal conflict between technology (numbers) and culture (letters). The criticism of concentrating and standardizing cultural objects has its roots in Adorno and Horkheimer's *Dialectic of Enlightenment* (1944), since when this tension has been constantly re-

1 Bootz and Laitano (2014) put forward the same name to designate a data visualization model based on Spinoza's ontology. My idea, however, points to the crossover and simultaneity of two methods (close- and distant-) that, in the historiography, have been seen as opposites.

2 One could also write *literary dataism* or *computational literary criticism*, but these terms are more restrictive than the one I propose, which is more labile and versatile. This modality would fall within the epistemology of 'Critical Data Studies', based on the study of data and their different critiques, systematized by Dalton and Thatcher (2014).

peated one way or another in "literary culture" (Gallego Cuiñas 2022). Adorno, remember, takes up Benjamin's thesis on the alienation of a work of art through its commercial reproducibility – circulation – which would have to be in opposition to its *aura*, the Kantian authenticity, of a cultural object. The attack on mass culture and mercantile utilitarianism is evident, as is Benjamin's romantic and idealized view, by Adorno himself and later by Guy Debord in *The Society of the Spectacle* (1967). Today, the consumption of Art reveals itself in its pure contingency more than ever (Gallego Cuiñas 2021), in that being-in-the-moment and in the ephemeral that the School of Frankfurt and its followers despised. Therefore, if we wish to read from the present and out of contingency, as a privileged mode of literary production, in the studies of literature we would have to include a sociological focus and big data, the volume and speed of which have increased exponentially over the last decade with an impact in the culture sphere that is both material and symbolic, and which we cannot avoid:

> In the arts and humanities, the notion of big data is still in its embryonic stage, and only in the last few years, arts and cultural organizations/institutions, artists, and humanists are starting to investigate, explore, and experiment the deployment and exploitation of big data as well as understand the possible forms of collaborations. (Schiuma and Carlucci 2021: xxiv)

In the last five years, studies on the humanities and quantitative methods have been appearing with much more frequency, particularly in English-speaking academia, followed by the French, who historically have had more porous borders between the Humanities and the Social Sciences than the Hispanic world, which facilitates the transdisciplinary crossover. In the specific case of literary studies, the most distinguished researchers in data or computational criticism are North American: Paul Delany and George Landow, Matthew Jockers, Andrew Piper and Ted Underwood. In the Iberian-American world, some names of note are: Belén Gache, Claudia Kozak, Carolina Ferrer, Carolina Gainza, Germán Ledesma, Diana Roig-Sanz, Germán Sierra and Alex Saum-Pascual.[3]

This leads me to the second aspect I would like to make clear from the start: the utilization by sociology and data science in literary studies transcends the concept of *Digital Humanities*. I agree with Underwood that this label is more of a reaction to a tactic – which has strengthened the use of digital technology and open science, essentially through the idea of the archive, digitalizing and cataloguing historical texts that are difficult to access – than referring to an area of

3 It needs to be made clear that Iberian-American literary criticism of data has left the sociology of literature to one side, which I believe is fundamental for the material understanding of the literary, not only the aesthetic understanding.

knowledge in itself. What has undoubtedly occurred is a *digital turn* (Dobson 2019), which in the second decade of this century has become a *computational turn* whose epistemic value is founded on the ecosystem of quantitative methods that the sociology of culture has traditionally used. Today these methods for measuring have also become methods for mediation – we are all now digital researchers – at the mercy of the aforementioned phenomenon of the intensification of datafication and of the advances in AI for creation and cultural consumption.

1.1 Literature, Market and Artificial Intelligence: Some Hypotheses

There are three main ideas that currently underpin the production, circulation and consumption of literature: the figure of the writer, the literary work and the reader. In criticism, these categories have become "zombies", insufficient to express the "new" state of what is literary (Gallego Cuiñas 2019) in which orality and print, the human and the digital, literature and literary culture live together. The space and the computational techniques have broadened the creative experience, to the point at which the digital environment has itself become a medium for the production and distribution of literature, with ever more enthusiasts. At the same time, we rely more and more on the predictive potentiality of big data (Sádaba Rodríguez 2020), both in reception studies and in the creative industry, to evaluate trends in the literary and artistic market, the quality of products and the degree of user satisfaction (Piper and Portelance 2016; Schiuma and Carlucci 2021). How therefore does Big Data and AI affect the ontology and epistemology of the literary?

Starting with this question, I propose some core ideas for reflection on the use of data analysis in the sociological and materialist approach of literature, which (re-)opens several lines of research for today's academia:

From author to '(artificial) writer'. Big Data is a highly advantageous instrument for the development – in literary studies – of what I call *writer criticism* (Gallego Cuiñas 2022), which is based on the sociological, materialist and aesthetic analysis of the figure of the writer, utilizing new methods and elements that have not received sufficient attention: the use of digitalized biographical archives; the production and reception of writer bots; the authorial image on social media; *bookporn*, interaction with other mediators of literary culture; performativity in the public sphere; the extent of education and literary professionalization, and so on.

However, in the creative sphere, *artificial creation* or the literary production of texts by an AI has imperilled the pristine category of 'author'. This takes us back to the same conundrum that the Frankfurt School detected regarding the

loss of the work of art's aura in the first half of the twentieth century, now applied to the anthropocentric notion of the author figure as the intellectual property holder of a text, and questions the author's hegemony and validity (Badía Fumaz 2012; Berti 2015; Herrmann et al. 2022). In the same way, the romantic idea of the author as genius creator, based on the symbolic value of human and individual literary creation, which is difficult to put a value on, has become unsustainable. In contrast, a mode of artificial creation, collective in origin, is growing, with a symbolic value that is more easily quantifiable in material and economic terms. In both cases, in the literary field the trend is to talk of the (artificial) 'writer' rather than the author.

To this we can add the commercialization of the aforementioned creation of *artificial works* – to give them a name – made by an algorithm, most in open access, and aimed at mass consumption, which until now have been 'overseen' by human writers. As in other areas of creative industry (i.e. music, art, photography, graphic design, et cetera), this new mode of (digital) literary production is being proclaimed as an attractive mode of exploitation and extremely beneficial for the cultural industry and institutions, which, with the control and use of an AI in the creation of an artistic work, can also become the co-authors – not merely co-producers – of texts.

Artificial creation, therefore, generates at least three theoretical and political problem areas:

(i) First, the entry into crisis that I have already mentioned – a new death? – and resignification of the notion of authorship tied to the concepts of 'authenticity' and 'intellectual property', which would shift from being individual to collective (the final text is the result of an algorithm that works with the big data obtained through millions of works), from being human to technological. The literary algorithms, avatars or bots (Olaizola 2018; Sierra 2022: 13) that automatically create literary content are also an example of the way in which digital production contributes to the performativity of the category of author, which changes to that of 'writer' (Gallego Cuiñas 2022), given that the capitalist notion of authorship is being displaced, and it is becoming very difficult to distinguish intellectual property.[4]

(ii) Second, the place of the writer in the creative process transforms and shifts from the romantic value of the genius who produces an original and unique work, to the pre-capitalist value of co-creation, appropriation and communi-

4 Literary (ro-)bots are algorithms that produce content, above all on social networks: "What distinguishes bots from other types of software is that they interact with and or produce content for human users, often taking on a human personality" (Olaizola 2018: 239).

tarian transmission of the work produced by an AI. Therefore, the writer would have to become a craftsman or a mixer, mediator or *gatekeeper* (Gallego Cuiñas 2022) of the resulting artificial work.[5]

(iii) Third, this mode of digital (re-)production results in the loss of bibliodiversity (cultural, of genres, authors, et cetera), and in the dangerous increase in colonial and gender biases (the majority of the works collected in databases are written by men and edited in cultural systems of the north – that is, of hegemonic cultural systems), unless, in the artificial creation phase we find ourselves in, there is a gatekeeper or guarantor of these egalitarian, decolonial and inclusive values.

From the work to the '(artificial) work under construction'. We first need to distinguish between the works that are born and are (re-)produced in the digital medium,[6] those that are hybrid (print publishing and digital technology), and those that are digitalized. However, they all share five essential traits: "digital or numerical representation, modular composition, variability, automatization, and transcoding" (Berti 2018: 139). Second, the alphabetic and binary codes are interpretable aesthetically and computationally based on common notions such as: mutability, contingency, collectivity, anonymity, fragmentation, brevity, materiality, and gamification. Based on these premises, I have come up with three hypotheses:

(i) From the creative point of view,[7] it is clear that the production of digital literature[8] is a modality that has been growing in Ibero-America[9] in recent years, mainly through the practice of poetry (cf. Gache 2006; Kozak 2010 and 2017; Berti 2015; Gainza 2019; Ledesma 2022; and Saum-Pascual 2022). This literature experiments with the signifier, with multimedia elements and with the archive[10] through algorithms (cf. Bolter 1991; Hayles 2008, or Côrtes Maduro 2017), which is why it is often associated with the notion of

5 It is clear that this mode of artificial production of literature recovers – literally – the notion of tradition as the great producer of texts, something that Borges stood for.

6 The large majority of texts today naturally appear digitally before in print form.

7 Remember that it was precisely with the publication of Mary Shelley's *Frankenstein* when creativity became a great (anthropocentric) value associated with divinity.

8 To mention a few notable names in Spanish-language digital literature: Belén Gache, Iván Marino, Luis Espinosa, Marina Zerbarini, Mariano Sardón, Gustavo Romano and Alex Saum. Argentines lead the list in numbers, followed by Spanish writers.

9 It has been developed and studied more in the English-speaking world. See the following databases: Electronic Literature Knowledge Base https://elmcip.net/; Electronic Literature Collection https://collection.eliterature.org/ or NETescopio https://proyectoidis.org/netescopio/.

10 On the one hand, digitality is a mode of production and a materiality, and on the other, digital literature works with the existing tradition, with what is repeatable becoming prime material.

avant-gardism[11] or with experimentalism[12] – not only aesthetic (of a formalist stamp) but also technical.[13]

The *live* creation of artificial work – cyberwriting (i.e. twitterature, instapoetry, literary memes and avatars, transmedia narrative or playwriting, Wonderbook, and literature made on WhatsApp or Wattpad) – has also been gaining greater visibility. This generates literary value that one could call *relational*, based on the participation of the reader – hence playing with Didi-Huberman's phrase "work under construction" (2015:16) – and on the 'live' or 'serialized' consumption of *readable* fiction, simple and direct.

Lastly, there is a boom in artificial works that are recycled or reworkings (remix or sampling), *works of works*, that incorporate GPS, audiovisual content or QR codes, where the extraordinary transmediality and performativity of the literary in the 21st century is at the forefront.

(ii) From the critical point of view, computational technology helps to probe into the nature of the written work, both from past eras and in the present: types of language, styles, biases, techniques or genres that have been adopted the most over the years and across cultures, which are easily studied through the mass digitalization of literary texts (i.e., Books Ngram Viewer, Blatt 2017 or CATMA). The data analysis of digitalized works lends itself to a predictive criticism that can calculate the success of a text, construct series (of texts) and evaluate their level of innovation.[14]

(iii) From the material point of view, the artificial literary work transcends the book-object as the receptacle of the text (Striphas 2011). The machines for making and selling literature cease to be the printing press and the distributors, for now it is the digital medium and its new formats that create and distribute it on platforms and social networks. The tools of production, publishing, reading and conservation of digital literature have changed radically in the second decade of this century, to the point at which some computer skill is required for it, although programs are being designed that are ever easier and more democratic to use.[15]

11 I am thinking about visual, concrete and sound poetry and their performative performance.

12 See the anthology of experimental literature compiled by Tomás Vera Barros (2014).

13 Rafael Pérez y Pérez is one of the most outstanding researchers in computer creativity and has produced several books with AI (see http://www.rafaelperezyperez.com/). Questions abound: Do algorithms have an aesthetic? What form of appropriation is it by the author regarding the product generated by AI? If the authorship is held by the publishers, are we returning to a hegemonic authorship? To a post-human authorship?

14 In this regard, in a few years Wattpad will be able to write its own stories based on the big data it has obtained from the success of certain stories on the platform.

15 María Goicoechea de Jorge explains: "These types of programs have enabled a greater number of authors to access this genre who are not necessarily connected to the academic world or

From reader to (digital) 'prosumer'. The reader of digital literature is always a co-producer or a prosumer (Villanueva 2022: 5), because the interaction with the work is a consubstantial part of the reading process. In fact, the artificial work acts as a kind of toy that is both literary and computer (with readings coded according to the text and to the use of technology), which goes along with a contingent and non-standardized use. The temporary nature of digital reading is manifold (it goes backward and forward, it ends, it breaks into parts), transmedial and simultaneous, non-linear, and successive like printed reading. But reading is also conceived in series and intermittency – like the nineteenth-century serialized novel or mass-culture subscription-based instalments – from the same digital setup, as occurs with Serial Box or in Spanish with the platform Black & Noir, which operate like distributors on mobiles and tablets of serialized literature.

Furthermore, Big Data has been heavily used in the study of audiences, with methods based mainly on Singular Value Decomposition (SVD).[16] The data analysis of reception measures the way in which rating patterns change and how literary prestige is formed and circulated. The 'stock exchange' that underlies aesthetic judgement has barely altered over the last century (Underwood 2019), since it has always been in the hands of the same authorities of the market and academia: institutions, universities, publishers, prizes, critics, et cetera. However, the democratization of taste that has gone hand-in-hand with digital and technological progress has undeniably impacted the appraisal of literary value, in such a way that not only are we witnessing an unprecedented proliferation of producers of literature and literary products, but also of readers/consumers, 'digital prosumers' who rate literary value on platforms such as the aforementioned Wattpad, or Goodreads, a social network of readers and writers who act as critics of other books and who influence the prescription of taste (Bourdieu 2002).[17] In the wake of this shift from academia and the cultural press as agents of literary value, we find *booktubers*,

to research. The following are three of the currently most popular programs with their most notable characteristics and differences: Twine, created by Klimas in 2009 with a free software licence; Inklewriter, a tool created by the games company Inkle, co-founded by the British mathematician and writer Jon Ingold; and Undum8, created by Millington in 2010 with an MIT licence (Figure 4). The importance of these types of programs is that they have democratized the use of this genre of digital literature, since advanced programming knowledge is no longer needed to write a narrative or interactive game" (2019: 175).

16 Singular Value Decomposition (SVD) is a technique used in 2009 to predict user ratings for films on Netflix.

17 Up until now, there are hardly any comments on self-published books on Amazon and similar platforms. Instead, the majority are from the usual publishers. We should also consider the content – book – recommendations that users make based on their consumer experiences or the data platforms based on algorithms (Vanoli 2019: 27–34).

bookstagrammers and Amazon algorithms – "symbolic expropriation" is what Jorge Carrión has come to call the *modus operandi* of the online retail site – which condition the book choices of (digital) mass culture upon the basis of patterns of consumption carried out using Big Data. Thus, Amazon acts like a virtual and democratic bookshop (Lefort-Favreau 2021: 79) that prescribes taste and dictates the norm – being a new instance of value appraisal that is consumerist and populist in style – by virtue of the quantitative concentration of information, which always entails a certain standardization: the tyranny of the masses, which replaces the former tyranny of the elite, of the mesocratic bourgeoisie that has dominated the construction of literary value in the modern world (Gallego Cuiñas 2019).

To finish this section, we cannot forget that the opposite phenomenon also exists: the appraisal of literary value with digital parameters that perpetuate and defend an elitist community, a ghetto, of prosumers of literature: "The electronic art and visual poetry market have adopted the NFT (non-fungible token) as the preferred format for diffusion and sale. But it no longer only applies to the visual arts, but also to texts, mixed artworks and even novels" (Sierra 2022: 13).

1.2 Literary Criticism and Big Data: A New Challenge for the Sociology of Literature

The sociological study of literature, which was prevalent in the 1960s and 1970s in Latin America, today only makes up between one and two percent of academic publications on literature in the Spanish language (Gallego Cuiñas 2022). The majority today take Pierre Bourdieu's perspective,[18] yet few dare to use quantitative and computational methods for literary analysis (i.e., Roig-Sanz 2019 and Gallego Cuiñas 2022). This shows that the question about the nature of the sociology of literature and about which methodological instruments they should use continues to be relevant and highly debated today. In this context, the use of a *critical dataism* seems like a very fertile space of expansion and experimentation both for the sociology of culture and for the literary studies of the future.

Let us remember that the articulation of the sociology of literature as framework of thought dates back to Marxist structuralism, but it did not begin to be developed as a discipline until the sixties, with the Birmingham School. Subsequently, in the seventies and eighties, a generation of cultural and literary sociologists emerged that carried it on into its brightest period. The Marxist approach was, from the 1990s, then displaced by the advance of New Materialism and the

18 See Moraña (2014) and Maltz (2020) on the colonization of Bourdieuan thought.

application of an anti-hermeneutic and anti-aesthetic methodology that Moretti called "distant reading". The weaknesses of this clearly positivist method have already been pointed out, although this does not, in my opinion, invalidate the idea that the sociology of literature and dataist technique, with statistical and computational methods, could prove politically advantageous for twenty-first century literary criticism, given that literature is historical and ideological merchandise, tied to the real economy, and it depends, in its dispositions, sociabilities and affects (Brouillette 2017: 280), on the numerical logic of the economy, on the possibilities of the digital and on the big data of the market. Value can undoubtedly be extracted from these – they give us a pattern, new forms of production, association and a forecast – because they represent and transform, materially and symbolically, literary taste. In other words, this new form of producing knowledge can lead us to the configuration of a new epistemic field.

How can we therefore give legitimacy to a sociology of literature based on dataism today? There is no avoiding the fact that one of the most important problems in literary criticism is precisely the legitimacy of the method or theoretical approach of the researcher. In reality, this is a question of strategy in the academic struggle for intellectual capital, safe from self-absorbed and centripetal methodological trends, where what is really at stake is the professional standpoint of the critic, not the conceptual make-up of a field (the knowledge or cognition), but recognition (Morgan 2013) in a sparsely populated ivory tower of specialists. Hence quantification – being associated with the social sciences and positivism – becomes a twofold enemy for the critic and theorist of literature, since it presupposes both the mix with sociology and an attack or questioning of the qualitative methods inherent in the humanistic field. However, one thing does not exclude the other, and in the third decade of the 21st century we cannot keep turning our backs on the myriad of resources that digital culture makes available for the theoretical, sociological and historical study of literature. Its use offers us a tool, not a substitute for but a complement to criticism and creation, already commonplace in the sphere of linguistics and historiography (Lemercier and Zalc 2019) – which are highly familiarized with working with corpus and archive – and becoming more so in other arts, although up until now they have barely used the databases, sources and samples of data on a large scale in the humanities.

To speak plainly, many humanists question the validity of quantitative methods, which they brand as neoliberal,[19] without understanding them or having tried out their uses politically, which in some cases represent real challenges for left-

19 We cannot deny that only universities in the northern world can use these highly expensive methods, which enable access to data. In the end, information is power.

wing, materialist literary criticism. Neither numbers nor quantifications are intrinsically objective or bad: they are merely signs, and as such, depend on (ideological) interpretation. This is why "numbers are becoming more useful in literary study for reasons that are theoretical rather than technical" (Underwood 2019: xi). Why? What can data science provide literature with? The techniques of quantification expand the scope of our study toward new forms of representation – such as data visualization (Karsdorp et al. 2021) – and toward new ends, at the same time as providing the aforementioned *relational value* – as Saussure and structuralism understood it – through the configuration of different statistical, digital and computational 'models' or 'structures', which strengthen theoretical and critical analysis, focused on themes, problems, genre(s), characters, periods, authors, et cetera (Piper 2017). Likewise, the access and handling of big data (re-)opens sociological lines of research – not widely explored in Hispanism – that can be developed through this approach, without giving in to data fetishism. The three that I believe have the most political repercussion are:

i. **Study of invisibilized works.** This is the area of interest of Moretti (2016), focused on the possibility of accessing the big data provided by texts that have been marginalized – historiographical blind spots and gaps – by the hegemonic mechanisms of recognition, which have generated the canon of literature and its modes of representation (Bode 2017; Roig-Sanz 2019).
ii. **Study of taste**. There are two options here: one aimed more at academic criticism, which Carolina Ferrer calls "criticometría" ("criticometry") and which entails the bibliometrical or citation analysis of certain critical theories and trends in academic publications – in different times and spaces – on the database of semantic associations, of repetition and generalization, in the geopolitical context of their utterance. This helps to trace the global map of academic geopower relations, and of their capital, in every era (Goldstone and Underwood 2014; Ferrer 2015; Espino 2020). The second option is focused on the cultural field, through the analysis of newspapers, notes, journals, digital content, prizes and other discourses that can contribute to learning the way in which literary value has been appraised and how social prestige is constructed outside the academy (Underwood 2019, 69),[20] also taking into account its variations in time and space (Martínez-Gamboa 2016; Posada 2019). One example is the book by Archer and Jocker, *The Bestseller Code: Anatomy of the Blockbuster Novel* (2016).[21]

20 For example, in his study Underwood shows with quantitative methods that the way we judge a literary work generally changes every thirty years.

21 The problem is that up until now, this type of study has omitted the material analysis of gatekeepers – publishers, translators, agents, etc. – which is essential, from my point of view, for con-

iii. **Study of figurations and networks of sociability**. In the area of research pioneered by de Nooy (1991), analyses have been carried out – first in psychology and then in sociology (Lemercier and Zalc 2019, 101) – of networks in this new era of Big Data (i.e., Jean So and Long 2013; Gallego Cuiñas et al. 2020) on the connections that are produced in order to build *value networks* on digital platforms, using content and profiles on social media (Twitter, Facebook, Instagram, Linkedin). This seems to be a highly productive opportunity for understanding the way in which the figures and figurations of the writer are currently constructed, as I stated earlier, but also for examining the role that algorithms, avatars, bots and intermediaries (i.e. publishers, agents, other writers, Granta, festivals, et cetera) take in the promotion of a work, a genre or an author, and their symbolic and financial resources.

Despite the new research areas that this sociology of literature based on the analysis of big data and on AI opens up, we cannot ignore the fact that, currently, the traditional *close reading* is still predominant in academic publications, and therefore provides much more professional assurance than this new agenda that, at the moment, does not enjoy the same prestige in our field. The price a humanist has to pay for expanding their discipline's horizons is high, since not only are they faced with another discipline that they have to learn but also with institutional and material problems deriving from the lack of technical training and infrastructure,[22] as well as the academic loss of worth as judged by the agents who control the field (Underwood 2019: xviii): journals, publishers, assessment agencies, departments, institutes, and so on. In short, the impact of *new* methods and study aims is always slow, and at first incurs rejection in the disciplines of origin, thus making the decision to opt for this type of research evidently riskier and more unrewarded.[23]

structing literary value in the contemporary world. The combination of both perspectives would give a more thorough and complete interpretation of the modes of production and circulation of value in the literary field.

22 Thus the wealthier northern academia, with better material conditions, are always the pioneers in taking up innovative methodologies.

23 This is why much of computational criticism carried out on humanist objects is being done by computer scientists, specialists in information and communication sciences, economists, engineers, and so on (Schiuma & Carlucci 2021).

2 Use of Literary Categories in Data Science

There is no doubt that the fact of dataism needs an interpretation, needs a *situated* narrative meaning. This assertion opens the door to the possibility that literary criticism has something to contribute to data science and not only the other way around.[24] What am I referring to? That we find theoretical categories and critiques of analysis – principally from Russian formalism and from (post-)structuralism – that help to explain the functioning of the algorithm and to articulate a kind of *Big Data hermeneutics* that will illuminate the critical and political thinking of computational methods and results. Namely:

i. **Close Reading** (Empson and Richard). This eminently literary strategy is fundamental for supervising the algorithms and for data interpretation that guarantees the certainty and efficacy of the results (Koskimaa 2005). The creation of the algorithm is also a *reading machine*, to use the Deleuzean metaphor; in other words, it is a model for reading. Hence every data reader is a co-producer of a *significant* structure, which comes from micro-thinking, not only macro-thinking: "thinking small in order to think big" (Piper 2018: 9). As well as knowing how to read between the lines, this involves acting as a kind of mediator or guardian of knowledge – that is, a gatekeeper (Gallego Cuiñas 2019) – that vouches for the value of the knowledge generated. This entails the differentiation and discerning of information, the removal of bias, and ensuring the quality of data: *Smart Data*. Thus, in computational criticism, the humanist (the ethnographer, the philosopher, and the philologist) becomes a gatekeeper because the authority, "the law" – in the Kafkian sense from the parable "Before the Law" – is still essential not only for giving *meaning* but also to *situate* and make visible gender, geopolitical and colonial inequalities that the algorithms do not *see*:

> The proficient and valuable use of big data needs the personal and organizational capacity of asking the right questions and in the right way. Big data is powerful only if it is generated, combined, or supported by the creation of strong narratives, organizationally and contextually framed. This means that the big data has to be "thick", i.e., not only quantitative but most importantly qualitatively relevant (Schiuma and Carlucci 2021: xxv).

24 "The hypothesis of the mutation of art due to digital transformation has been widely accepted, but it is also worth reversing these suppositions. As Kenneth Goldsmith states, 'if one thinks about it, the engine that drives the internet is literature [...]. It gives the possibility of cutting, copying and pasting, imitating the movements of language. Language has never been moved in the way that we are moving it today (14 February 2014)'." (Helgueta Manso 2022: 43).

ii. **Series and construction** (Tynyanov). These concepts belong to Russian formalism. The former refers to the property that the literary text has of breaking down into different units of meaning, the same procedure that algorithms use today. The latter refers to the "constructive function" and "relational function" of literary works, texts and units in similar "series" or systems of correspondence, as occurs in computing. Obviously the selection of texts, topics or units of meaning has a subjective or immanent component in literary criticism, as the algorithmic training of data processing also has, which presupposes a "value in itself" of the elements (i.e., *Topic Modelling*). This is why human – and humanist – readers are needed, to supervise the constructed computational models, since the aforementioned *relational value* is responsible for the recontextualization of texts in series – and one must remember here that a context is a point of view – as well as for its decontextualization.
iii. **Intertext** (Bakhtin and Kristeva). The theory of intertextuality, structuralist in origin, is based on the assumption that all text refers to other texts (in ideas and statements, in diachrony and in synchrony), in a more or less evident way (Pozuelo Yvancos 1994). This *value* of repetition or of the quotation also works as the constituent principle of algorithms that work with big data to come up with the correlations of a "series" or accumulation of meanings. Moreover, it is interesting to bring in here other literary notions such as "parody" and "irony", which require human involvement for their interpretation: machines operate with quotations or literal reproductions and this distorts the meaning.
iv. **Rhizome and Diagram** (Deleuze and Guattari). The epistemological definition of rhizome is well known and appeals to concepts that explain, many years in advance, the functioning of data science: multiplicity, modification, lines of flight, the calque and replication, connections and associations, as well as the absence of a centre and of a hierarchical model. The same occurs with the Deleuzean notion of diagram, which is chaos and seed, a "possibility of fact" and a "modulation". Both ideas appear to me to be fundamental for thinking theoretically about the form and procedure of *Big Data.*

To conclude: methods of analysis "tend to be concealed, are legitimized as neutral in themselves, as supposedly independent" (Rodríguez 2011: 95), but they are not. The problem lies in that one must know how to grasp those interdependent and transdisciplinary relations, which are often "invisible" (cf. Merleau-Ponty 1979). It is humanists who can do this, because they are the ones who have the competence of *crossed* and *situated* reading, although the task represents an epistemic and academic challenge. I am convinced that these days there is no sense in separating literary criticism – its ideological construction – and data analysis – quan-

titative and computational methods – although the former deals with the object in a simultaneous order – synchronic and micro – to unravel its principles and its limits, while the latter does so in a chronological order – diachronic and macro – to situate specific literary productions in a historical process that answers to a given social matrix, not exempt from colonial and gender biases. Computational criticism supplies the appropriate set of tools for the theoretical, historical, material and aesthetic knowledge of the literary work, but in turn this science is modified and is augmented with humanistic, philosophical, feminist and decolonial tools. Ultimately, the relationship between literature, big data and artificial intelligence does not only point to *other* forms of knowledge and representation, but to new *crossovers* between the theory and the praxis that create value: social, cultural and academic.

Bibliography

Altamirano, Carlos y Beatriz Sarlo (1991): *Literatura y sociedad*. Buenos Aires: Centro Editor de América Latina.

Archer, Jodie y Jocker, Matthew L. (2016): *The Bestseller Code: Anatomuy of the Block-buster Novel*. New York: St. Martin's.

Bajtín, Mijail (2012): *Estética de la creación verbal*. México: Siglo XXI.

Berti, Agustín (2015): *From Digital to Analog. Agrippa and Other Hybrids in the Beginnings of Digital Culture*. Nueva York: Peter Lang Publishers.

Badía Fumaz, Rocío (2012): "Muerte del autor y literatura digital", in *Eikasia: revista de Filosofía*, 44, pp. 113–127.

Blatt, Ben (2017): *Nabokov's Favorite Word Is Mauve: What the Numbers Reveal About the Classics, Bestsellers, and Our Own Writing*. New York: Simon & Schuster.

Bloom, Harold (2009): *La ansiedad de la influencia. Una teoría de la poesía*. Madrid: Trotta.

Bode, Katherine (2017): "The Equivalence of 'Close' and 'Distant' Reading; or, Toward a New Object for Data-Rich Literary History", in *MLQ*, 78, pp. 77–106.

Bolter, Jay David (1991): *Writing space: The computer, hypertext, and the history of writing*. Hillsdale, NJ: Lawrence Erlbaum.

Bootz Philippe and Maria Ines Laitano (2014): "Cross-Reading: un outil de visualisation de close readings", in *Revue Formules*, 18, pp. 271–288.

Boudieu, Pierre (2002): *Las reglas del arte*. Barcelona: Anagrama.

Brouillette, Sarah (2017): "Neoliberalism and the Demise of the Literary", in Michum Huehls and Rachel Greenwald (eds.). *Neoliberalism and Contemporary Literary Culture*. Baltimore: John Hopkins University, pp. 277–290.

Côrtes Maduro, Daniela (ed.) (2017): *Digital Media and Textuality. From Creation to Archiving*. Bielefeld: transcript.

Dalton, Craig and Jim Thatcher (2014): "What does a Critical Data Studies look like and why do we care? *Society + Space*". https://www.societyandspace.org/articles/what-does-a-critical-data-studies-look-like-and-why-do-we-care

Debord, Gay (2005): *La sociedad del espectáculo*. Valencia: Pre-textos.
Delany, Paul, Landow George P. (eds.) (1991): *Hypermdia and Literary Studies.* Cambridge/London: MIT Press.
Deleuze, Gilles (2008): *Pintura. El concepto de diagrama*. Buenos Aires: Cactus.
De Nooy, Wouter (1991): "Social networks and classification in literature", in *Poetics*, 20, pp. 507–537.
Didi-Huberman, Georges (2015): *En la cuerda floja*. Santander: Shangrila.
Dobson, James E. (2019): *Critical Digital Humanities*. Urbana: University of Illinois Press.
Empson, William (1966): *Seven Types of Ambiguity*. New York: New Directions.
English, James and Underwood, Ted. (2016): "Turbulent Flow: A Computational Model of World Literature", in *MLQ*, 77, pp. 345–367.
Espino, Francisco (2020): "Big data, criticometría y el estudio de las literaturas nacionales en la bibliografía crítica: el caso excepcional de la literatura cubana", in *Revista de Humanidades Digitales*, 5, pp. 66–85.
Ferrer, Carolina (2015): "Digital Humanities, Big Data, and Literary Studies: Mapping European Literatures in the 21st Century", in *Rupkatha Journal on Interdisciplinary Studies in Humanities*, 1, pp. 1–11.
Gache, Belén (2006): *Escrituras nómades*. Gijón: Trea.
Gainza, Carolina (2019): *Narrativas y poéticas digitales en América Latina: producción literaria en el capitalismo informacional*. México: Secretaría de cultura / Centro de Cultura Digital.
Gallego Cuiñas, Ana (2019): *Las novelas argentinas del siglo 21. Nuevos modos de producción, circulación y recepción*. New York: Peter Lang.
Gallego Cuiñas, Ana, Esteban Romero-Frías y Wenceslao Arroyo (2020): "Independent publishers and social networks in the 21st century: the balance of power in the transatlantic Spanish-language book market", in *Online Information Review*, 44, pp. 1387–1402.
Gallego Cuiñas, Ana (2021): "La cuestión de la literatura latinoamericana y española en el siglo XXI", in *Novísimas. Las narrativas latinoamericanas y españolas del siglo XXI*. Madrid: Iberoamericana, pp. 11–41.
Gallego Cuiñas, Ana (2022): *Cultura literaria y políticas de mercado. Editoriales, ferias y festivales*. Berlin: De Gruyter.
Goicoechea de Jorge, María (2019): "La literatura digital y los nuevos formatos de la edición literaria", in *Revista de Humanidades Digitales*, 4, pp. 162–186.
Goldstone, Andrew and Underwood, Ted (2014): "The Quiet Transformations of Literary Studies: What Thirteen Thousand Scholars Could Tell Us", in *New Literary History*, 45, pp. 359–384.
Halavais, Alexander (2015): "Bigger sociological imaginations: framing big social data theory and methods", in *Information, Communication & Society*, 5, pp. 583–594.
Haraway, Donna (1995): *Ciencia, cyborgs y mujeres. La reinvención de la naturaleza*. Madrid: Cátedra
Hayles, Katherin (2008): *Electronic Literature: New Horizons for the Literary*. Indiana: University of Notre Dame Press.
Helgueta Manso, Javier (2022): "Antiguos y modernos en la comunicación multimedial de la literatura. Citas, videoreseñas, tuiteros y booktubers", in *Insula*, 907–908, pp. 43–46.
Herrmann, Sebastian et al. (eds.) (2022): *Beyond Narrative. Exploring Narrative Liminality and Its Cultural Work*. Bielefeld: transcript.
Horkheimer, Max y Theodor W. Adorno (19989: *Dialéctica de la ilustración*. Madrid: Trotta.
Jean So, Richard and Long, Hoyt (2013): "Network Analysis and the Sociology of Modernism", in *Boundary*, 40, pp. 147–182.
Jockers, Matthew L. (2013): *Macroanalysis: Digital Methods and Literary History*. Chicago: University of Illinois Press.

Karsdorp, Folgert et al. (eds.) (2021): *Humanities Data Analysis. Case Studies with Python*. Princeton: Princeton University Press.
Kittler, Friedric (2010): *Optical Media*. Cambridge: Polity.
Kristeva, Julia (2002): *Estética I*. Madrid: Fundamentos.
Koskimaa, Raine (2005): "Close reading: hipertextos de ficción", in Laura Borrás (ed.). *Textualidades electrónicas. Nuevos escenarios para la literatura*. Barcelona: Editorial UOC, pp. 177–191.
Kozak, Claudia (2012): *Tecnopoéticas argentinas. Archivo blando de arte y tecnología*. Buenos Aires: Caja Negra.
Kozak, Claudia (2017): "Esos raros poemas nuevos. Teoría y crítica de la poesía digital latinoamericana", in *El jardín de los poetas. Revista de teoría y crítica de poesíaLatinoamericana*, 4, pp. 1–20.
Ledesma, Germán (2022): *El susurro de los mercados. Capitalismo financiero y literatura digital*. Rosario: UNR Editora.
Lemercier, Claire and Zalc, Claire (2019): *Quantitative Methods in the Humanities. An Introduction*. Virginia: University of Virginia Press.
Maltz, Hernán (2020): "Discusión sobre sociología de la literatura", in *Políticas de la Memoria*, 20, pp. 261–271.
Martínez-Gamboa, Ricardo (2016): "Big Data en humanidades digitales: de la escritura digital a la "lectura distante"", in *Revista Chilena de Literatura*, 94, pp. 39–58.
Merleau-Ponty, Maurice. Merleau-Ponty (1968): *The Visible and the Invisible*. Evanston: Northwestern University Press.
Moraña, Mabel (2014): *Bourdieu en la periferia. Capital simbólico y campo cultural en América Latina*. Santiago de Chile: Cuarto Propio.
Moretti, Franco (2016): *Lectura distante*. Buenos Aires: FCE.
Morgan, Nick (2013): "¿Olvidar el latinoamericanismo?: John Beverley y la política de los estudios culturales latinoamericanos", in *Cuadernos de Literatura*, 34, pp. 18–45.
Olaizola, Andrés (2018): "Bots sociales literarios y autoría. Un aporte desde la retórica digital", en *Virtuales*, 9, pp. 237–259.
Piglia, Ricardo (2005): *El ultimo lector*. Barcelona: Anagrama.
Piper, Andrew (2017): "Think Small: On Literary Modeling", in *PMLA*, 132, pp. 651–658.
Piper, Andrew (2018): *Ennumerations. Data and Literary Study*. Chicago: University of Chicago Press.
Piper, Andrew and Eva Portelance (2016): "How Cultural Capital Workds: Pizewinning Novels, Bestsellers, and the Time of Reading", in *Post45*, https://post45.org/2016/05/how-cultural-capital-works-prizewinning-novels-bestsellers-and-the-time-of-reading/.
Posada, Adolfo R. (2019): *Big Data*: el crítico frente al algoritmo (Lectura distante de la narrativa española editada en el año 2019), in *Colindancias*, 10, pp. 17–40.
Pozuelo Yvancos, José María (1994): *Teoría del lenguaje literario*. Madrid. Cátedra.
Rodríguez, Juan Carlos (2011): *Tras la muerte del aura (En contra y a favor de la Ilustración)*. Granada: Editorial de la Universidad de Granada.
Roig-Sanz, Diana and Reyne Meylaers (2019): *Literary Translation and Cultural Mediators in 'Peripheral' Cultures: Customs Officers or Smugglers? (New Comparisons in World Literature)*. New York: Palgrave MacMillan.
Rosseti, Miguel (2014): "A contraluz: World Literature y su lado salvaje", in *CHUY*, 1, pp. 60–93.
Richards, Ivor A. (2004): *Practical Criticism: A Study of Literary Judgment: A Study of Literary Judgement*. New York: Routledge.
Sádaba Rodríguez, Igor (2020): "El no tan nuevo espíritu del predictivismo: de la estadística moderna al *Big Data*", in *Revista Crítica Pneal y Poder*, 19, pp. 57–77.

Saum-Pascual, Alex (2022): “Materialidad digital, algoritmos y otras abstracciones modernas”, in Daniel Escandell (ed.). *Escrituras hispánicas desde el exocanon*. Madrid: Iberoamericana, pp. 23–36.

Schiuma, Giovanni and Carlucci Daniela (eds.) (2021): *Big Data in the Arts and Humanities*. London/ New York: CRC Press.

Sierra, Germán (2022): “Caos frio”, in *Insula*, 907–908, pp. 11–14.

Striphas, Ted (2011): *The Late Age of Print: Everyday book culture from consumerism to control*. New York: Columbia University Press.

Tinianov, Iuri (1972): *El problema de la lengua poética*. México: Siglo XXI.

Underwood, Ted (2019): *Distant Horizons. Digital Evidence and Literary Change*. Chicago: University of Chicago Press.

Vanoli Hernán (2019): *El amor por la literatura en tiempos de algoritmos. 11 hipótesis para discutir con escritores, editores, gestores y demás militantes*. Buenos Aires: Siglo XXI.

Vera Barros, Tomás (comp.) (2014): *Escrituras objeto. Antología de literatura experimental*. Buenos Aires: Interzona.

Villanueva, Darío (2022): “Muerte de la literatura, posliteratura, literatura digital”, in *Insula*, 907–908, pp. 2–7.

Azucena G. Blanco

Epistemology and Big Data: From Grand Narratives to Big Data

This study sets out to reflect on the change of episteme that the transformations in power-knowledge relations have brought about through the contemporary epistemological model of Big Data. I propose a work in process – it cannot be any other way due to the intense topicality of the phenomenon, which is constantly being developed and expanded. The main objective is to think about the power that is associated with what we can know, and the way in which we know, in the Big Data episteme. So, although its nature as a scientific tool can project – as has already happened with other positivist methodologies before – the idea that its scientific knowledge is universal and blessed with transhistorical unity, the truth is that the very object of knowledge of Big Data – let us not forget – is historical, given that it is translating a cultural language into data. In other words, it is cultural and humanist knowledge, and therefore radically historical.

In 1975, Foucault published *Discipline and Punish*, in which he presented the panopticon as a historical-epistemological model of disciplinary power. From that year until today, one could say that the models of disciplinary power have followed the line of an ever more methodical and global panopticism, in which Big Data is presented as a network of networks of power based on statistical data. This is how Byung-Chul Han conceives it in *The Transparency Society* (2012). Unlike Baudrillard in *Simulacra and Simulation* (1978), he considers that:

> at the moment, we are not experiencing the end of the panopticon, but rather the beginning of an entirely new, *aperspectival* panopticon. The digital panopticon of the twenty-first century is aperspectival insofar as it no longer conducts surveillance from a central point, with the omnipotence of the despotic gaze. The distinction between center and periphery, which is fundamental to the Benthamian panopticon, has disappeared entirely. The digital panopticon functions without any perspectival optics. That is what makes it efficient (45).

Note: Where texts cannot be found in English, the translations are mine.

Note: This publication is the result of the R&D project of the Ministry of Science and Innovation Procesos de subjetivación: biopolítica y política de la literatura. La herencia del primer Foucault ("Procesos de subjetivación: biopolítica y política de la literatura. The In a footnote, add: This publication is the result of the R&D project of the Ministry of Science and Innovation Procesos de subjetivación: biopolítica y política de la literatura. La herencia del primer Foucault ("Procesos de subjetivación: biopolítica y política de la literatura. The legacy of the first Foucault" PID2019-107240GB-I00).

https://doi.org/10.1515/9783110753523-004

As Han says, the absence of perspective makes it more efficient because it can be produced simultaneously, from all perspectives. In the disciplinary society, the imprisoned cannot communicate with each other nor see each other: "For the purpose of improvement, one reads in Bentham, they are exposed to isolation" (46). For Han, however, the digital panopticon is different from Bentham's because it is connected and communicates with itself: "Not lonesomeness through isolation, but hypercommunication guarantees transparency" (46).

Far from the exhaustion of this device of power that Baudrillard predicted at the end of 1970s, the panopticon – as per Bernard Harcourt in *Exposed Desire and Disobedience in the Digital Age* (2015) – has attained such a current relevance that it has become one of the most recurrent metaphors in the studies on this issue (along with others such as Orwell's *Big Brother*, or that of the Matrix). Today, it has even come to develop legal aspects:

> The metaphor has reached into the legislative and constitutional debate as well, especially in the controversy surrounding the constitutionality of the NSA's bulk telephony metadata program. In the related judicial opinions and commission reports, the new "surveillance state" is being variously described as a great protector and selfless warrior by one federal judge in New York, as Big Brother by another federal judge in Washington, D.C., and as a New Deal–like administrative savior by President Barack Obama's advisers (Harcourt 2015: position 958 of 7602).

The author adds a new metaphor to this pile, elaborating on Han's idea of the society of transparency: "our mirrored glass pavilion". This lies at the basis of what he proposes as the birth of the expository society in the digital era:

> Part crystal palace, part high-tech construction, partly aesthetic and partly efficient, these glass and steel constructs allow us to see ourselves and others through mirrored surfaces and virtual reflections. They are spaces in which we play and explore, take selfies and photograph others. At times they resemble a fun house; at other moments they make us anxious. They intrigue and amuse us. They haunt us. And they hide pockets of obscurity (Position 1890 of 7602).

I would like first to consider some questions that are related with these new developments of disciplinary power that enable a more effective panopticism, unchained from the traditionally disciplinary spaces such as the workplace, hospital, school, prison, and so on, until reaching what has traditionally been called the space of "private life". For **is the power related to dataistic society a disciplinary power, or does this epistemological model and its decentralization from the aforementioned spaces toward the space of privacy combine it with another type of power in our societies?**

Furthermore, I would like to question whether this epistemological transformation in 21st-century society – or what we define as a change of episteme "from

grand narratives (or master narratives) to big data" – also opens up other ways of resistance, "machines of resistance" (Deleuze/Foucault), using literary and artistic hermeneutic models. For the first question, I join Harcourt in replying that we are facing a new model of power, which Foucault defined as "the power of the shepherd", and which other authors have already called "digital pastoral power". For the second question, I will test out some of the proposals made by James C. Scott in *Domination and the Arts of Resistance: Hidden Transcripts* (1990), a work that does not address the questions of digital society but which, given that Scott starts out from generalizable anthropological hypotheses, should give us resources for our question, **can we resist (against) digital pastoral power?** Because, as Cigüela Sola states: "it is no longer only a question of a set of tools to analyse more data more quickly, but that it is producing highly significant effects in our personal, social and political life, in the way we inhabit our own body (think about the data-gathering techniques on our health) and our own cities (think *Smart Mobility*)" (Cigüela Sola, 2017: 36).[1]

1 Pastoral Power in the Digital Society: Power in Your Living Room

Continuing with Harcourt's metaphor from above, we could add that the power enters your living room and turns it into a fairground . . . Jokes aside, one of the key questions in the definition of the concept of disciplinary power is that, as I have said, it separated some individuals from others; it left them incommunicado. This was the model upon which Bentham imagined the panopticon: the subject who was to be disciplined could not communicate but is observed at all times. In the transparency society, in Han's terms, we are not incommunicado or isolated, or at least not in appearance; rather, the channels of communication have multiplied, diversified and branched out, and they are realized from inside the private space. And surveillance, likewise, acquires the complexity of a hall of mirrors, as Harcourt indicated.

1 Original: "ya no se trata solo de un conjunto de herramientas con las que analizamos más datos y más rápido, sino que está produciendo efectos muy significativos en nuestra vida personal, social y política, en el modo que habitamos nuestro propio cuerpo (piénsese en las técnicas de recogida de datos sobre nuestra salud) y nuestras propias ciudades (piénsese en la Smart Mobility)".

In *Security, Territory, Population* (1978),[2] Foucault introduces the concept of **pastoral power** – a type of previous government that in the modern state would be called *governmentality*[3] and that is at the basis of a "society of security".

Pastoral power has its origin in a religious type of power: "the pastoral relationship in its full and positive form is therefore essentially the relationship of God to men. It is a religious type of power that God exercises over his people" (170). This is a power that is no longer characterized by isolating the citizen-subject, but "is exercised over a multiplicity on the move" (Foucault 2006: 171), which lets itself by guided by the pastor. For this power is not exercised with violence but concerns itself with ensuring the life of all and every one of the members of the multiplicity. It is a *benevolent* power that watches over the "security" of everyone. We should remember at this point that our first steps toward the loss of our privacy came with the request for our data, validated by the investigations that took place after the 9/11 terrorist attacks. President Bush spoke of the need to look after innocent citizens who, being innocent, had nothing to hide. Only those who were hiding wrongdoing that affected all society had reason to fear, because they would be removed from the flock.

Similarly, pastoral power is founded on its mobile capacity and, therefore, on the need to guide this shift – that is, it is exercised by shift and conduction. In the words of Cigüela Sola: "it has to be a dynamic power that *moves* the flock to certain places and certain practices, at the same time as removing them from others" (2017: 47).[4] This is thus the type of power that the global world demands: techniques of pastoral control – power of *conduction*.

The question we now need to pose is, **how is this benevolent conduction possible in our society?** According to Foucault, pastoral power is related to a politics of truth: pastoral power is exercised in a daily way; it is a necessary guide.[5]

These types of power relations are so asymmetrical, argues Harcourt, that they do not even require coercion or discipline, and one cannot even talk precisely of *removal*: on the contrary, "we are giving it freely and voluntarily, with love, desire and passion" (Cigüela Sola 2017: 41).[6] It is a veridical participation, in

2 A study prior to his work on biopolitics, which he would develop in the following course of lectures in 1979/80, and has been published under the title, *The Birth of Biopolitics*.

3 Foucault defines governmentality as "the way in which the conduct (*conduite*) of a set of individuals became involved, in an increasingly pronounced way, in the exercise of sovereign power" (364).

4 Original: "ha de ser un poder dinámico que mueva al rebaño a determinados lugares y a determinadas prácticas a la vez que los saca de otros."

5 Note that the techniques of confession, in Christianity, are aimed precisely at a dependence on the confessor/guide.

6 Original: "la estamos proporcionando libre y voluntariamente, con amor, deseo y pasión".

that the observed subject constantly forgets that they are being watched. And, in part, we could consider that this comes from most people's incomprehension of the complexity of the algorithms. As per Harcourt, taking the Derridean concept, they are given a "mystical foundation of authority", because we carry on immersed in a hauntology.[7] The justification of the marketing of our data is also based on this same spectral ontology. **Joseph Vogl,** in *The Specter of Capital* (2010), argues that the statement "the market knows better" is a secular version of Adam Smith's faith in the market's "invisible hand", his economic interpretation of the providentialist theodicy of the 18th century, which later hardened into an *oikodicy*, an unquestionable belief in the beneficence of the self-regulation of market forces. Vogl shows that the financial theory, aided by mathematical modelling and digital technology, in essence works like a "hidden hand", pushing economic reality toward unknown territory.

We can therefore **conclude up to this point** that, if we consider daily communication through networks and in the constant checking of messaging apps, the lack of communication – a defining trait of disciplinary power previously – is not a characteristic of our situation. It is a fact that we communicate with each other (though what we communicate about is another matter, as is how we do it and whether this communication is free). The digital society provides us, therefore, with the platforms of communication, selling the idea of a "secure communication" space, as though it were a confessionary or a psychiatrist's couch, and it observes, registers and classifies all this information as "conduct knowledge of the citizens". Think of the apps that intend to care for our health: counting steps, measuring heartrate, and encouraging us to overcome our personal goals, almost like a mother who is watching over our individual achievements.

As Han argues in *Psychopolitics: Neoliberalism and New Technologies of Power*, its "friendliness is what makes surveillance so efficient", precisely because this is what legitimizes its invasive and constant nature (2017: 39). According to Han, the current regime of knowledge is a "microphysics of power", in which the political or economic power knows so much about the citizens that they are able to mould their "offer" to their subjectivity (62).

7 Derrida developed the concept of mystic authority in *The Specters of Marx* (1993) and in *The Force of Law: The "Mystical Foundation of Authority"* (1994).

2 Digital Affectivity, Privacy and Crisis of Exteriority: Power Asks for your Things, Pretty Please

We can state, without any worry of being wrong, that we have developed an ***affective*** dependence on these modes of communication. As Olga Subirós and José Luis de Vicente, curators of the exhibition "Big Bang Data" (2014), declared: "we are data". And, as Byung-Chul Han explains in *The Transparency Society*:

> the particularity of the digital panopticon is that its inhabitants actively collaborate in its construction and maintenance by putting themselves on display and baring themselves. They display themselves on the panoptic market. [. . .] The society of control achieves perfection when subjects bare themselves not through outer constraint but through self-generated need, that is, when the fear of having to abandon one's private and intimate sphere yields to the need to put oneself on display without shame. (46)

When I say affective dependence, I also mean it in the sense that Remedios Zafra (2017) has of enthusiasm, in her work of the same name: the enthusiasm with which we give up our data to friendly questioning. Zafra has tackled in depth how capitalism has instrumentalized the principles of will: creativity and enthusiasm, in contrast to other historical periods when art was considered "a socially non-productive activity". According to Zafra, capitalism has appropriated the machinery of enthusiasm – that is, "induced enthusiasm", fuelled by the logic of the market (31). And she goes on to write: "The reason for their incentive can be found in the fact that this induced enthusiasm has become a capitalist tool that makes it possible to keep up the pace of productivity, hide conflict beneath a mask of motivation capable of maintaining the demands of production at less cost"[8] (31). We could, ultimately, call this **enthusiastic shepherding**.

Indeed, it seems that we have turned a deaf ear to one of the warnings Foucault gave us about fascism in his introduction to *Anti-Oedipus* by Deleuze and Guattari: "**do not become enamoured of power**", because "it is the connection of desire to reality that possesses revolutionary force". We were also warned against this fascism – which Antonio Méndez Rubio calls "low-intensity fascism" – by Pier Paolo Pasolini in his *Lutheran Letters*:

8 Original: "La razón de su incentivo puede encontrarse en que este entusiasmo inducido se ha convertido en herramienta capitalista que permite mantener la velocidad productiva, esconder el conflicto bajo una máscara de motivación capaz de mantener las exigencias de la producción a menor coste".

> Consumerism can create "social relations" which are not subject to modification; in the worst case creating a new techno-Fascism in the place of the old clerico-fascism (which could probably come about only if it were to call itself anti-Fascism) or, as is now more probable, by creating a context for its own hedonistic ideology a context of false tolerance and of false laicism: that is to say, the false attainment of civil rights (Pasolini 1987: 124).

To these words we could add: a context of false confidence and of freedom to speak. Both Remedios Zafra and Carissa Véliz – who recently published her best-seller *Privacy is Power: Why and How You Should Take Back Control of Your Data* (2020) – both argue for an exit from electronic systems. Because, Véliz says, "any social system depends on the cooperation of the people. [. . .] If we stop cooperating with surveillance capitalism, we can change it" (78). The aforementioned exhibition, "Big Bang Data", gave its catalogue the title, *Anonimízate. Manual de autodefensa electrónica* [*Go Anonymous: Electronic Self-Defence Manual*] (2015), and it was presented in the style of a manual with techniques of resistance to digital surveillance.

However, the question we now ask is: **is it possible to leave the system, if our spectral ontological condition currently depends on our constant exposure on social networks, on teaching platforms, on online communication apps (*are we data)*?** If the opinions of citizen-subjects are configured through politicized messages that direct their opinions to where they meekly want to go, **is leaving the device a real exit from the system of the computerized world?** Street cameras, the mobile devices that can record us without our knowledge, at a concert, at a demonstration, or in our classrooms . . .

3 The Art of Resistence: Hiding in Plain Sight

In *Domination and the Arts of Resistance: Hidden Transcripts* (1990), James C. Scott distinguishes between "public discourse" and the "hidden discourse". Like Carissa Véliz, Remedios Zafra or the Groupe MARCUSE, among others, in James C. Scott we find a praising of exteriority: in his work on cultures, he observes how the suppressed, the dominated, *hide* to express themselves with freedom. "Finally, subordinates in such large-scale structures of domination nevertheless have a fairly extensive social existence outside the immediate control of the dominant. It is in such sequestered settings where, in principle, a shared critique of domination may develop" (xi). Thus, resistance, marginal dissidence, can be expressed hidden from power relations. As Scott says:

> Every subordinate group creates, out of its ordeal, a "hidden transcript" that represents a critique of power spoken behind the back of the dominant. The powerful, for their part,

> also develop a hidden transcript representing the practices and claims of their rule that cannot be openly avowed. A comparison of the hidden transcript of the weak with that of the powerful and of ***both*** hidden transcripts to the public transcript of power relations offers a substantially new way of understanding resistance to domination (xii).

In other words, resistance is often carried out in hushed tones, because there are certain things that cannot be spoken directly to power. These forms, according to Scott, take on the hue of literature. In the same way that literature was a form of critical writing, as per Bakhtin, in popular culture as well the modes of resistance are related to the carnivalesque principle: "We are saved from throwing up our hands in frustration by the fact that the hidden transcript is typically expressed openly – albeit in disguised form" (xii–xiii). It is "the infrapolitics of the powerless" (21). Therefore, Scott considers that, against the idea of conformism, of false class consciousness or of hegemony, in a society that does not appear to repress our freedom violently, the forms of resistance are, again, in a kind of codified language:

> When it comes to understanding why the Western working class has apparently made an accommodation with capitalism and unequal property relations despite its political rights to mobilize, one finds, again, both thick and thin accounts of ideological hegemony. The thick version emphasizes the operation of what have been called "ideological state apparatuses," such as schools, the church, the media, and even the institutions of parliamentary democracy, which, it is claimed, exercise a near monopoly over the symbolic means of production just as factory owners might monopolize the material means of production. Their ideological work secures the active consent of subordinate groups to the social arrangements that reproduce their subordination (Scott 2004: 100).

Scott's hypothesis is, therefore, that resistance is possible as a capacity of denial in the hidden discourse of a group against forced submission (experiences of slaves, castes, workers), because even when it is accepted, since the class difference is "crushing", this does not mean that a class conflict is not generated.

Two humanist and enlightenment hypotheses persist in Scott's argument, which he himself argues, via Sharon S. Brehm (*Psychological Reactance: A Theory of Freedom and Control*, 2013). First, that "there is a human desire for freedom and autonomy that, when threatened by the use of force, leads to a reaction of opposition" (109) – which corresponds to the principles of *liberté and egalité*. And second, "the essential point is that a resistant subculture or countermores among subordinates is necessarily a product of mutuality" (119) – which corresponds to the principle of *fraternité*.

Therefore, resistance occurs: (1) in plain sight; (2) creating an exteriority in the very interior of the system, with traits of the carnivalesque principle; and (3) guided by the principle of class solidarity. This resistance is what defines the characteristics of this hidden discourse:

1) the hidden discourse is a social product and therefore the result of the power relations between subordinates.
2) as popular culture, the hidden discourse does not exist as a form of pure thought; it exists only insofar as it is practised, articulated, manifested and disseminated within marginal social spaces.
3) the social spaces where the hidden discourse grows are, by themselves, a conquest of resistance, which is won and defended in the jaws of power (175).

The subordinated classes thus appropriate the **classic hermeneutic principle** and turn it on its head (**carnivalesque principle**): there is a hidden truth, which is not visible to power, but is the truth of the resistance. The resistance reuses places and discourses from which domination and subordination were traditionally exercised. Scott gives us several examples, one of which is of the slaves in the southern USA before the civil war. These slaves practised a form of Christianity as resistance: while the preachers, tied up with the interests of the masters, emphasised the New Testament passages on meekness, on offering the other cheek or making more effort, "the offstage Christianity, as we know, stressed the themes of deliverance and redemption, Moses and the Promised Land, the Egyptian captivity, and emancipation. The Land of Canaan, as Frederick Douglass noted, was taken to mean the North and freedom" (116). It was their way of showing their "disagreement" within the very discourse of power, using the hermeneutic principle of inversion, whereby they used Christian readings to empower the enslaved population.

4 From Grand Narratives to Big Data: The Ideological Construction of Truth or Verisimilitude

According to Foucault, power and knowledge, as historical production of the truth, are always intertwined. **The truth** is of this world and is produced here thanks to many impositions. That is, here it has regulated effects of power, by which every society has its regime of truth, its "general politics of truth": the types of discourse that they accept and make work as truths.

The ideological constructing of truth is something that the system has learned, and all the all the agencies of political campaigns of recent years (the examples of post-truth in Trump's campaigns, among others, were very much sounding in that direction, while the uses of quantification and analysis in the pandemic are all too familiar).

Post-truth is based on what we could call a "social verisimilitude", that is to say, detecting and working on those stories that are not true but are plausible for a majority of the population, and which Big Data helps to illustrate. According to Boyd and Crawford in "Critical Questions for Big Data" (2012), Big Data is a phenomenon that is not only concerned with the technological and analytical – its more obvious functions – but also with the mythical: in the way we construct the truth about ourselves and about our societies, in the way we associate certain views, both utopian and dystopian, with this phenomenon, and in the way in which we think of it as the solution to immemorial problems and also as the origin of new threats: "This is where, therefore, the *mythical* nature of this technology lies: Big Data currently forms part of the process of *mystification* typical of all authority that makes decisions self-referentially" (Cigüela Sola 2017: 40).[9]

Big Data, therefore, is a fundamental tool in the politics of truth in our time. And, like all tools (*tekné*), it can be used in different ways in the processes of veridiction (ways of telling and constructing the truth).

5 Notes for A Hermeneutics of Data: Modes of Reading as Modes of Resistance

As we have seen, the age of digital power is that of a power of digital shepherding, which mobilizes its population by analysing the traits and limits of their social verisimilitude. The arguments extolling exteriority seem difficult to accept, insofar as the system of digital surveillance extends its mechanisms of quantification beyond personal liberty, gathering data that we do not give up voluntarily. For this reason, it seems to me that James C. Scott's proposal of a countercultural resistance is more productive, as it allows every citizen-subject "to hide in plain sight". In his study, Scott shows us two modes of resistance: that of subjects who "appropriate" typical elements of power in order to invert them *carnivalesquely*, to twist their subordinated meaning and direct it toward a silent subversion that is woven into a network of solidary collaboration; and a methodological mode, which he puts forward with these modes of reading.

Taking these suppositions, I would like to conclude with the following three ideas, which are a working principle:

9 Original: "Es ahí, por tanto, donde radica el carácter mitológico de esta tecnología: el Big Data forma parte actualmente del proceso de mistificación propio de toda autoridad que decide autorreferencialmente".

1) The need for a hermeneutics of data that is in itself a subversion of quantifying reading (dataism).
2) The need for forms of narrating the world that we inhabit, with Big Data as a tool. I consider that Big Data does not only offer useful quantification data for a bolder and more effective capitalism, but also new strategies of reading (Little Big Data), forms of articulating new narratives, and new grammars (José Luis de Vicente and Olga Subirós*: Big Bang Data*, CCCB 2014).
3) **How to read data?** I began this article pointing out that the knowledge derived from Big Data is always historical and subject to interpretation. It would thus be necessary to introduce specialized corpora to respond to diachronic questions, conscious of the historicity of our readings. This would enable, for example, etymological reflections typical of phenomenological hermeneutics, though this hermeneutics would be based on rhizomatic analysis (Deleuze), intertextualities (Bakhtin) and grafting (Derrida): a historical hermeneutics of dissemination. Although the media patterns are finite, virtual readings are historically variable, and infinite in their epistemes (virtuality as potentiality).

I will conclude, therefore, that "knowing how to read is knowing how to resist". In this way, the readings that we propose should be a thorough examination of the present and, at the same time, an archaeology of the future.

Bibliography

Anderson, Chris (2008): "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete": *Wired* (23-6-2008). <https://www.wired.com/2008/06/pbtheory/> (10-11-2021).

Bantjes, Marian (2008): "The Petabyte Age: Because More Isn't Just More–More Is Different": *Wired* (23-6-2008). < https://www.wired.com/2008/06/pb-intro/> (10-11-2021).

Bauman, Zigmunty David Lyon (2012): *Liquid Surveillance: A Conversation*, Cambridge: Polity Press.

Bigo, Didier (2006): "Globalized (in)Security: the Field and the Ban-opticon. <http://www.people.fas.harvard.edu/~ces/conferences/muslims/Bigo.pdf> (10-11-2021).

Bollier, David (2010): *The Promise and Peril of Big Data*. Washington: Aspen Institute.

Boyd, Danah & Kate Crawford (2012): "Critical Questions for Big Data", in *Information*, *Communication & Society*, 15, pp. 662–679.

Braidotti, Rosi (2015): *Lo Posthumano*. Barcelona: Gedisa.

Bratton, Benjamin (2015): *The Stack. On Software and Sovereignty*. MIT Press, London: Cambridge.

Cigüela Sola, Javier (2016): "Poder y Subjetividad en la era digital (Reseña de Bernard E. Harcourt), *Expos ed. Desire and Disobedience in the Digital Age*", in *Isegoría*, 55, pp. 735–739.

Cigüela Sola, Javier (2017): "*Big Data*, saber-poder y pastoreo digital: sobre el fundamento mitológico de la autoridad", in *Foro interno. Anuario de Teoría Política*, 17, pp. 35–54.

Danah Boyd & Kate Crawford (2012): "Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon", in *Information. Communication & Society*, 15.

Deleuze, Gilles (1999): “Post-scriptum a las sociedades de control”, in *Conversaciones*. Valencia: Pre-textos, pp. 277–286.
Derrida, Jacques (1993): *Specters of Marx: The State of the Debt, the Work of Mourning and the New International* (ed. Trans. 2006). London: Routledge Classics.
Derrida, Jacques (1994): The Force of Law: The “Mystical Foundation of Authority”, in Gil Anidjar (ed.), *Acts of Religion*. London: Routledge (ed. Trans. 2001).
Han, Byung-Chul (2015): *The transparency society*. Stanford: Stanford University Press.
Han, Byung-Chul (2017): *Psychopolitics: Neoliberalism and New Technologies of Power*. Edinburgh: Verso Books.
Harari, Nuval Yoah (2016): *Homo Deus. A Brief History of Tomorrow*. London: Random House.
Harcourt, Bernard E. (2016): *Exposed. Desire and Disobedience in the Digital Age*. Cambridge: Harvard University Press.
Foucault, Michel (1980): *Power/Knowledge. Selected Interviews & Other Writings, 1972–1977*. New York: The Harvester Press/Vintage Books.
Foucault, Michel (1983): “Introduction to the Non-Fascist Life”, in Gilles Deleuze & Felix Guattari, *Anti-Œdipus. Capitalism and Schizophrenia*. Minneapolis, University of Minessota Press: pp. 11–4.
Foucault, Michel (2006): *Security, Territory, Population. Lectures at the Collège de France, 1977–78*. London: Palgrave Macmillan.
Foucault, Michel (2008): *La verdad y las formas jurídicas*. Barcelona: Gedisa.
Foucault, Michel (2009): *Vigilar y Castigar. Nacimiento de la prisión*. Madrid: Siglo XXI.
Foucault, Michel (2014): *Wrong-Doing, Truth-Telling. The function of avowal in Justice*. Chicago/London: University of Chicago Press.
Grupo Marcuse (2019): *La libertad en coma. Contra la informatización del mundo*. Madrid: Ediciones El Salmón.
Grupo surrealista de Madrid (2012): *Crisis de la exterioridad. Crítica del encierro industrial y elogio de las afueras*. Madrid: Enclave de libros.
Luhmann, Niklas (1995): *Das Recht der Gesellschaft*. Frankfurt: Suhrkamp.
Luhmann, Niklas (2010): *Organización y decisión*. México D.F.: Herder/Universidad Iberoamericana.
Manovich, Lev (2011): “Trending: The Promises and the Challenges of Big Social Data”. <http://manovich.net/content/04-projects/067-trending-the-promises-andthe-challenges-of-big-social-data/64-article-2011.pdf> (11-10-2021).
Manovich, Lev (2013): *El software toma el mando*. Barcelona: UOC Press.
McLuhan, Marshall (1996): *Comprender los medios de comunicación. Las extensiones del ser humano*. Barcelona: Paidós.
Méndez Rubio, Antonio (2020): *Fascismo de baja intensidad*. Santander: Textos (in)surgentes.
Pasolini, Pier Paolo (2010): *Cartas luteranas*. Madrid: Trotta.
Scott, James C. (2018): *Domination and the Arts of Resistance. Hidden Transcripts*. New Haven/London: Yale University Press.
Subirós, Olga y José Luis de Vicente (2014): *Big Bang Data*. <http://bigbangdata.cccb.org/es/tema-somos-datos/>
Véliz, Carissa (2021): *Privacidad es poder. Datos, vigilancia y libertad en la era digital*. Barcelona: Debate.
Vogl, Joseph (2010): *The Specter of Capital*. Stanford: Stanford University Press (ed. Trad. 2014).
Willke, Helmut (2007): *Smart Governance: Governing the Global Knowledge Society*. New York: Campus Verlag.
Zafra, Remedios (2017): *El entusiasmo. Precariedad y trabajo creativo en la era digital*. Barcelona: Anagrama.

José Antonio Pérez Tapias

Is 'An-other Humanism' Possible through the Folds of Big Data?

Big Data Through the Ambivalence of the Technology that Drives the Information Revolution

The information revolution has had such a thorough effect on our culture that we can fairly say that we live in a digital culture. It has brought us, with its successive revolutionary stages within itself, to the era of so-called *big data* and artificial intelligence. In this era, having the *mass data* that computing and telematics provide in the most varied of fields, new paths have opened up not only to learn more about our reality in its different dimensions, but also to affect it in such a way that we are seeing transformations of such magnitude and depth that they constantly give rise to an overpowering vertigo, even when these changes can be valued positively – which, however, is not always the case.

Undoubtedly, the capacity of technology is astonishing, and at the same time overwhelming: to handle millions and millions of data, in quantities that are easy to put into words but difficult to imagine, in order to extract from them, by means of algorithms that continuously expand the sphere of artificial intelligence, information capable of being converted into knowledge, whether for scientific progress, for greater financial gain, or for political manipulation. It is well known how having at one's disposal work that is well-oriented and effective with big data has a positive impact on the development of biotechnology. This has led to impressive results in genomics, for example, and spectacular applications, as has occurred in what is called biosurgery, hand in hand with nanotechnology. Nevertheless, even in fields such as these, we can see the ambivalence of the technologies that revolve around big data. For what is revealed with them can both help to deal with diseases that are difficult to treat, and provide data – and predictions – about health, including proclivity to certain pathologies, for millions of people. This can easily lead to medical practice, social guidelines or financial decisions that would be damaging for these people in light of the predictions made. We know how and why such harm is socially concentrated at one point: the increase in inequalities, whether by how the information accumulated in this way about individuals, social classes and groups susceptible to (even more) discrimination is handled, or by the actual difficulties of accessing these sources of infor-

https://doi.org/10.1515/9783110753523-005

mation, or by the way in which the data they handle are made available (Eubanks 2021).

It is nothing new that computer and telematic technologies show both positive and negative possibilities in terms of their application. This has occurred and continues to occur with all the techniques and technologies that humanity has brought into being. What is new here, indubitably, is the weight of that ambivalence in these new technologies; just as their positive effects can be great, so can their negative consequences be immense. In addition to this ambivalence is the way in which, from their beginnings, many of the data that are subjected to algorithmic processes to extract the required information are obtained. When these data concern what takes place in the public sphere, from how the financial markets work to what the dominant trends in the literary field are – something of particular interest for the digital humanities – the obtaining of data should not be marred by processes that violate the freedom of individuals and their discretion with regard to the privacy of their lives. However, when the data are obtained using the footprint that all of us leave when we use digital resources, when we browse the internet, when we write emails, when we look up websites, when we interact through social networks, et cetera, then that is a whole other matter. In such cases, even while we as individuals may be aware that through these practices we are promoting the sale – without any profit to ourselves – of information about our habits, our convictions, our most personal decisions, and our most intimate messages, it is clear that we are faced with a serious, unresolved problem regarding ethical and – where applicable – legal limits concerning the obtaining and use of these data, just to prevent abuses.

The debates that revolve around such a thorny issue have, moreover, become especially vital since it has been shown how such use and abuse of what is done with the big data obtained in this way brings with it pernicious effects, whether in economic dynamics or in the political life of our societies, as well as possibly entailing harmful consequences for individuals. It is a clear fact that the algorithmic treatment of mass data provides valuable information for economic activity, from which, moreover, the most powerful businesses profit, starting with the very technological companies that dominate this same field. From the point of view of the market, it turns out that the conditions for competition are seriously affected. Intense concentrations of economic power have been facilitated, with strong monopolistic tendencies, which at the very least end up in oligopolistic conditions that have an enormous effect on the dynamics of capitalism today. This capitalism, which has been characterized by a dynamic marked by the primacy of finance, is now also being reshaped as “surveillance capitalism” – because capital gains now gravitate towards a new merchandise: the data that,

individually and collectively, we offer up to the large companies that dominate the digital realm (Zuboff 2020).

To be concise, the appearance of new fronts of economic activity due to the use of a massive growth of digitally available information seems to favour the addition of new "entrepreneurs" to the business sphere; yet it also reinforces the expansion of capitalist logic to fields of activity that heretofore had remained untouched. If people's data, people's lives, become merchandise whose commercialization – irrespective of the people – provides high added value, activities that until recently belonged to the sphere of privacy in individual lives, such as travelling or owning one's own home, are now fully part of the dynamics of intensive economic exchange. What was known as the "sharing economy" ends up becoming pure capitalist economic exchange beneath a label that has cooperative connotations but which is actually one of subterfuge or concealment. The "uberization" of many activities confirms that capitalism is still omnivorous and voracious in the age of universalized digitalization.

From a political point of view, the handling of big data has introduced new ways of acting that have a large impact on the dynamic of our societies. If the available mass data makes it possible to have extremely precise knowledge about social trends, states of opinion, political preferences, and so on, and all this enables decision-making with a greater margin for political accuracy, then it is that same availability that gives rise to the distortion of politics. This leads to anti-politics, to serious interferences in actual electoral processes from the moment that certain messages are spread on the internet and the various social networks, which manipulate information and thus harm or benefit particular candidatures or parties. And ultimately it gives rise to the spread and diffusion of the perverse cognitive dynamic that we have come to call *post-truth*, which is devoted to sowing lies and to consolidating the deceit that is expressly produced for political profit – including, since big data makes it possible to know the inclinations and emotional states of citizens, the cynical creation of supposed "alternative truths".

Digital Humanities in the Age of Big Data

For good or for ill, the huge impact of everything that big data makes available to us on our economic, social, political and cultural realities – taking into account how it can affect our individual existences and the collective life we form a part of – makes it an inescapable factor that we cannot ignore. This is also the case for the humanities, or forms of knowledge relative to our human realities as such, which we can see deployed in various areas, with different epistemic fields dis-

tributed around them. With an always notable common denominator, these have an enriching diversity in terms of knowledge about ourselves and the practices that we observe through them, be they favoured or questioned. The humanities encompass a wide range of fields, from the types of knowledge about languages and literary traditions that they have given rise to – we can say this of the philologies and studies of languages and linguistics as forms of knowledge of communication in diverse societies and ages – to philosophy as critical and argumentative knowledge regarding our forms of knowledge, aesthetic values and normative principles, to the search for meaning, passing through the different types of history, which as knowledge of memory constantly bring the knowledge of humanity's pasts to the present. And all this shares the company of disciplines such as cultural anthropology and geography, forms of knowledge with humanist roots that describe the plurality of cultures and spaces that humans inhabit.

The *humanities*, through their plurality and in their current state, neither can nor should be separated from what big data, and the digital culture to which it belongs, entails. They cannot, because they themselves are affected by the technological developments of our era: computing and telematics, which a few decades ago we began to call "new technologies of communication and information", have had a bearing on the *humanities*, introducing profound changes in their ways of working, with new epistemologies, and the issues addressed, and engaging with new perspectives (Vinck, 2018). For example, the study of languages makes use of the possibilities offered by data on linguistic uses in communities of speakers that would have been unimaginable previously. Philosophy itself has to deal with new moral dilemmas, such as in bioethics, wherein these "new technologies" have modified scientific knowledge and medical practices. The treatment of texts, the digitalization of documents, and the information accumulated about them by virtue of it, having impacted the *humanities* in general, have notably changed the ways of working in the field of history, including archaeology, with digital procedures applied to the information obtained in fieldwork, or as has happened in art history, with new knowledge that has led to spectacular innovations in the areas of conservation and restoration of cultural assets.

If, by virtue of the aforementioned changes and the reassessments made in the humanities as a result, we can speak of the *digital humanities*, encompassing all the new epistemic developments that have taken place, not to mention the promising nature of many of their approaches, then no less noteworthy is the fact that the humanities must also deal with questions of digital culture that are unavoidable, both in the study of the facts and the processes that they fall within, and also from a normative point of view, whether epistemological or ethical. Big data – to give an example – can provide us with a huge amount of information on the habits and behaviours of millions of people, which would support studies on

the construction of identities and processes of subjectivation that are most certainly novel. But at the same time, as I noted earlier, big data assists highly refined marketing or facilitates the gross manipulation of opinions that distort politics, occasionally coming close to breaching – or even overstepping – individuals' right to privacy, and even breaking the most basic legal requirements regarding freedom of expression and information. No approach to the humanities can avoid such tendencies, which are most noticeable in relation to problems such as those considered by the perverse cognitive dynamic that we find underlying the label *post-truth*, with negative political consequences (Pérez Tapias, 2018: 163–180).

Keeping in mind normative criteria when considering what can be done with big data, not only ethical criteria but also epistemological criteria are relevant, even essential. Digitalization provides new resources, through greater information, to store more knowledge and strengthen diffusion via new communication routes. Furthermore, it opens the way to generating knowledge in another way, and this is what is boosted many times over thanks to the use of mass data, its algorithmic treatment and the application of artificial intelligence. Hence one cannot lose sight of a fundamental epistemological question that, though it has been dealt with at length, is still of the upmost importance. This concerns being aware that the mere accumulation of data, however massive it may be, does not produce knowledge by itself. Obviously, the handling of big data has to be well guided, from search and selection with precise criteria, to the unequivocal formulation of the problems that need investigating or of the hypotheses that need addressing. Put concisely, having a lot of data is no guarantee at all that inductive strategies will successfully lead to the knowledge we desire and the conclusions we seek. Without clear questions there can be no satisfactory answers.

Delving deeper, where ethical and epistemological questions intertwine, we have what for the humanities is never unwelcome – quite the contrary, it is what we refer to when we talk of the *question of meaning*. The humanities, given that their *objects* of study concern humans as *subjects*, must always meet the need to move constantly between the interrelation of *explain* and *understand*, emphasized since the epistemological contributions of hermeneutics formulated in contemporary philosophy from Dilthey to Gadamer, Ricoeur and Apel. If sound *explanations* increase our knowledge of human realities with new *meanings* thanks to their articulation in well-founded theories, and also by being a component of empirical comparison, as is widely present in the social sciences, the humanities cannot give up trying to *understand* what such realities encompass, including what is relative to the *meaning* with which at their core humans live their existence.

Therefore, the *digital humanities*, which though *digital* must still be *humanities*, should not – and this is an epistemic task with an ethical dimension – lose sight of the *question of meaning* (Pérez Tapias, 2003). Moreover, they must address this question with reflexive contemplation, in terms of the most genuine meaning of the expression, as well as how to consider everything related to *meaning* in digital culture. And, more specifically, they must think about what it means to be human in the digital medium, when mediated digitally, and how such *mediation* comes about, critically addressing when and in what ways it becomes *mediatization*, through the *big data* with which we operate in our world.

The *Meaning* of What is Human and the Humanist Tradition

The humanities are committed to addressing the meaning of what is human. To this we can add the consideration that, in the existence of all humans, what is key is how we manage, individually and collectively, to travel along the paths that go from the hominization we stem from to the humanization we must cast ourselves toward. Moreover, it is in what we recognize as the humanist tradition that the developments focused on it have hastened into – at least in the vectors that we consider shapers of the humanist tradition identifiable as western, however much it may harbour universalist pretensions. It should be stated, therefore, that those humanities in which that tradition reaps its harvest cannot be disassociated from the humanism that has been forged in them through the various contributions that have enriched it. Hence if we speak of *digital humanities*, we are obliged to consider which humanism it is that they maintain or promote. Furthermore, if we were to conclude that they are contrary to continuing to weaving the thread of an unrenounceable humanism – clearly needful of radical reconsideration – then we would be at the point at which it would no longer do to talk of *humanities*, however much we wished, by making them *digital*, to save an epistemic space for the forms of knowledge that have constituted them.

At this point in time we cannot allow ourselves any naïveté when speaking of *humanism*. Although the roots of its intended meaning are found in Graeco-Latin thought, one should not disregard the humanist components of other traditions, such as Enrique Dussel with regard to the Semitic world and, more specifically, the Hebrew tradition, or as Erich Fromm has shown of the presence of humanist components in different cultures. Yet though we may underscore that statement by Protagoras, long established as a mandatory humanist reference, in that “man is the measure of all things”, and highlight the contributions of major figures

such as Cicero or Seneca, expanding the conception of the human to some more effectively universalist terms – the *humanitas* that every individual intensively bears, widely recognized by all members of humanity – we are not exempt from critically confronting what underpins humanist discourse, even by those who in the Renaissance eagerly took up that thread, such as Petrarch or Pico della Mirandola. Indeed, such a requirement for critical reception is accentuated with regard to how humanism has become reformulated in modern philosophy.

After what was described as the *anthropological turn* of the Renaissance, the protomodernity that began to excel in the culture of the Baroque – which in the thought of the Spanish Baroque found expression in the work of authors such as Francisco Suárez and Baltasar Gracián – was able to consolidate its humanism in a new anthropological conception, certainly, and in those ideas of *ius naturale* that used it to support a whole legal architecture around human dignity (Bloch, 2011). Such an ethical-political core would come to be a common element in all the humanist conceptions that followed, no matter that many of their constructions came to be the object of criticism due to their ethnocentric biases or ideological functions that were precisely contrary to the demands of that postulated dignity.

Modernity, which on the plane of thought gathered strength with the metaphysics of the subject that began with Descartes, added the value of autonomy to that demand for dignity, which, stated first as belonging to consciousness in the exercising of rationality, began to forge ahead as moral autonomy – Kant being the culmination in this aspect – with the consequent requirements transferred to the political field as claims for rights that should accompany the formation of the condition of citizenry in what would in time be nascent democracies. While not diminishing the criticism Heidegger formulated of a humanism in debt to an onto-theo-logical view that it had not shed, trapped moreover in humanism's drift toward the nihilism that he himself wished to eliminate, we should not neglect the atheist humanism of Feuerbach, in the interest of saving human dignity by rescuing it from its bondage to religious alienation. Neither should we forget its legacy in a Marx that, on the same wave, maintained the humanist vector, repositioning it in his historical materialism.

The crisis of that humanism arrived, in anticipation of the crisis of modernity itself, after that boom of its versions incubated in the heat of existentialist currents, with Sartre and Camus at their head. Emerging from the same Marxist camp was a strong critique of what was presented as "socialist humanism", for considering it an ideological creation according to the concept of ideology originating with Marx: the structuralist thought of Althusser erected an antihumanist bastion – against even the humanism that could be found in Marx's earlier writings, since the later works were framed in a "scientific" paradigm that was alien

to the humanist corruption via Feuerbach, along with the legacy of Hegel. The rejection of humanism gained ground with Foucault in "the death of Man", a formula that echoed Nietzsche's "death of God" and with which there was a radical questioning of a conception of man that, upon the pedestal of modern subjectivity, elevated the human being to an unsustainable deified condition, just as had been advocated by a philosophy that was both anthropocentric and idealist, with the social sciences themselves being affected by this concept since their outset, including versions of them in the Marxist field. Such Foucauldian anti-humanism prepared the way for post-humanism, in which many philosophical positions have grounded themselves, and in the sphere of the *humanities* themselves, since the crisis of modernity began to evolve into postmodernity. The questioning of the *subject*, the critique of a strong concept of reason, the objection to a view of history according to a mythicized progress and a way of thinking often unfolding in the shadow of Nietzsche, given the context of a culture permeated by nihilism, frame the criticism of a humanism for which the few proposals proffered for its recovery appeared unviable.

When, in the crisis of modernity, the criticism of humanism intensified, the questioning of it due to its connection to the metaphysics of subjectivity was added to entrenchment in the vector that was a response to a critical radicalization of ideologies: anti-humanism came to highlight how the humanist discourse has fulfilled certain functions of covering up and justifying a social order with a great deal of dehumanization. The general exposition of a conception of humanism linked to an idea of "human nature" that gave favourable scope to conservative political and religious approaches made humanism lose the emancipating potential that it had had when it was a bastion for the defence of human dignity. Humanism as an ideology became vulnerable to the most conservative interests present in society. Furthermore, since the last decades of the twentieth century, culturalist awareness has increased and feminist sensibility has strengthened, and so criticism has intensified, accusing western humanism of ethnocentrism and patriarchalism. Thus two fronts have opened up through which humanism is undermined, ending up as a mainstay of a false universalism, and simultaneously of a machismo underpinned by an androcentric view of the human.

With this kind of questioning of humanism, the *humanities* have been constantly impacted by the criticisms that have been heaped upon them. Though these criticisms are still relevant for the *digital humanities*, the latter finds itself open to another front of criticism: the accusation that humanism is succumbing to technocratism – or, phrased another way, to *technological fetishism* – as a consequence of a development of computing and telematics that is at the mercy of an instrumental reason that lacks purpose. The idolatrization of technology produced in such a case is what can give rise to the production of new applications

of the pragmatist maxim that "one can do, or even should do, everything that it is technically possible to do," without further consideration about aims or a supposed morally legitimate use of means in the handling of big data, for example. If this is so, the *meaning* of the human becomes strangled between the algorithmic *folds* that mass data are being hurled at to produce calculations in virtue of which rules are established to be followed by humans or regarding them in some way.

Is it Possible for Humanism to Recover, and also Recover the *Meaning* of the *Humanities* Themselves as *Digital Humanities* of a Neo-Baroque Age?

The question that makes up the above heading contains a supposition, which could well be considered a cryptically communicated enthymeme. It is this: if we cease to sustain an approach that is somehow recognizable as humanist, it no longer makes sense to talk of *humanities*. I personally think that for different reasons we still need to use the denomination "humanities" for the types of knowledge that I briefly alluded to earlier, which we also, incidentally, call *Arts* ("*Letras*" – literally "letters" – in Spanish). And to this I would add that the use of the same word "humanities" becomes somewhat inconsistent if it is not accompanied by humanist thought – a humanist thought that must be reimagined in order to survive once the criticisms made against previous versions of humanism, which can no longer be defended due to the contradictions they contain or the epistemic shortcomings they have accumulated, have been confronted and overcome.

The defence of the *humanities* and the proposal of a humanism that is sustainable with good arguments, which is the heir to a tradition, but which at the same time involves an *excess* in what is transmitted that exceeds what is captured by ideological mechanisms, must be carried out in the context of the digital culture we are immersed in. This can be seen from another perspective as the culture of a Neo-Baroque age. This age shows many *symptoms* of the Neo-Baroque, which should come as no surprise since the Baroque was the cultural movement of a previous era of crisis – which marked the beginning of modernity – and that if we now speak of neo-baroque it is precisely in the midst of the crisis, after a few centuries, of that very modernity (Pérez Tapias, 2019: 297–312).

Baroque culture catalysed the crisis, between the end of the sixteenth century and the start of the eighteenth century, that was produced by the collision and

resulting vacuum between the old culture of Christianity and the new culture that took off in modernity, with the Renaissance transition in between. We can add to that collision of Baroque Europe – with particular prominence at first of the Spanish Baroque or, more widely, the Iberian Baroque – with the clash between the European world of the conquistadors and the world of the indigenous cultures in America that were invaded and subjugated by them. The current crisis of the end of the twentieth and start of this century, meanwhile, is a crisis in which we clearly see the emptiness of questioned social and political institutions and secular ideologies of the modernity that is already breaking apart, overrun by the economic processes, socio-political events and cultural phenomena of our societies. It is worth pointing out some vectors in which all this takes place: the computing revolution; economic globalization; states being exceeded by the market (crisis of democracies under the neoliberal paradigm); the digitalization of culture; the correlation between identitarianism and nihilism; and the environmental crisis that has gathered around what we call climate change. On top of that, there is the COVID-19 pandemic that since the beginning of 2020 has ravaged humanity around the whole planet, affecting ways of life, the economy, social life, political dynamics and the way we understand ourselves through a heightened consciousness of vulnerability.

In the midst of these circumstances, new practices and new ways of thinking have emerged and are being developed, which we can aptly call neo-baroque. At the same time that we are seeking answers to the crises we are going through, from ecological and economic answers to healthcare, we continue trying to explain the realities surrounding us, and ourselves in them, reconstructing resources to address, however *fragmentarily* – which is so baroque! – the nihilism that invades us. This is the gravest cultural problem, with excrescences of cynical behaviours everywhere, analogous to how in the seventeenth century our predecessors of the beginning of modernity dealt with the scepticism that then became ubiquitous.

Whether with efforts still based on theological survival, or with creations that were exclusively based on independent reason, the thinkers of the Baroque Seicento attempted to come up with solutions to their crisis. One way or another, in this new view of the world, they had philosophical-anthropological developments of a humanist nature at their disposal (though these had differing degrees of coherence, particularly regarding their compatibility with universalizable requirements of respect for human dignity, for example for women, people considered heretical, or Indians and blacks, who were subjected to exploitation or slavery). Such contributions are of great value for comparing similarities and differences between their baroque and our neo-baroque age, between their search for answers in a world in crisis and ours in a world no less beset by new crises.

Between the *Folds* of Leibniz as Baroque Philosopher and the *Folds* of Big Data in our Neo-Baroque

When searching for comparable references from the Baroque of modernity that help us to consider ourselves in our Neo-Baroque crisis of modernity, it is Leibniz who, from the end of the seventeenth century, gives us a body of work that is particularly suited to the task. For Leibniz, moreover, there is the additional circumstance of his having been an exceptional mathematician, the creator of infinitesimal calculus and inventor of a suitable notation for it (invented simultaneously but wholly distinct and apart from the similar intellectual feat by Newton, as is well known). Between his infinitesimal calculus and his metaphysical thought, in which ontology and theodicy are combined, there is an interesting correlation: an ontology in which force displaces extension when thinking of reality, and in which matter, insofar as force, is assimilated to spirit[1] – an ontology that has a notable structural correlation with his mathematical achievements in terms of conceptual development.

Giving thought to the finite-infinite relationship in a construction of explanations of reality capable of opening up pathways in the search for meaning, Leibniz offers a solution in a great metaphysical construction. This has two parts. First, the ontology describes a reality made up of monads, some separated from others, but each one with their own perspective on the world, and in such a way that in turn each monad is a constructive result of other monads, according to a *principle of compossibility*, by virtue of which the real world is formed and, thanks to each and every one of the monads, is continually updated. Second, it leads to a theodicy – a justification of God (in view of the glaring problem of evil in the world) – that aims to demonstrate that this world is, thanks to that God, the effective realization of the best of possibilities that can be conceived. Such is the *sufficient reason* – the *principle of sufficient reason*, indubitably the "unifying element of the Leibnizian system" (Saame, 1988: 125), which obtains both for *truths*

1 Thus leaving behind Cartesian dualism, as Leibniz had already emphatically underlined in his *Discourse on Metaphysics*, which preceded the great works of his philosophical thought (it was written in 1686 but not published until 1846), notably in paragraph 18 on the importance of force as opposed to extension, and the paragraph, following those in which he outlined his concept of individual substance, in which he called it monad, differentiating it from Descartes' concept of substance (Leibniz 1983: 72 ff.).

of reason and for *truths of fact* or *contingent truths*[2] – which makes it possible to give an account of reality and its meaning, in close relation to the *compossibility* thanks to which the combinatorics of monads is considered the basis for justifying this world as it is, in which its *meaning*, since the presence of evil becomes neutralized, as a lesser evil, as not contradictory – *principle of non-contradiction* – is at the core of the reality that is given as the best possible.

The sufficient reason that the compossibles provide is, therefore, the keystone to regaining a questioned meaning, when not lost, in the midst of the infinite folds of reality – "pleats [*replis*] of matter", "folds in the soul", signs of identity likewise of baroque thought (Deleuze, 2009: 11 ff.) – which is coiled in each monad and which finds its unfolding in the forming of an order in which the positions of every monad in the flow of the series they are found in leave space for the human being to find their place and live their freedom through conditions that are more and more enlightened by a reason that sheds light on the need that emerges from those conditions. The ontology that speaks of a reality constituted by monads, and monads of monads as dynamic substantial entities, opens up to a humanist view in which humans find their place in the constant flow between the *folds of folds* of a reality of unending complexity that, nevertheless, is in accordance with the "harmony pre-established" by a God who can only want the best. This God's existence is (supposedly) proved out of what is in effect truly best, in a reformulation of Anselm's ontological argument. Through the same divine freedom, to the rhythm of the principle of reason, human freedom is saved in that history of the world in which the possible – including the maximum good – and the real – where the actual minimum of evil counts – are joined thanks to the intelligence and will of God, thus writing, as is stated in the *Theodicy*, that "novel of human life" that is effectively universal history (Givone, 2006: 308–309).

Deleuze's reading of Leibniz's thought as a thinking of the *fold* has its correlate in the force and presence of the fold and the measureless *fold of folds* not only in thought but also in Baroque painting and sculpture, in architecture and even in music (Chambers, 2006: 101–130). Reality, and the human being at its heart, is a monadological kaleidoscope, in which each part (monad) reflects the whole, although the whole is not perceivable from and by any part. This can only be done by the God that is indicated by a thought that tries to save reality in its immanence based on a *hubris* of metaphysical principles through which the transcendence of that same absent God is retained.

---

2 As Leibniz states in his *Monadology* (par. 32–38) and in the corresponding paragraphs of his *Theodicy* (for example, par. 44, 280–282 and 340–344).

What our reality suggests is to associate the Leibnizian *folds of folds* with the folds of an Artificial Intelligence, which could be considered a reiteration of infinitesimal calculus. Now it is a question of *statistical folds* in which the *unfolding/deployment of big data* takes place to *coil/refold* them on the individualist condition of human beings with a virtual perspective on a world whose reality nobody encompasses. Because it is only for the "great algorithm" which, transcending the materiality of data, tells us that this world is the one that exists without any alternative – the "there is no alternative" attributed to Margaret Thatcher. This is the discredited discourse in which the digital variant of the preestablished harmony is reformulated, with the conclusion that this is the best of all possible worlds because there is no possibility of another – the digital successes do not annul the neoliberal paradigm and its cognitive (that is, ideological) effects. It is thus as a *hubris* of data – the *dataism* Yuval H. Harari discerned as counterpoint to the atheism that the atheism of modern humanism culminated in, now promoted like a new religion, so necessary for the transhumanist faith (Harari 2017: 400 ff.) – that it carries with it the danger of big data as a threat to all humanist ambition, including the commitment to the dignity of each and every human being.

The Proposal for "*An-Other Humanism*", Also Through *Big Data*, Opposed to *Dataism*

Being able to establish parallels in this way between the baroque folds of Leibniz and the folds, with their unfolding and refolding, of *big data*, the limit of these parallels becomes apparent as soon as one observes that the nihilism of our technological civilization is not capable of harmonious development in which *meaning* can shine, as Leibniz still intended, albeit with his theodicy, for his humanism, running through all the complexity of his ontology. Today not only do we know that theodicy is impossible, but that we prove daily that the "unbearable lightness of being" – as per Kundera – in absence of theodicy, provides scandalous scope for cynicism that appears in the various spheres of our lives (Pérez Tapias 2016: 410–417).

Is it possible to save *sense* without God, through a maze of algorithms in which there is no Ariadne's thread? Modern humanism, when all is said and done, attempted it, but the very criticisms of humanism showed its failure. The truth of these anti-humanisms concerns the shortcomings that they revealed of prior forms of humanism. And the humanism that can be found in Leibniz not only is not free of this diagnosis, but also contributed in part to bringing about such criticism. In his book, *The Era of the Individual*, Alain Renaut points this out, showing how Leibniz's thought, as well as that committed reference to transcen-

dence, sees its humanist *pathos* due to the extreme individualism of his monadology (Renaut 1993: 60 ff. and 131–175). Hence, a non-individualist humanism is necessary – although there is no reason why this should not aim to be metaphysical, a point that Renaut himself comes to recognize, without it necessarily having to accompany the rejection of individualism. Is this possible? As long as the aim is to address the *question of sense*, metaphysics appears. Therefore, the answer we are looking for would have to be provided by a humanism that entails an alternative metaphysics with respect to previous iterations – ancient, premodern and even modern; in other words, *another sense paradigm*. It is to such a need that Levinas's metaphysics of alterity responds, and it is because of this that the French-Jewish philosopher can speak of a "humanism of the Other man", for which recognition of the alterity of the *other human* through the constituent responsibility of moral conscience, in which freedom is justified, is key (Lévinas 1993).

Taking Levinasian humanism as a starting point, rejecting individualism, one can move on to reconsider the autonomy of the subject that has been inseparable from modern humanism, beginning with that paradoxical heteronomy that Levinas highlights as the seat of autonomy itself for those who must earn it in response to another's – or others' – interpolation in the interrelation of co-subjects in which demands of justice manifest themselves. It is true, however, that when that autonomy matures and is exercised as responsibility, in the face of others and against otherness – including nature as otherness that calls us to responsibility – the matter of anthropocentrism that humanism had historically borne with it returns under a new light. This must be transmuted from *anthropocentrism of control* to *anthropocentrism of responsibility*, which is a touchstone for combining the same relationship of humans with animals without having to sacrifice necessary humanism to a supposedly possible animalism.

There is still some way to go in what could be considered a rehabilitation of humanism – analogous to what might be done with the very concept of "human nature" – in order to be able to talk of "an-other humanism": and we should make clear that this rehabilitation cannot be limited to creating one more variant among the known forms of humanism, based on fiddling with the details. What we need is precisely a reformulation of humanism so that it is not ideological cover for capitalism, neocolonial practices, patriarchalism, hidden forms of racism or cultural supremacism, and so on. Such an "other humanism" could follow the lead of decolonial thought and of the epistemologies of the South advocated by Boaventura de Sousa Santos (2019), when they speak of "an-other thinking" or of thinking through "an-other paradigm" – a common thread of the contributions collected in *El giro decolonial* [*The Decolonial Turn*] (Castro-Gómez and Grosfoguel, 2007). I should qualify that it is not a question of sweeping away the entire

humanist tradition, rather of tidying up the excess once the relevant criticisms have been made.

It is time to pay attention to the persistence of the legacy and need for humanism, even by those who have gone through anti-humanism and, furthermore, not remaining merely negative in respect to it, have positioned themselves within the parameters of a post-humanist thought. A particularly significant case is that of the Italian philosopher Rosi Braidotti, who on the one hand insists on the rejection of humanism, for the reasons already given, but on the other hand recognizes that there is a kind of humanist urge that we cannot free ourselves from – that we do not want to be free of, such as when we again take up the question of the subject, in a manner that recalls the later Foucault with the processes of subjectivation, after his watertight critiques of the *subject* (Braidotti 2020: 59 ff.). Leaving to one side Braidotti's untenable excess, as performatively self-contradictory, when she not only speaks of posthumanism but also of the "posthuman" (2015), in order therein to set up and place thought itself in that supposed position, the case she represents serves as a contrasting reference to support the proposal of the "other humanism", which we refer to in the Aristotelian way as the "*humanism we seek*".

When through the current world and culture we advocate "an-other humanism" – accompanied by a dialogic universalism, not imperialist, sexist or racist . . ., but quite the contrary – it must be done without demonizing the technological resources that computing and telematics have placed in our hands, and, at the same time, resisting the fetishism with which they are often treated. The aforementioned *dataism* is a result of this: it is this cult that incentivizes the excesses of the datum, both fanciful and humanly detrimental, which we see in the sphere of transhumanism. Critical assessments of this phenomenon, such as Luc Ferry's *La Révolution Transhumaniste*, are much needed. When transhumanism reaches an *inhuman* concept of the human being, which, as well as being destructive for the individual who accepts it against the evidence of their own finitude, even aiming for immortality, is radically anti-egalitarian in how it understands the relations between humans and supposed transhumans (Pérez Tapias 2020), we find even more reasons to take the side of Jacques Rancière from the moment that he also turns his gaze upon humanism for always having contemplated, in the best versions of itself, the equality of all humans – that ontological equality that moral exigencies must be based on in terms of equality of treatment and the political objectives of social and gender equality (Bodas 2012: 185–204). The *question of sense*, as a metaphysical matter that demands ethics, makes it necessary for the rehabilitation of humanism as "another humanism" to use an ontological approach regarding equality so that the *humanization* that all human beings have the right – and duty – to access is not tangled up with all kinds of conditions that make it impossible. We must therefore cultivate "an-other humanism" that, emerging from the crisis of modernity, points

to the *transmodernity* that Dussel and others consider when in theories and practices they set forth toward new inter-individual and inter-cultural relationships through an "other paradigm" (Dussel 2005: 257–294). Such "an-other humanism", being necessary, is the humanism that is proving possible in a digital culture in which the *humanities*, without succumbing to the tyranny of the algorithm, are still capable of putting all their epistemic might to the service of the dignity of each and every human being.

Bibliography

Bloch, Ernst (2011): *Derecho natural y dignidad humana* [1961]. Dykinson: Madrid.

Bodas, Lucía (2012): "Humanismo, humanidad y humanitarismo. Los avatares de la idea de los derechos humanos en Jacques Rancière", in Cereceda, Miguel y Menegazzi, Tommaso (eds.), *Humanismo/Animalismo*. Madrid: Arena Libros, pp. 185–204.

Braidotti, Rosi (2015). *Lo posthumano* [2013]. Barcelona: Gedisa.

Braidotti, Rosi (2020): *El conocimiento posthumano* [2019]. Barcelona: Gedisa.

Castro-Gómez, Santiago y Grosfoguel, Ramón (eds.) (2007): *El giro decolonial. Reflexiones para una diversidad epistémica más allá del capitalismo global*. Bogotá: El Siglo del Hombre.

Chambers, Iain (2006): *La cultura después del humanismo. Historia, cultura, subjetividad* [2001]. Valencia: Cátedra.

Deleuze, Gilles (2009): *El pliegue. Leibniz y el Barroco* [1988]. Barcelona: Gedisa.

Dussel, Enrique (2015): *Filosofías del Sur. Descolonización y transmodernidad*. Madrid: Akal.

Eubanks, Virginia (2021): *La automatización de la desigualdad. Herramientas de tecnología avanzada para supervisar y castigar a los pobres* [2019]. Madrid: Capitán Swing.

Ferry, Luc (2017). *La revolución transhumanista. Cómo la tecnomedicina y la uberización del mundo van a transformar nuestras vidas* [2016]. Madrid: Alianza.

Givone, Sergio (2006): "El bibliotecario de Leibniz", in VV.AA. (2006), *El yo fracturado. Don Quijote y las figuras del barroco*. Madrid: Pensamiento, pp. 301–318.

Harari, Yuval H. (2017): *Homo Deus. Breve historia del mañana* [2015]. Barcelona: Debate.

Leibniz, Georg W. (1983): *Monadología, Discurso de metafísica, Profesión de fe del filósofo*. Barcelona: Orbis.

Lévinas, Emmanuel (1993): *Humanismo del Otro hombre* [1972]. Madrid: Caparrós.

Pérez Tapias, José A. (2003): *Internautas y náufragos. La búsqueda del sentido en la cultura digital*. Madrid: Trotta.

Pérez Tapias, José A. (2016): *La insoportable contradicción de una democracia cínica*. Universidad de Granada: Granada.

Pérez Tapias, José A. (2018): *Universidad y humana dignidad. Verdades de las Letras frente al mercado de la posverdad*. Universidad de Granada: Granada.

Pérez Tapias, José A. (2019): *Ser humano. Cuestión de dignidad en todas las culturas*. Madrid: Trotta.

Pérez Tapias, José A. (2020): "De la muerte del Hombre" al transhumanismo. La *parrhesía* foucaultiana ante la pretensión de *Homo Deus*", in *Pensamiento*, 290, pp. 657–677.

Renaut, Alain (1993): *La era del individuo. Contribución a una historia de la subjetividad* [1991]. Barcelona: Destino.

Saame, Otto (1988): *El principio de razón en Leibniz* [1987]. Barcelona: Laia.
Santos, Boaventura de Sousa (2019): *El fin del imperio cognitivo. La afirmación de las epistemologías del Sur* [2018]. Madrid: Trotta.
Vinck, Dominique (2018). *Humanidades digitales. La cultura frente a las nuevas tecnologías*. Barcelona: Gedisa.
Zuboff, Shoshana (2020): *La era del capitalismo de vigilancia* [2019]. Barcelona: Paidós.

2 Methodological Issues

Wenceslao Arroyo-Machado and Nicolas Robinson-Garcia

Towards a Science of Humanities: How Big Data can Solve the Limitations of Scientometrics

1 Introduction

Fields in the humanities have historically been neglected and mistreated in research evaluation systems (Nederhof 2006). The indicators used in these assessments were conceived and designed based on communication practices from the natural and life sciences (Merton 1973; Price 1963). Since its conception, scientometrics, – a quantitative field devoted to the study of science as an informational process (Nalimov & Mulchenko 1971) –, has encountered great difficulties on studying communicational processes in the humanities (Garfield 1996). These limitations are normally attributed to differing publication and communication patterns (Hicks 2005; Hicks 1999) and, paraphrasing Manovich, a 'surface data' approach in which bibliometric methods were simply applied without any type of verification or 'translation' to the fields' practices (Manovich 2012).

But the computational advancements taken place in the last few decades have transformed the processes by which new knowledge is created, shared and discussed within and beyond academia (Wouters, Zahedi & Costas 2019; Peng 2011). Big Data techniques have introduced greater capabilities on the tracking and monitoring of the scholarly activities, revolutionizing the field of scientometrics, which has expanded its toolbox beyond the development of indicators based on journal publications and citations.

The launch of the search engine Google Scholar or the scientific database Scopus in 2004, ended with a long-standing monopoly held by the database Web of Science, the main data source used for quantitative studies on scientific communication. Since then, the array of sources to study science in general, and the humanities specifically, has greatly expanded. From using library holdings (Torres-Salinas & Moed 2009; Linmans 2010) or loan statistics (Cabezas-Clavijo et al. 2013), to introducing books and monographs in citation indices (Torres-Salinas et al. 2013) or the use of social media metrics, known as altmetrics (Hammarfelt 2014; Kousha & Thelwall 2009). Still, despite the excess of data, there is a lack of consensus and evidence on how these approaches can be of use (if they should be used at all) to better understanding and assessing scholarly communication in the humanities (Pedersen, Grønvad & Hvidtfeldt 2020; Franssen & Wouters 2019; Thelwall & Delgado 2015).

https://doi.org/10.1515/9783110753523-006

In this chapter we argue that big data can indeed help us better understand the dynamics of the humanities. To do so, we call for a more data intensive approach on the technical side, and a more sociologically driven notion on the theoretical side. Humanities has been addressed since the 1980s as a technical limitation to be disclosed rather than as a scientific enquiry to be explored (Franssen & Wouters 2019). With notable exceptions (Ochsner, Hug & Daniel 2013; Hammarfelt 2016), most of the limitations observed on the use of scientometric methods to understand these fields, were considered technical rather than conceptual. From issues with publication types (Hicks 2005) to data coverage (Hicks 1999) or the lack of infrastructure (Kulczycki et al. 2018). As Franssen and Wouters (2019) state, this has largely to do with an urge to evaluate and monitor scholarly activity for research policy purposes.

However, to address policy-related questions, more fundamental ones must be answered first. How do humanists disseminate their outcomes? How can they be characterized? How is new knowledge produced and research topics shaped in these fields? How do humanities and society interact and cross boundaries to shape each other?

This chapter illustrates some of the possibilities Big Data techniques offer to researchers interested on understanding the dynamics of humanistic studies. For this, two case studies are discussed. The first case study describes the possibilities for accessing and merging large datasets of academic literature to identify and analyze the oeuvre of humanists. We discuss a specific case in which natural language processing techniques are used to identify humanists from Spanish speaking countries from two major international databases, and then both sets of data are merged into a unique one. The second case study makes use a machine learning technique called archetypal analysis (Cutler & Breiman 1994) in order to identify the publication profile of researchers in different fields. In this second case, the goal is to discuss how machine learning techniques can help us delve into big datasets to better characterize the humanities.

2 Big Data and the Identification of Spanish Speaking Humanists Worldwide

2.1 From Publications to People: The Power of Author Identifiers and Name Disambiguation Algorithms

One of the main fundamental shifts that the era of big data has brought to the field of scientometrics is the change on the unit of analysis from publications to people. Author name disambiguation is one of the most fundamental challenges

to which any information retrieval system is confronted. The task of disambiguating author names has been traditionally related to the field of library and information science. The great difficulties encountered in bibliometric studies to accurately assigning corpuses of publications to single authors has discouraged author-level analyses until quite recently (Ruiz-Pérez, López-Cózar & Jiménez-Contreras 2002; Costas & Bordons 2005).

The development of name disambiguation algorithms along with the expansion of author registries in recent years now make this type of approaches feasible (Costas, Corona & Robinson-Garcia forthcoming; Tekles & Bornmann 2020). Still, scientometric studies devoted to the study of publication patterns in the humanities still adopt a publication-level perspective. Following, we will show how these is possible also at the author level by combining information from two unique data sources: Dialnet and the Open Researcher and Contributor ID platform also known as ORCID. What makes these two databases unique is that they are both of free access and include an author registry with their publication records along with biographical data (e.g., institutional affiliation, educational record).

2.2 Brief Overview of Dialnet and ORCID

2.2.1 Dialnet

Dialnet is one of the larges bibliographic databases, containing scientific literature from Spanish-speaking countries in the fields of the Humanities, Social Sciences and Legal Sciences. It is hosted by Fundación Dialnet, a non-profit organization belonging to the University of La Rioja, in Spain. Originally launched in 2001, Dialnet is maintained jointly by university, public and special libraries from all over Spain and Latin America. Librarians from the different partner institutions oversee the retrieval and processing all the records included in the database, as well as cleaning and managing author profiles.

For each author profile, information on additional author identifiers (including the ORCID, which we later discuss), affiliation data, research discipline and their publication record. Disciplines are organized based on the 190 knowledge areas established by the Spanish Ministry of Education and Vocational Training. 40 of these disciplines belong to the Humanities fields. These are grouped into 13 major areas: History, Philology, Arts, Philosophy, Archaeology, Language & Linguistics, Music, Anthropology, Literature, Translation & Interpretation, Geography, Paleontology, and Cultural Studies.

For these areas a total of 57,851 scholars associated with 752,423 unique publications were identified. Dialnet divides scholars' outputs into three document types: articles, book chapters and books. A fourth document type was identified, proceedings papers. These were located using text mining techniques and applying them to book titles. Our final corpus of publications consisted of 416,537 journal articles, 213,088 book chapters, 89,195 books, and 33,603 proceedings papers. Figure 1 offers an overview of their distribution by major field and time evolution by document type.

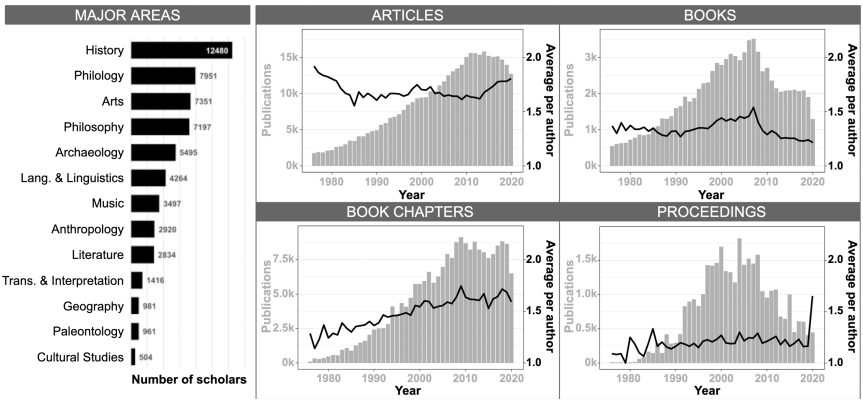


Figure 1: Publication trends of Spanish-speaking academics by documental type according to Dialnet. 1976–2020 period. Bars indicate the total number of unique publications; lines are the average number of publications per author.

2.2.2 ORCID

ORCID is an open author registry which assigns a unique identifier to each scholar signed up in the platform. Using self-reported data as well as connecting with scientific and scholarly publishers and databases, ORCID creates an author profile highlighting information related to their education, employment, scholarly outcomes, funding peer review activity among others. As a user platform, ORCID relies entirely on authors to register, meaning that its coverage is limited by its use. A more thorough description of ORCID and its possibilities for bibliometric analyses is offered by Costas et al. (forthcoming). In this case, the identification of scholars from the humanities was not as straightforward and advanced techniques were needed. These are described in the next section.

2.3 Identification of Scholars from the Humanities in ORCID

For this analysis, we downloaded in 2020 the complete dataset of over 9.5 million records from the ORCID API (https://info.orcid.org/documentation/features/public-api/). This dataset includes all users at the time of the download regardless of their field of research. In order to identify scholars from the humanities, we queried author keywords and affiliation departments. In both cases, this information is self-reported, and non-mandatory. This already imposes an important limitation as only 588,794 included keywords (6.14% of total), while 1,943,623 (20.28%) included information on their affiliation. In total, 2,164,093 (27.20%) users were included in our analysis. This already shows one of the challenges when dealing with large datasets: the messiness and incompleteness of datasets.

Furthermore, we filtered only for scholars affiliated to Spanish speaking countries based on their affiliation data. A total of 170,177 user profiles were identified. To identify which of these authors belonged to the humanities, we adopted stepwise approach. Table 1 summarizes each step and describes the big data technique used.

Table 1: Summary of the ORCID humanist identification process and data science methods employed.

	Description	Result	Method
1	Create a co-occurrence network based on ORCID profile keywords	Thematic landscapes	Social network analysis
2	Detect the main topics to select those related to humanists	Identification of fields	Community detection algorithm
3	Query departments which may belong to fields in the humanities	List of departments	Text mining

Step 1. Creation of thematic landscapes. We used text mining techniques to normalize authors' keywords. Through regular expressions, special characters (e.g., accents) and errors in text strings (e.g., spaces at the beginning and/or end) were fixed and removed. A vector of keywords was obtained for each ORCID record. As authors may describe their research either in Spanish language or English language, both languages were considered when identifying author keywords. Figure 2 visualizes the so-called thematic landscape of humanist authors' keywords for both English and Spanish languages, separately. By thematic landscape we refer to a visualization of terms based on their co-occurrences in each author's profile. That is, each node in the map refers to a keyword used by an author in their ORCID profile.

Nodes' size reflects the number of users who have included that keyword to describe their work. Distance between nodes reflect how many times they co-occur in a user profile, that is, the number of times both keywords are included together in the same profile.

Step 2. Detection of fields. Colors in Figure 2 reflect the field to which they belong. Fields are identified also using a machine learning technique, in this case a community detection algorithm (Traag, Waltman & van Eck 2019). This algorithm is both, used to filter to only fields from the humanities and to identify each specific field. In the case of the landscape including English keywords (Figure 2A), the network is composed of 2,079 keywords, and 8 disciplines are identified: Arts and Cultural Studies, Sociology and Gender Studies, Law, Political Science, Archeology and Anthropology, Philosophy, History, and Social Media. For the Spanish landscape (Figure 2B), 1,075 keywords are included, and 8 disciplines are identified: Social Anthropology and Gender Studies, Arts and Patrimony, Social Media and Communication, History, Literature and Linguistics, Philosophy, Archeology, and Performing Arts.

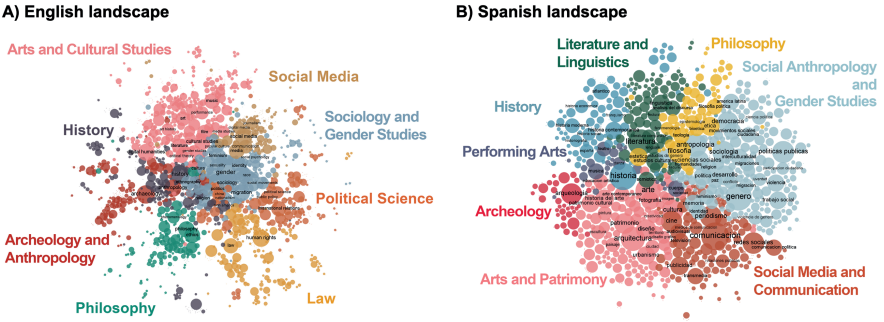


Figure 2: A) English and B) Spanish thematic landscapes of humanists in ORCID. The maps show author keywords co-occurrence networks in the humanities.

The main areas of the humanities are covered in both cases, with some differences such as Law and International Relations. In this sense, the way in which they are grouped and related also differs. For example, while the Arts appear together in Figure 2A, they appear dispersed in several communities in Figure 2B. Likewise, while English terms are more generic in scope, Spanish terms are more specific.

Step 3. Identification of departments in fields from the humanities. In addition to the detection of humanists based on the keywords specified in the ORCID record, the name of the department in which the scholars work or have worked

was also used. This search was performed on the basis of department names in both English and Spanish. Due to the diversity of departments, the search was performed using regular expressions instead of a search based on exact names. Regular expressions allow to define much more specific searches to locate those text strings that follow a certain pattern, for example department names that include the terminological root “phylolog” or that include both “geography” and “planning”. Based on the fields identified previously as well as looking at the structure of different universities both in Spain and Latin America, we searched for departments related to art, anthropology, antique, archeology, classical studies, dance, geography and spatial planning, history, human geography, humanities, translation studies, language (and some major ones such as English, Spanish, French or Russian), literature, music, paleontology, philology, philosophy, religion, theater or theology, among others.

After a first search, a second filter was applied in which departments retrieved in the wrong way or not clearly related to the humanities were eliminated. We prioritized precision over exhaustiveness, eliminating clear cases of wrong assignment. For instance, when querying for departments which included the word ‘language’ in their name, we would retrieve i.e., department of Language and Linguistics, but also department of Languages and Computer Systems.

A total of 33,491 scholars were identified as being associated either through their keywords or affiliation data to fields related to the humanities. Of these, 16,198 were located by keywords and 27,556 by affiliation. 20,361 (60.8%) of them include at least one work in their publication record. In total, after preprocessing the data, 409,189 unique works were identified. 356,122 are publications (87.03%), 38,600 conferences (9.43%), 607 intellectual property (0.15%), and 13,860 others (3.39%).

2.4 Integrating Datasets from Different Databases

A key challenge when merging datasets belonging from different data sources is the unification of records (that is, rows in a data table) and variables (i.e., columns). This process can be especially complicated and limiting, as each data sources will have its own structure and identifiers of records. Furthermore, some records may be present in both data sources but including different levels of completeness. E.g., a scholar may be present in both Dialnet and ORCID but have different publication outputs included in each database.

In our case, we overcame these limitations by using different approaches for merging the data. First, we merged publication records which were both in Dialnet and ORCID. First, we unified and grouped the number of document types in

both databases to journal articles, books, chapters, conferences (in the case of ORCID it is the sum of abstract, paper and poster of conferences) and others (these are document types included in ORCID but not in Dialnet).

Second, we merged publications which were present in both datasets. For this, the use of publication identifiers (mainly Digital Object Identifiers or DOIs, and ISBNs) is again indispensable (Sandberg & Jin 2016; Mayernik & Maull 2017). Figure 3A shows the overlap of records between the two databases. As observed, there is an overlap of 43,105 publications. That is 5.73% of all records retrieved from Dialnet and 10.53% of those retrieved from ORCID. While most of the publications identified as articles, books and chapters came from Dialnet, more than have of the conference proceedings identified came from ORCID. This already reflects the problems of coverage that can be present in any study using only one data source to analyse scholars' outputs.

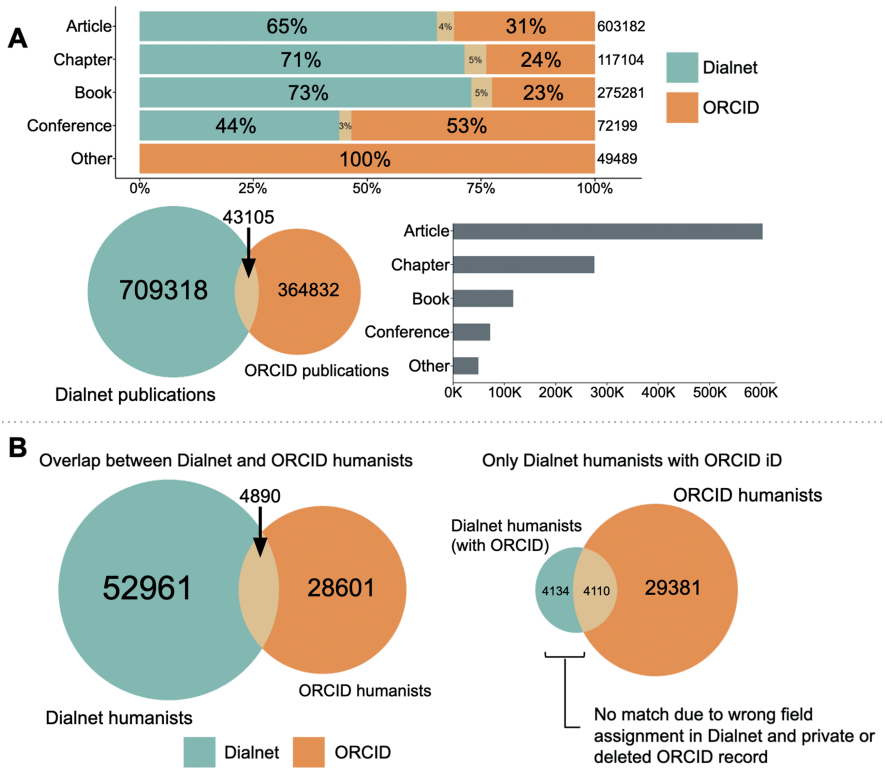


Figure 3: Merging of Dialnet and ORCID datasets for A) publication records and B) scholars.

Once identified the overlap between datasets and the publication level, we must identify overlapping profiles of scholars. To do so we combined two different approaches. First, we linked profiles which had common identifiers. As both Dialnet and ORCID include a field of 'other researcher identifiers', we matched profiles of users which included the Dialnet ID or ORCID in both databases. 8,244 Dialnet authors include an ORCID, 5,347 include a Scopus Author ID (the identifier used by the scientific database Scopus), and 2,210 include a Researcher ID (the identifier used by the scientific database Web of Science) (Boudry & Durand-Barthez 2020). In ORCID, the presence of other identifiers is scarce. As a tool that uses self-reported information, it seems to be much more incomplete than Dialnet, which is fed and updated by librarians and information professionals. Only 276 of the identified humanists include the Dialnet author identifier in their profile. Using this approach we identified up to 4,184 authors present in both sets.

The second approach used consisted of the identification of authors who shared the same name and surname. This approach is not free of limitations as there may be cases in which two researchers may share the same name, while in other cases they are the same individual. To avoid problems derived from ambiguity we use the set of common publications to identify possible matches. Name initials and surnames are matched, having previously unified the formatting in both datasets to facilitate this process (i.e., eliminating accents, transforming the text to lowercase and removing hyphens). We found that 629 scholars present in both Dialnet and ORCID who do not only matched by their name and surname, but also had at least one publication present in both platforms. These profiles were also merged in our final dataset.

Our final dataset included a total of 86,452 scholars, out of which 4,890 are present in databases (Figure 3B). 85.56% of the overlapping researchers were identified through the matching of author identifiers, while 14.44% were merged by name matching. Figure 4 provides an overview of the distribution of scholars by field and country.

3 Profiling Types of Scholars by Their Publication Patterns

In the last decades, one of the topics of concern to scientometricians in relation to the humanities, has been the publication patterns and forms of production of knowledge (Franssen & Wouters 2019). This second study identifies the different profiles of humanists according to their publication patterns. With this case study we aim at responding at the following questions. Can we identify different pro-

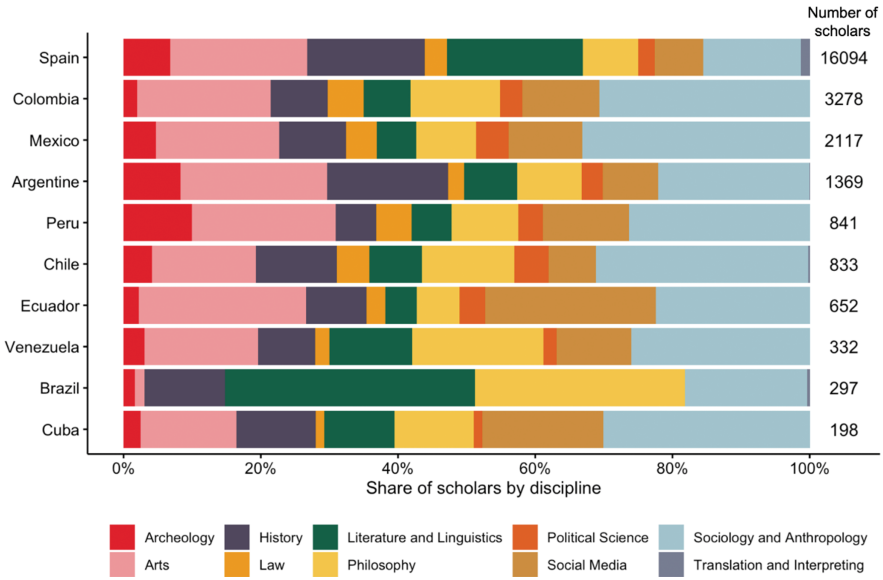


Figure 4: Distribution of humanist scholars identified in both Dialnet and ORCID for the top 10 countries by major field.

files of humanists based on the types of outputs they produce? How are humanists distributed among these profiles or archetypes? Do we observe differences by discipline? This analysis is only possible when analysing large datasets in order to observe reliable and robust patterns, and large datasets can only be exploited by using Big Data techniques. Following we describe the methodology followed to identify profiles of scholars and the data processing and methodological design followed. We conclude by showing the results of our case study and discussing the findings.

3.1 Archetypal Analysis

Machine learning methods are commonly categorized into two types: supervised learning and unsupervised learning (Soni 2018). Supervised learning methods are those that aim at predicting either values or categories. Unsupervised learning methods aim at learning the structure or features of a dataset.

In our case, we wish to identify types of scholars based on their publication patterns, hence an unsupervised learning method should be applied. In these cases, normally dimension reduction or clustering methods are considered. However,

these can be problematic as they simplify reality and force cases into categories as if they were alike. To avoid this limitation, we propose conducting an archetypal analysis (Cutler & Breiman 1994). The archetypal analysis identifies different archetypes or profiles that emerge from a given multivariate dataset. Archetypes are combinations of extreme observations represented as convex combinations of the observations in the dataset that result from a least squares problem.

While not as popular as other machine learning methods, archetypal analysis has been previously used in scientometrics for similar problems as the one showcased here. For instance, Seiler and Wohlrabe (2013) applied it to identify archetypal scientists based on a set of publication and citation variables. Robinson-Garcia et al. (2020) used archetypal analysis to identify types of researchers based on the types of contributions they did in publications when collaborating. More recently, Ramos-Vielba, Robinson-Garcia and Woolley (2021) applied archetypal analysis to a dataset combining scientometric, altmetric and survey data to better understand science-society interactions.

The most interesting characteristic of archetypal analysis is that, contrarily to clustering techniques, it does not group cases, but shows the distance of each case to each of the identified archetypes. This distance is called α score and provides a value that ranges from 0 to 1, being 1 a complete resemblance with a given archetype. As an example, lets considered a dataset of individuals for which we have the total number of journal articles and books that each individual has produced. As observed in Figure 5, the first stage will consist on determining the number archetypes identified in the data. For this, we will first calculate the residual sum of squares (RSS) is used, which indicates how well the individuals fit the archetypes. The lower the value the better the fit. In general, the RSS is obtained from several models, each with a different number of archetypes, and the one with the lowest value is selected. However, given that the greater presence of archetypes usually leads to a reduction in this value, we have used the so-called “elbow criterion”, selecting not the lowest value but the one where there is a significant reduction and from which there is a flattening of the residual sum of squares of a multivariate dataset. Then, using an “elbow criterion”, we will establish which is the most appropriate number of archetypes.

In Figure 5, bottom-right, we observe a scatterplot with a triangle overlaid on it. Each dot represents an individual, while the axis of the triangle represents our three archetypes. The α score of each individual will represent the distant from the individual to each archetype. Furthermore, for each archetype we can extract the expected values for each variable (in our case, books and articles) that a complete resemblance to the archetype would have, facilitating the interpretation of each of the archetypes.

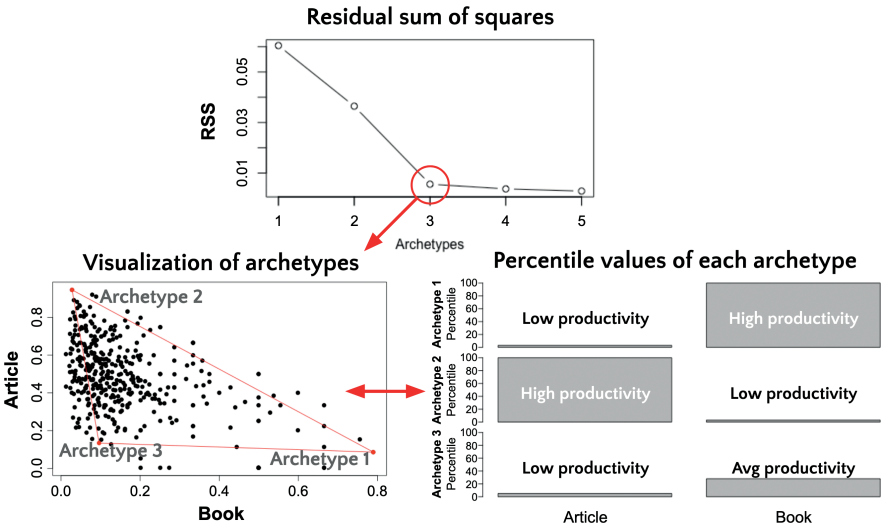


Figure 5: Process of identification, visualization and interpretation of archetypes.

3.2 Data Processing and Selection of Variables Under Analysis

In this case we will work the set of humanist scholars identified via Dialnet, incorporating for those also identified in ORCID, the output identified in the latter database. In order to facilitate the interpretation of the findings, we have removed extreme cases that may bias our results and have only included scholars who have published between 4 and 400 outputs. This leaves us with a total of 33,773 scholars.

Table 2: Variables used for the archetypal analysis, definition and data source.

Variable	Definition	Source
Books	Share of edited or authored books from their total output	Dialnet; ORCID
Book chapters	Share of authored book chapters from their total output	Dialnet; ORCID
Journal articles	Share of journal articles from their total output	Dialnet; ORCID
Proceedings papers	Share of proceedings papers from their total output	Dialnet; ORCID
Non-scholarly documents	Share of non-scholarly publications from their total output	ORCID

Table 2 (continued)

Variable	Definition	Source
International output	Share of publications indexed in Scopus or Web of Science from their total output	Dialnet; ORCID; Scopus; Web of Science
Publications	Total number of outputs	Dialnet; ORCID

For each scholar we include 7 variables as defined in Table 2. Non-scholarly documents include reports, online resources, translations, artistic performances, encyclopedia entries, manuals, websites, dictionary entries, data sets, registered copyright, research tools, disclosures, patents, standards and policy, inventions, softwares, and press articles. In the case of international publications, DOIs and ISBNs were used in order to identify publications which were also indexed in Web of Science or Scopus.

3.3 Identification and Interpretation of Profiles

The archetypal analysis was performed using the statistical programming language R (R Core Team 2021) and the archetype package (Eugster & Leisch 2009). We performed the analyses for all scholars, as well as for those belonging to the fields of Archeology, Philology and Philosophy. The results are shown in Figure 6. As observed, the number of identified archetypes varies depending on the dataset or subset used.). For each analysis we show the normalized value expected for each value per archetype (left-side) and the distribution of scholars based on their α score.

Figure 6A shows the results for the complete dataset. Overall, three different profiles were identified. The most common profile is that of humanists who publish papers and who have a greater international projection, although their productivity is average. The second most common profile is that of those who publish books and chapters, mostly national, and with higher productivity. The third and most minority profile is the most mixed, standing out above all for publication in conferences and non-scholarly materials, also with high productivity.

Figure 6B shows the archetypes for archaeologists. Again, three archetypes emerge. Archetype 1 is quite similar to archetype 2 in Figure 6A, and as it happened, quite rare. The archetype to which more scholars resemble is archetype 2, characterized by an average productivity of mainly journal articles published in international venues. This focus on international journal articles is also observed

for philologists (Figure 6C, archetype 1), who also seem to publish book chapters but do not rely on other types of materials. Finally, the most diverse of the three fields seems to be Philosophy (Figure 6D), for which four archetypes are identified. Scholars in this field are mostly characterized by archetypes 1 and 2. In the former case, these are scholars with low productivity values who mainly publish book chapters and proceedings papers. The latter is formed by scholars with an average productivity of mainly journal articles indexed in international venues.

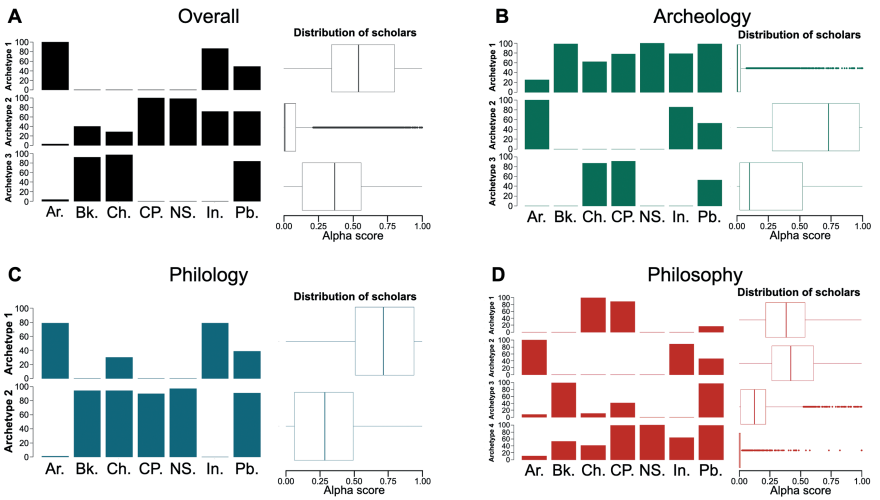


Figure 6: Percentile values of three archetypes and distribution of humanists for each one.
Note: Ar. (Article), Bk. (Book), Ch. (Chapter), CP. (Conference proceedings), NS. (No scholarly), In. (International), Pb. (Publications).

4 Towards a Science of the Humanities

In a relatively recent study, Hammarfelt discussed 'the possibilities of establishing a bibliometrics for the humanities' (2016: 116) and concluded that 'evaluations that use bibliometrics might provide a valuable complement to traditional peer review' (2016: 127). Here we paraphrase Derek de Solla Price (Price 1963) who favored the term of a humanities of science over a science of science, and advocate for a Science of the Humanities. We do this in the belief that beyond assessments and policy issues, both for those interested on the quantitative study of science and humanists. This of course goes beyond traditional scientometric approaches, as big data and machine learning techniques are introduced and more fundamental and theoretical questions are presented.

This chapter shows two examples that illustrate how big data can help better understand the knowledge production mechanisms of the humanities. But these are just glimpses of the many opportunities presented to those interested on this. Natural language processing techniques allow us analyze how research interests and topics evolve in the humanities, while altmetrics can serve to identify societal perceptions on culture (see the chapter by Gallego-Cuiñas and Torres-Salinas in this same volume). Furthermore, emerging fields like Cultural Analytics (Manovich 2020) can greatly benefit of these methodological innovations plus the expertise already developed in the field of scientometrics.

Impact is one of the dimensions not considered here but that deserves further attention and in which qualitative approaches combined with machine learning techniques can serve to better understand how this is operationalized in the humanities (Ochsner, Hug & Daniel 2013). Not only considered as citation impact, but also introducing other metrics such as altmetrics, that is those derived from social media mentions.

Another field of interest is the boundary between what is local and international literature in the fields of the humanities and the interaction between local stakeholders with global interests and the role played by multilingualism. In this work we have analyzed the production of Hispanic humanists, differentiating only between international publications, which has been useful in the identification of humanist profiles. But this could be better fine-grained and analyzed. We encourage readers to further explore some of these topics and many others inspired by the ideas presented in this chapter.

Bibliography

Boudry, Christophe & Manuel Durand-Barthez (2020): "Use of author identifier services (ORCID, ResearcherID) and academic social networks (Academia.edu, ResearchGate) by the researchers of the University of Caen Normandy (France): A case study", in *PLOS ONE. Public Library of Science*, 15. e0238583. doi:10.1371/journal.pone.0238583.

Cabezas-Clavijo, Álvaro, Nicolás Robinson-García, Daniel Torres-Salinas, Evaristo Jiménez-Contreras, Thomas Mikulka, Christian Gumpenberger, Ambros Wemisch & Juan Gorraiz (2013): "Most borrowed is most cited? Library loan statistics as a proxy for monograph selection in citation indexes", in *arXiv*.1305.1488 [cs].

Costas, Rodrigo & María Bordons (2005): "Bibliometric indicators at the micro-level: some results in the area of natural resources at the Spanish CSIC", in *Research Evaluation*, 14, pp. 110–120. doi:10.3152/147154405781776238.

Costas, Rodrigo, Carmen Corona & Nicolas Robinson-Garcia (forthcoming): "Could ORCID play a key role in meta-research? Discussing new analytical possibilities to study the dynamics of science

and scientists". In Alis Oancea, Gemma E. Derrick & Nuzha Nuseibeh (eds.), *Handbook on Meta-Research*. Cheltenham: Elgar Publishing, pp. 1–20.

Cutler, Adele & Leo Breiman (1994): "Archetypal Analysis", in *Technometrics*, 36, pp. 338–347. doi:10.1080/00401706.1994.10485840.

Eugster, Manuel J. A. & Friedrich Leisch (2009): "From Spider-Man to Hero — Archetypal Analysis", in R. *Journal of Statistical Software*, 30, pp. 1–23. doi:10/ggck2c.

Franssen, Thomas & Paul Wouters (2019): "Science and its significant other: Representing the humanities in bibliometric scholarship", in *Journal of the Association for Information Science and Technology*, 70, pp. 1124–1137. doi:10.1002/asi.24206.

Garfield, Eugene (1996): "Citation indexes for retrieval and research evaluation. *Consensus conference on the theory and practice of research assessment*", in *Capri*, 7.

Hammarfelt, Björn (2014): "Using altmetrics for assessing research impact in the humanities. *Scientometrics*", 101, pp. 1419–1430. doi:10.1007/s11192-014-1261-3.

Hammarfelt, Björn (2016): "Beyond Coverage: Toward a Bibliometrics for the Humanities". In Michael Ochsner, Sven E. Hug & Hans-Dieter Daniel (eds.), *Research Assessment in the Humanities*. Cham: Springer International Publishing, pp. 115–131. doi:10.1007/978-3-319-29016-4_10.

Hicks, Diana (1999): "The difficulty of achieving full coverage of international social science literature and the bibliometric consequences", in *Scientometrics*, 44, pp. 193–215.

Hicks, Diana (2005): "The Four Literatures of Social Science". In *Handbook of Quantitative Science and Technology Research*. Cham: Springer, pp. 473–496.

Kousha, Kayvan & Mike Thelwall (2009): "Google book search: Citation analysis for social science and the humanities", in *Journal of the American Society for Information Science and Technology*, 60, pp. 1537–1549. doi:10.1002/asi.21085.

Kulczycki, Emanuel, Tim C. E. Engels, Janne Pölönen, Kasper Bruun, Marta Dušková, Raf Guns, Robert Nowotniak, et al. (2018): "Publication patterns in the social sciences and humanities: evidence from eight European countries", in *Scientometrics*, 116, pp. 463–486–24. doi:10.1007/s11192-018-2711-0.

Linmans, A. J. M. (2010): "Why with bibliometrics the Humanities does not need to be the weakest link", in *Scientometrics*, 83, pp. 337–354. doi:10.1007/s11192-009-0088-9.

Manovich, Lev (2012): "Trending: The promises and the challenges of big social data". In Matthew K. Gold (ed.), *Debates in the digital humanities*, 2, pp. 460–475. Minneapolis, MN: University of Minnesota Press.

Manovich, Lev (2020): *Cultural analytics*. Cambridge: Mit Press.

Mayernik, Matthew S. & Keith E. Maull (2017): "Assessing the uptake of persistent identifiers by research infrastructure users", in *PLOS ONE*, 12, e0175418. doi:10.1371/journal.pone.0175418.

Merton, Robert K. (1973): *The sociology of science: Theoretical and empirical investigations*. Chicago: University of Chicago press.

Nalimov, V. V. & Z. M. Mulchenko (1971): *Measurement of Science: Study of the Development of Science as an Information Process*. Washington DC: Foreign Technology Division.

Nederhof, Anton J. (2006): "Bibliometric monitoring of research performance in the Social Sciences and the Humanities: A Review", in *Scientometrics*, 66, pp. 81–100. doi:10.1007/s11192-006-0007-2.

Ochsner, Michael, Sven E. Hug & Hans-Dieter Daniel (2013): "Four types of research in the humanities: Setting the stage for research quality criteria in the humanities", in *Research Evaluation*, 22, pp. 79–92. doi:10.1093/reseval/rvs039.

Pedersen, David Budtz, Jonas Følsgaard Grønvad & Rolf Hvidtfeldt (2020): "Methods for mapping the impact of social sciences and humanities – A literature review", in *Research Evaluation*, 29, pp. 4–21. doi:10.1093/reseval/rvz033.

Peng, Roger D. (2011): “Reproducible research in computational science”, in *Science (New York, Ny)*, 334, pp. 1226–1227. doi:10.1126/science.1213847

Price, Derek John de Solla (1963): *Little science, big science*. Columbia University Press New York.

R Core Team (2021): *R: A Language and Environment for Statistical Computing*. https://www.R-project.org/: R Foundation for Statistical Computing.

Ramos-Vielba, Irene, Nicolas Robinson-Garcia & Richard Woolley (2021): “A value creation model from science-society interconnections: Archetypal analysis combining publications, survey and altmetric data”, in *SocArXiv*. doi:10.31235/osf.io/tkzc5.

Robinson-Garcia, Nicolas, Rodrigo Costas, Cassidy R Sugimoto, Vincent Larivière & Gabriela F. Nane (2020): “Task specialization across research careers”, in Peter Rodgers & Allison Morgan (eds.), *eLife*. *eLife Sciences Publications*, Ltd. 9. e60586. doi:10.7554/eLife.60586.

Ruiz-Pérez, R., E. Delgado López-Cózar & E. Jiménez-Contreras (2002): “Spanish personal name variations in national and international biomedical databases: implications for information retrieval and bibliometric studies”, in *Journal of the Medical Library Association*, 90, pp. 411–430.

Sandberg, Jane & Qiang Jin (2016): “How Should Catalogers Provide Authority Control for Journal Article Authors? Name Identifiers in the Linked Data World”, in *Cataloging & Classification Quarterly*, 54, pp. 537–552. doi:10.1080/01639374.2016.1238429.

Seiler, Christian & Klaus Wohlrabe (2013): “Archetypal scientists”, in *Journal of Informetrics*, 7, pp. 345–356. doi:10.1016/j.joi.2012.11.013.

Soni, Devin (2018): “Supervised vs. Unsupervised Learning”, in *Towards data Science* <https://towardsdatascience.com/supervised-vs-unsupervised-learning-14f68e32ea8d> (06.05.2022).

Tekles, Alexander & Lutz Bornmann (2020): “Author name disambiguation of bibliometric data: A comparison of several unsupervised approaches”, in *Quantitative Science Studies*, 1, pp. 1510–1528. doi:10.1162/qss_a_00081.

Thelwall, Mike & Maria M Delgado (2015): “Arts and humanities research evaluation: No metrics please, just data”, in *Journal of Documentation*. doi:10.1108/JD-02-2015-0028.

Torres-Salinas, Daniel & Henk F. Moed (2009): “Library Catalog Analysis as a tool in studies of social sciences and humanities: An exploratory study of published book titles in Economics”, in *Journal of Informetrics*, 3, pp. 9–26. doi:10.1016/j.joi.2008.10.002.

Torres-Salinas, Daniel, Rosa Rodríguez-Sánchez, Nicolás Robinson-García, J. Fdez-Valdivia & J. A. García (2013): “Mapping citation patterns of book chapters in the Book Citation Index”, in *Journal of Informetrics*, 7, pp. 412–424. doi:10.1016/j.joi.2013.01.004.

Traag, V. A., L. Waltman & N. J. van Eck (2019): “From Louvain to Leiden: guaranteeing well-connected communities”, in *Scientific Reports*, 9, pp. 5233. doi:10.1038/s41598-019-41695-z.

Wouters, Paul, Zohreh Zahedi & Rodrigo Costas (2019): “Social media metrics for new research evaluation”. In Wolfgang Glänzel, Henk F. Moed, Ulrich Schmoch & Mike Thelwall (eds.), *Handbook of Quantitative Science and Technology Research*. Cham: Springer, pp. 687–713.

Francisco Benítez and Esteban Romero

What is Blockchain and How can it Help the Humanities?

1 Introduction to Blockchain and its Philosophy

First of all, a short definition of blockchain to introduce what this technology is. A blockchain is a distributed ledger that allows for the storage and transmission of information over the Internet in a transparent and secure manner without the need to rely on a trusted third party. The database contains transactions that are publicly auditable, validated, executed and saved in a chronological tamper-resistant manner by a distributed network of computers.

A blockchain is to a transaction, as the Internet is to information. Its qualities are attributed to it by its applications creating a network of value more than a mere information deployed in the network. The idea behind is to transfer information from simple networks to smart networks creating new added value (Swan & De Felipi 2017: 605).

This goes beyond the Internet Revolution and, for the first time in the history of technical revolutions, a technology has the capacity to affect the vertical and centralized power of states with regard to the economy: money, banks and financial transactions. But also we can define a decentralization in regard to energy, electricity, properties, and social and political institutions.

2 But, What is Blockchain?

From a technological point of view a blockchain is a distributed database, which is shared and agreed upon in a peer-to-peer network. It consists of a linked sequence of blocks, containing a timestamp (for each of the blocks) and the transactions secured by a cryptographic public key and verified by the entire network community. Once an item is added to the blockchain, it cannot be altered, becoming an immutable record of past activity.

The previous definition is the simplest way with which we have been able to define one of the technologies that is called to change society and, especially, the way in which we handle data of any kind. Despite this simplicity in the description, it hides many complexities, which make it abstract for the general public and, on many occasions, difficult to incorporate into the processes of companies and institutions.

https://doi.org/10.1515/9783110753523-007

Distributed ledger technologies (DLTs), currently make up a broad and complex ecosystem, which has multiple definitions, and according to literature reviews is quite inconsistent. As in the rest of all blockchain technology, we can affirm that there is still a lack of terminological standardization (Rauchs et al. 2018: 11).

As has happened in the history of computer science, before the emergence of a new technology or paradigm shift, such as the Internet, there have been previous works that paved the way for developing new disruptive models. Currently there are a large number of DLTs with different configurations and typologies, which on many occasions, make it very difficult to establish a clear taxonomy of how they operate and are constituted. There are a lot of new ways to build up new DLTs more efficiently and with new perspectives of use, but referring to current projects with a clear economic value, the two more used are Bitcoin and the Ethereum ecosystem. Both of them are just the starting point, and Ethereum was also the first platform to use smart contracts, the base of the actual *tokenomics* world (the tokenization of the crypto economic and social projects).

But before defining what is a smart contract, we want to define what is a DLT. The concept of distributed ledger technologies (DLT) has been established as a general term to designate multi-party systems that operate in an environment with no operator or central authority, even though the parties involved may be unreliable or malicious and in harsh environments. Blockchain technology is considered a specific subset of the broader DLT ecosystem, using a particular data structure consisting of a chain of linked data blocks with cryptographic hashing functions. Conceptually, DLTs were first described in 1982,[1] and the concept of blockchain in 1991 (Haber and Stornetta 1991). However, we are in their deployment phase, before they are massively incorporated into society.

It is necessary to clarify that a hostile environment in a DLT is characterized by the presence of malicious actors within the system or network, who undermine it by using it in a way that it was not intended. The prototype adversary in a DLT system is an entity that attempts to exploit consensus rules to transfer assets without authorization, censor the transactions of others, or otherwise disrupt or destroy the network. Adversaries can operate both inside (on-chain) and outside the system (off-chain). For all these reasons, the governance schemes to establish the management framework seem crucial in the management of any type of platform (Brown and Grant 2005).

What is a Smart Contract? Actually, the term refers to any script (the so-called smart contract), which is executed by itself, automatically, and without the

1 http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.12.1697.

need for any intermediary entity. Despite the "intelligent" name, no type of Artificial Intelligence process is involved in its execution. This is written as a computer program (the script) that is self-executing, instead of being written in the customary legal language of the physical world. The script can define processes, rules and strict consequences in the same way that it operates on a legal document. But unlike these, a smart contract can manage information that is provided to it externally in order to complete its routine successfully. That is, to effectively terminate the processes for which it has been programmed.

There are a lot of more properties and items in a DLT as we explain in a recent publication (Romero-Frias et al. 2021), but understand what is a blockchain, a DLT and a smart contract is the basis to understand which means the concept of tokenization to understand what is a Non Fungible Token.

3 NFTs and Tokenization

In recent years, there are hundreds of cryptocurrencies on the market. Only very few get a quotation base with fiat money on it. Most of them have been implemented on the Ethereum platform, and the way in which they manage the mining of their cryptocurrencies is through Smart Contracts that generate a token, and each of them can present different properties, according to the rules that have been applied, and they have been defined in the governance system of its "White paper". The standard token is the ERC-20, which is also the most popular within the aforementioned platform. The work of Chen et al. (2020) reviews the ICOs that have been based on this token, which exceed 80%, which demonstrates the weight that this standard has within the Ethereum platform.

It is important to point out the differences between cryptocurrency and token. Terms that are often confused, due to the influence of fintech solutions in the current blockchain market. Cryptocurrencies are the form of digital money that are created by blockchain solutions, while the token represents an asset or a utility that has a specific value (tangible or not) within the community that has created it. They are usually transferable goods that can range from loyalty points, game bonuses, or future rights to a service that can be redeemed when the agreed result occurs.

What is a token? We can define that **a token is a digital asset that operates "on top" of a cryptocurrency or a blockchain**, which often runs as a programmable asset thanks to a Smart Contract, to be used within a project or a dApp.[2]

When we consider that cryptographic tokens represent the right to something, we are defining the tokenization of a digital asset. Tokenization is a way of turning the rights of something or someone into a digital artifact, which takes on the digital format of a token. With crypto tokens, the benefits of tokenization lie primarily in greater versatility, greater liquidity, improved programmability, and immutable proof of ownership (di Angelo and Salzer 2020). However, there is still a great lack of tokenization standards and, above all, in most states there is also a lack of a legal infrastructure and a legal framework that regularizes and legally defines the concept of tokenization.

Following the taxonomy of Ritchey (2005) we will find the perspective of fungibility and regarding this, in economics, fungibility refers to the interchangeability of each unit of a product with other units of the same product. Examples of these could be durable goods, such as precious metals or fiat money. Expendable assets have two key properties: a) only quantity matters, which means that units of expendable assets of the same type are indistinguishable; and, b) any amount can be merged or split into a larger or smaller amount, making it indistinguishable from the rest. Fungible crypto tokens can represent any physical or digital asset that is identical to each other and therefore can be easily replaced. They are not unique and are perfectly interchangeable with other tokens of their type. If two parties have the same amount, they can exchange it without losing or gaining anything. Unique tokens, by contrast, are not fungible. Examples of this are identification cards, a token that represents ownership of a house, car, work of art or membership in a club, community or entity. If you lend a non-fungible token that is transferable to someone, you would expect them to return the same token, with the inherent properties that it confers. So, we are defining a Non Fungible Token (NFT).

What is a NFT? A NFT is a kind of ERC token. The most popular ERC token in Ethereum is the ERC-20 because it is widely used in the crypto market, but we are referring to a fungible token.

---

2 A decentralized Application (dApp) is an application that is stored and executable in a distributed environment. Early examples may be found in the blockchain environment of widespread systems like Bitcoin and Ethereum. While users may access a DApp like a traditional app via a user interface (front-end), the program logic and the data are not located on a centralized server, but rather on a peer-to-peer network (P2P), such as a blockchain or a DLT. Thus, dApps require no centralized services or platforms, which implies that no intermediary is necessary. In the financial context, many dApps have emerged under the umbrella of decentralized finance.

But a NFT is referred to as the ERC-721, the Non Fungible Token Standard. According to Entriken et al. (2018), it refers to the token that is different and distinctive from the rest and therefore allows the tracking of distinguishable and unique assets. Each asset must have its property tracked individually and indivisible. This standard requires compatible tokens to implement 10 mandatory functions and three events, which are associated with their execution.

Which means "standard" in this definition? A standard interface allows wallet/auction applications to work with any NFT on Ethereum. The smart contracts used track an arbitrarily large number of NFTs. This standard is inspired by the ERC-20 token standard.

A NFT (ERC-721) standardizes a safe transfer function within its framework in order to secure the transfers (transactions) in the applications that use a large number of NFTs. Note that we are talking about only the transactions, not about how to transmit the final possession of the original asset, and how to securize the transmission of this asset from A to B. The ERC-721 token would play a role very similar to that of title or writing, which assigns ownership to whoever owns it.

So, the objective behind this is to develop unique tokens, where their intrinsic value is given by their weirdness. This property makes an ERC-721 a collectable token.

The first project to use the ERC-721 was the CryptoKitties[3] card collectible platform (using Metamask[4] as the wallet to store and sell/buy these tokens), it was officially launched on December 3, 2017. CryptoKitties makes you a collector of virtual cats that base their value on its rarity (The CryptoKitties Genesis Card was sold on December 2, 2017 by an amount of 247 ETH).[5] And after that it was used also in Decentraland,[6] a project based in Ethereum where the users can create and trade with their NFTs and also this entire platform is owned by them. This platform was the first one where the users can trade their creations and artworks as digital assets. Following this market OpenSea[7] was created, where actually the most of the trade of NFTs is deployed.

3 https://www.cryptokitties.co/.

4 https://metamask.io/.

5 The ETH is Ether, the cryptocurrency of the Ethereum platform. In April of 2022 a ETH has an approximate value of 2.900 US$.

6 https://decentraland.org/.

7 https://opensea.io/.

4 Web4: Blockchain, NFTs and their Impact in Arts

Nowadays, NFTs are scaling into the limelight, inspiring a surge of enthusiasm and money around the technology, and the NFT community is still navigating its move into the mainstream. Estimates of total NFT sales in 2021 vary from $25[8] billion to $41[9] billion both a tremendous increase from 2020, when sales totaled more than $250[10] million. So, definitely we are not talking about just economic hype. The study of Nadini et al. (2021) is the most comprehensive work to date on its economic impact and future predictability.

But NFTs also have a dark side, not all the stories about NFT are about money and success. As they have become increasingly entrenched in the society, multiple digital platforms that facilitate the sale of NFTs must face allegations of fraud, plagiarism, service errors, and also environmental issues as described in the work of Rehman et al. (2021). Despite this, NFTs are an opportunity to represent a permanent shift in how artists, creators, and craftsmen can change their relationship with final consumers, avoiding third parties and adding new value to their work/creations.

One of the problems to face off is the gap to understand the complexity of NFTs in relation with blockchain technology. But also, the new regulations and how to create solid platforms that will avoid this complexity, not only for the consumers but also for creators as stated in the Vasan et al. (2022) work with the mapping of the Foundation platform.[11]

Actually, most NFTs trade in the marketplaces have zero value, because of the lack of security (technologically and in terms of regulation) on these platforms. Remember that value of a NFT is not in the transaction but in the security that it is not reproducible, that is it is unique and rare. So, securize a transaction in a block to transmit property from A to B without having the certainty that the property and the unique digital object are transmitted, is a problem for the future of this niche in Web3.[12]

---

8 https://www.reuters.com/markets/europe/nft-sales-hit-25-billion-2021-growth-shows-signs-slowing-2022-01-10/.

9 https://www.ft.com/content/e95f5ac2-0476-41f4-abd4-8a99faa7737d.

10 https://www.insider.com/nft-nfts-art-history-what-are-can-help-explain-hype-2021-3?amp.

11 https://foundation.app/.

12 This combination of the World Wide Web (WWW) and the third generation has evolved with decentralized technologies, such as blockchain and distributed ledgers. It recognizes the early phase of the WWW from 1992 until the beginning of the 2000s as first distributed ledgers from centralized platform providers to the users themselves, thus leading to more decentralization and democratization in the web. Web3 questions the role of established third parties such as

The combination of real products with a NFT to develop new characteristics of an item is the real base for the future of crafts and Arts. Web3 is an opportunity to add new value to any kind of data stored and signed by its owner. The path started by Toymint[13] is a good one to be deployed in future projects.

The possibilities that NFTs and new Web3 platforms have in the field of fine arts and crafts are endless, given the convergence between the creative capabilities of artists and the technological capabilities of distributed ledgers that we have only just begun to explore.

5 The Sociopolitical Properties of Blockchain

As we can see, blockchain is more than a promise (including its controversial topics: legal framework and its complicated understanding), offering various ways to imagine alternative models of politics and social schemes.

Blockchain appears as a powerful framework for a total decentralization with a great desintermediation, that includes an emerging tool beyond its actual use: the DAO, a decentralized autonomous organization. A DAO is an opportunity to transform the political institutions to develop new e-Voting or e-Participation (Benítez-Martínez et al. 2020) systems or to transform current procedures, as the procurement within the public administration platforms (Benítez-Martínez et al. 2022).

But, what is a DAO? We are defining a form of organization where multiple actors are organized by a decentralized software system. The concept has grown with distributed ledger technologies that include smart contracts for executing governance and the organizational rules. This allows many activities of a DAO to be carried out automatically without human intervention and without intermediaries. An example for new political bodies governed by a DAO is Bitnation,[14] creating a new model of "nation", the Decentralized Borderless Voluntary Nations (DBVN) within its Pangea Platform. Bitnation is self-considered as a new governance model. We are just seeing a new beginning.

banks, insurance companies and exchanges by replacing them with structures and processes of decentralized finance. Among the examples are electronic payments with cryptocurrencies, crowdfunding platforms, crypto exchanges for fungible and non-fungible tokens as well as decentralized organizations (DAO). However, it is still an open question to which extent Web3 solutions will replace existing structures.

13 toymint.co.

14 https://tse.bitnation.co/.

In the recent work of Bychkova and Kosmarski (2022) they consider how blockchain can affect in the fields of individual freedom, the consensus mechanisms, new methods for shared rules in a public ledger, and the transformation of common good with new political approaches. They concluded that blockchain can modify the *res publica* with more efficient processes to create new models of governance, some kind of a Democracy 2.0 (linked to the Web3).

Despite current projects, blockchain is an opportunity to build up new techno governance models, including the convergence of distributed ledger technologies with big data and artificial intelligence. The management of democracy procedures and eAdministration processes in complex scenarios needs new rules and cultural transformations. As Innerarity (2020: 339–348) pointed out, we need democratize Democracy creating a new cognitive infrastructure of Democracy; the blockchain can be used to do so.

We can affirm that there are some sociological properties of the blockchain that make it converge with any humanistic discipline and that they would be the following:

- Disintermediation. The blockchain makes it possible to dispense with intermediary or third-party entities that certify the content of the transactions or the data they store. In the field of documentation, education or philology, it allows the creation of networks of researchers and professionals who would not need to resort to third-party entities to validate or certify the contents, within a P2P[15] network.
- Immutability. This property has the security of the certainty that the data and processes created will endure over time. Their sealing and their verifiability is intrinsic to a blockchain network. In the legal field, in historical studies and in social sciences, it allows to determine with certainty when, how and where a known data flow was produced.
- Trust. From the social point of view, it is the property that weaves social ties and security in the management of institutions. Without a doubt, it is one of the properties of the blockchain that is helping to create new forms of governance, since being distributed on a peer-to-peer basis, the certainty that what is certified is unalterable and immutable, allows to create that trust in the stored data. and secured throughout the blockchain.
- Transparency. The blockchain has the ability for all transactions to be known throughout the network. From the point of view of political science and legal

15 Peer-to-peer (P2P) computing or networking is a distributed application architecture that partitions tasks or workloads between peers. Peers are equally privileged, equipotent participants in the application. It is a network of balanced and equal nodes.

science, all the processes in which this property is necessary to obtain the trust of the stakeholders involved could be reinforced. Not to mention how it could revolutionize the way scientific research is shared or the administrative processes of any institution in the world.

All these properties allow us to develop cooperative systems to innovate in the research fields of the humanities with new processes and new tools. The path of distributed registry technologies has barely begun, and in the course of this decade we will see how they mutate and generate new perspectives on their use. The possibilities that these offer have only just begun to be explored and therefore we understand that there are new fields of research that mix projects in the Humanities with the blockchain.

6 Towards a Disintermediated Governance: Adhocratic Relations and New Rules of Creation and Hybridization

As we remarked before, blockchain can be used as a way to decentralize and build new political frameworks. We are not advocating replacing democratic electoral systems with new and unknown decentralized structures. However, we do see an opportunity in the promise that a DAO has, to build new tools and processes that eliminate inefficiencies (that due to human error, omission or spurious interests) may occur.

Integrate in the eAdministration, public institutions or political parties new paradigms of interaction can develop a new common trust amongst citizens, politicians and public servants. This interaction can be constructed under the umbrella of the DLTs, with specific DAOs developed for it. That means we can automate with no human intervention a lot of processes and procedures. Por instance, in the eAdministration everything with tax collection processes, to reinforce the trust of citizens in local authorities we can use new e-Participation tools with tokenized models, and regarding political parties they can present themselves to elections with automated political programs triggered in an automatic DAO to be deployed automatically after the election of the institutional bodies where they must govern.

Of course, those actions mean that we have the opportunity to deploy a new governance model with new roles of all the stakeholders involved in this new framework. If we are talking about transparency, decentralization and disinter-

mediation, we are talking about adhocratic schemes of administration, creating new rules, roles, and actions. Blockchain can be a reactor not only for the digital transformation but for new democratic frameworks.

A new key concept: techno-governance hybridization. Actually in Political Sciences concepts as a hybrid regime is not a positive one. A hybrid regime is a mixed type of a political system that is often a result of an incomplete transition from an authoritarian regime to a democratic one, although there are regular elections, these regimes hold political repressions. So, we are defining a new context of polymorphic views about what we can consider a legitimate democracy. In our post pandemic context, we can observe a new kind of a wave of democracy,[16] emerging in the actual political context. On one side with the current situation of governments in Russia, Brazil or Hungary, but on the other side with the possibilities that new technologies such artificial intelligence and big data, together with blockchain offers a real digital transformation in public institutions to be more efficient, transparent, and cooperative.

Following Huntington's scheme, we could affirm that we are experiencing a new wave with two opposing directions. The one that is being directed by post-politics (disinformation, fake news, alternative facts, etc.) and the one that can develop a new type of more transparent governance, which is what we call techno-governance, thanks to the implication that the technologies of the 4th Industrial Revolution will have in the development of a new eDemocracy.

Building new processes and tools that allow the development of new and more resilient democratic scenarios, more participatory and more effective, is an opportunity that technologies such as blockchain offer us. Its ability to transform the channels of eParticipation (Benítez-Martínez et al. 2020), electronic voting (Holmes 2022), public contracting systems or the documentary certification of all eAdministration processes (Parenti et al. 2022), are already a reality to change the anti-democratic paths that some democracies are suffering in various countries of the world.

But disintermediation is not only a property that will impact political systems or the field of culture. Its ability to create new management systems, organize processes or develop new tools that allow resources and networks to be managed more effectively and efficiently, will be one of the fundamental premises of the impact of blockchain in multiple fields.

16 Wave of Democracy is a term that appeared in 1887 (Morse, "The Cause of Secession"), but popularized by Samuel P. Huntington in his article published in the Journal of Democracy and further explained in his 1991 book "The Third Wave: Democratization in the Late 20th Century". Democratization waves have been linked to sudden shifts in the distribution of power.

As we have seen before, this impact will create new fields of study, new formats and stories. The adhocratic capacity of the blockchain will allow the construction of new scenarios, following its sociological properties. And from these opportunities will arise to establish new epistemological paths in many disciplines of the Humanities and Social Sciences.

7 Blockchain and Humanities

As we have seen so far, the blockchain has great applications in the world of the arts or in the definition of new governance systems, from the point of view of political science or sociology, but there are many more fields where it can have a great impact. Let's see in some fields it can generate new possibilities.

Blockchain for Libraries

Regarding its properties, using blockchain in libraries can have a great impact in the digital preservation and tracking of books and digital copies with the tokenization of the digital assets – the cultural works – and the way as the community-based collections are shared.

The way that the inter library loan and actual voucher system can be shifted also, including a strong and updated verification of credentials via a dApp (instead the current library card).

Also we can organize the keeping of the corporate library records in a different manner, including the provenance and authenticity of valued items. And of course, all of these items allow data management (Frederick 2019) more efficiently and with no loss of information (or misinformation caused by a breach in data custody).

Blockchain for Scientific Publishing

Think about the opportunity to establish new models of scholarly publishing (and the way as this type of works are shown and recorded). With the new DLTs we can deploy easy tools for users with a low cost implementation, that will be an independently and verifiable method that could be widely and readily used to audit and confirm the reliability of scientific studies. This is because we treat information by creating a cryptographic hashing of every record (with plain text)

in a scientific work adding this on the blockchain. This creates a time-stamped record in the network, which other researchers can quickly verify in the future. No data of the document can be changed or altered from the stored record, so a verifiable network of scientists around the globe can store and preserve all this information in a P2P platform owned by researchers, librarians and academic institutions at the same time. With a framework like this no third parties are needed (publishers, for instance).

Therefore, this technology is the best fit for academia and gets a potential pace in libraries and authors, to connect them with its final users, and can be used to change the actual status quo within the scholars and publishers bringing the opportunity to the universities and authors to hold their own rights for their scientific contributions.

Blockchain in Museums and Archeology

The ability of the blockchain to be able to secure data and preserve it without the possibility of altering it has a great impact on museology and archaeology.

The possibility of being able to manage museum collections through a tokenized system that assigns the value of the data and the origin of each work or element of a museum is an intrinsic value for its chain of custody.

In addition to guaranteeing custody, by being secured in a blockchain network, the data will be more transparent than ever for any stakeholder that intervenes in the value chain of each of the pieces in custody.

Examples such as the one used in Indonesia with the Prabu Geusan Ulun Museum, using Hyperledger Composer (a certification system based on the Ethereum ecosystem) is a successful case study. To prove that the system worked, it was divided so that one part of it needed human intervention and the other part was fully automated to serve as a testing tool. (Hongo et al. 2021).

And in the field of archeology a based blockchain network can be used to assure the traceability of the archaeological remains from the field where they are collected to the laboratories where these items are going to be studied, dated and secured. Actually, there is not an ongoing project published in the major databases. But we are sure that blockchain will have a huge impact in this field regarding the promise of its properties.

8 Future Landscapes: Convergences

One of the great challenges that society faces is how the convergence of several of the technologies of the 4th Industrial Revolution will impact it. If we refer to Big Data, that is, the capacity that our society currently has to create, store and analyze huge amounts of data, the challenge is enormous.

Therefore, one of the main challenges is knowing how we are going to treat this data and how we are going to use it. The ethical, philosophical and political dimensions are a great challenge. In order to advance in solutions that restore confidence to citizens in how these data are treated and that they will not be used against them, the role of the blockchain is decisive, due to the properties it has.

In addition, being Big Data the fuel that feeds the tools and algorithms that are created through artificial intelligence, it is necessary to have technological tools that can guarantee us levels of security in the treatment and custody of data.

We are not only talking about a technological implication of this convergence between blockchain and AI, something that is being considered as the work of Ekramifard et al. (2020) pointed out. We are talking about the need in how the humanities, with philosophy at the forefront, have to help this convergence to occur in accordance with ethical principles that determine the regulatory framework, present and future.

The social and political implications of how we are going to work in this technological-humanistic convergence is decisive for the future of our societies. The trend towards dataism in our society cannot lead us to a datacism, where biases and programming failures can cause social or political gaps. Or what is worse, cause systemic failures that are difficult to repair or ignore.

We have to jointly build a “social algorithm” that is based on a clear (and safe) ethical framework. The challenges faced by the so-called Metaverse and the way in which decentralized digital identities are created (many of them based on AI engines, that is, avatars that will be a digital twin of us), pose an even greater challenge of how We must regulate and secure those artificial intelligences that will intercede for us, with an impact on a personal level that we are not yet able to fully imagine.

We have to be able to properly discern the advantages, benefits, threats and challenges that all these convergences entail. The power of Big Data and artificial intelligence is tremendous, but they need the blockchain and its ability to preserve, seal and secure data in a distributed way, so as not to cause a “dataclism”.

There is no clearly established path. Both blockchain and artificial intelligence capabilities are at an early stage to be democratized. That is, the possibility that they have a great social capillarity, between citizens and small businesses, so

that they do not cause social or economic gaps and are an engine of social and economic rebalancing.

9 Epilog: How Humanities can Help Blockchain?

In relation to what we have exposed in this work, we invert the order of the sentence in the title of this chapter and we consider how the humanities can help the blockchain.

The question is not trivial, since more than ever it is necessary to critically question the impact of technologies in our society, as we have already explained.

Indeed, the Humanities can (must) be the vector that does not allow the dehumanization of technology. In this sense, the intrinsic properties of the blockchain are of great contribution, due to the great sociological and political burden that its use entails.

Although the construction of the narrative of the blockchain policy is highly conditioned by the first blockchain network, bitcoin, and its impact on the creation of crypto finance, there is an economistic preconditioning of its philosophy. In the work of Golumbia (2016: 50–63) a very good approximation is made on this issue and how polarized this debate is in society.

But, we must go further and establish multidisciplinary and hybrid channels that allow us to develop a polyhedral conversation about the impact that the blockchain will have on society.

We cannot forget that this technology is very young (it was born in 2009) and that it is still developing its first steps. DLTs will have to adapt to provide solutions to new challenges and social and economic problems, and this implies constant adaptation and innovation. Beyond the need to create scalable and interoperable systems, there will be the need to analyze how problems are solved without creating new ones. The convergence here of sociology, political science, law, history, anthropology and the humanities in general with the blockchain seems to us to be something decisive for the future evolution of our society.

The challenge is enormous, and therefore this hybridization and generation of new contexts and meta-narratives between different fields of knowledge is more necessary than ever.

Bibliography

Benítez Martínez, F.L., Hurtado Torres, M.V. & Romero Frías, E. (2020): "The Tokenization of the eParticipation in Public Governance: An Opportunity to Hack Democracy", in Prieto, J., Das, A., Ferretti, S., Pinto, A., Corchado, J. (eds), *Blockchain and Applications. BLOCKCHAIN 2019. Advances in Intelligent Systems and Computing*, vol. 1010. Cham: Springer. doi: 10.1007/978-3-030-23813-1_14.

Benítez Martínez, F.L., Hurtado Torres, M.V. & Romero Frías, E. (2020b): "A neural blockchain for a tokenizable e-Participation model", in *Neurocomputing*, 423, pp. 703–712. doi: 10.1016/j.neucom.2020.03.116.

Benítez-Martínez, F.L., Romero-Frías, E., & Hurtado-Torres, M.V. (2022): "Neural blockchain technology for a new anticorruption token: towards a novel governance model", in *Journal of Information Technology & Politics*. doi: 10.1080/19331681.2022.2027317.

Brown, Allen E., and Gerald G. Grant. (2005): "Framing the frameworks: A review of IT governance research", in *Communications of the Association for Information Systems*, 15, 38.

Chen, W., Zhang, T., Chen, Z. et al. (2020): "Traveling the token world: A graph analysis of Ethereum ERC20 token ecosystem", in *WWW'20: Proceedings of The Web Conference 2020*, pp. 1411–1421. doi: 10.1145/3366423.3380215.

Di Angelo, M., and Salzer, G. (2020): "Tokens, Types, and Standards: Identification and Utilization in Ethereum", in *2020 IEEE International Conference on Decentralized Applications and Infrastructures (DAPPS)*, pp. 1–10. doi: 10.1109/DAPPS49028.2020.00001.

Ekramifard, A., Amintoosi, H., Seno, A.H., Dehghantanha, A. & Parizi, R.M. (2020): "A Systematic Literature Review of Integration of Blockchain and Artificial Intelligence". In Choo, KK., Dehghantanha, A., Parizi, R. (eds), *Blockchain Cybersecurity, Trust and Privacy. Advances in Information Security*, 79. Cham: Springer. doi:10.1007/978-3-030-38181-3_8.

Entriken, W., Shirley, D., Evans, E. & Sachs, N. (2018): "ERC-721 non-fungible token standard". <https://eips.ethereum.org/EIPS/eip-721> (03-03-2022).

Frederick, D.E. (2019): "Blockchain, libraries and the data deluge", in *Library Hi Tech News*, 36, pp. 1–7. doi: 10.1108/LHTN-09-2019-0059.

Golumbia, D. (2016): *The Politics of Bitcoin. Software as Right-Wing Extremism*. Minneapolis: University of Minnesota.

Haber, S., and Stornetta, W.S. (1991): "How to time-stamp a digital document", in Journal of Cryptology 3, pp. 99–111. https://doi.org/10.1007/BF00196791.

Hongo, S., Sudrajat, R., and Yulita, I.N (2021): "Design of Museum Historical Heritage Management System Using Blockchain Digital Certificate and Hyperledger Composer". *International Conference on Artificial Intelligence and Big Data Analytics*, 2021, pp. 231–235. doi: 10.1109/ICAIBDA53487.2021.9689733.

Innerarity, D. (2020): *Una teoría de la democracia compleja. Gobernar en el siglo XXI. III Democratizar la Democracia*. Barcelona: Galaxia Gutenberg.

Holmes, C. (2022): "How Blockchain Technology could support Democracy and E-Voting", in *The Journal of The British Blockchain Association*, 33960.

Nadini, M., Alessandretti, L., Di Giacinto, F. et al. (2021): "Mapping the NFT revolution: market trends, trade networks, and visual features", in *Sci Rep*, 11, 20902. doi: 10.1038/s41598-021-00053-8.

Parenti, C., Noori, N., and Janssen M. (2022): "A Smart Governance Diffusion Model for Blockchain as an Anti-corruption Tool in Smart Cities", in *Journal of Smart Cities and Society*, 1, pp. 71–92. doi: 10.3233/SCS-210122.

Rauchs, Michel, et al. (2018): *Distributed ledger technology systems: A conceptual framework*. Cambridge: University of Cambridge.

Rehman, W., Zainab, H., Imran, J. & Bawany, N. Z. (2021): "NFTs: Applications and Challenges", in *2021 22nd International Arab Conference on Information Technology (ACIT)*, IEEE. pp. 1–7.

Ritchey, T. (2005): "Problem structuring using computer-aided morphological analysis", in *Journal of the Operational Research Society*, 57, pp. 792–801. doi: 10.1057/palgrave.jors.2602177.

Romero-Frías, Esteban, et al. (2021): "Fundamentals of Blockchain and New Generations of Distributed Ledger Technologies. Circular Economy Case Uses in Spain", in Muthu, S.S. (ed.), *Blockchain Technologies for Sustainability. Environmental Footprints and Eco-design of Products and Processes*. Singapore: Springer. doi: 10.1007/978-981-16-6301-7_2.

Swan, Melanie, and Primavera De Filippi. (2017): "Toward a philosophy of blockchain: a symposium: introduction", in *Metaphilosophy*, 48, pp. 603–619.

Vasan, K., Janosov, M. & Barabási, AL. (2022): "Quantifying NFT-driven networks in crypto art", in *Sci Rep*, 12, 2769. doi: 10.1038/s41598-022-05146-6

Miguel Calderón Campos

Spanish Corpora: Big (Quality) Data?

1 Introduction

In Linguistics, reference to Big Data entails the reference to corpora and to their ensuing size, type, representativeness and sample selection. Figure 1 shows the tendency towards bigger and bigger Spanish corpora, from the early RAE projects of over 100-million words (CREA) to the macro-corpora of project *TenTenCorpora* aiming at over 10,000 million words. In the latter, the Spanish corpus, *EsTenTen18*, is close to 17,000 million words.

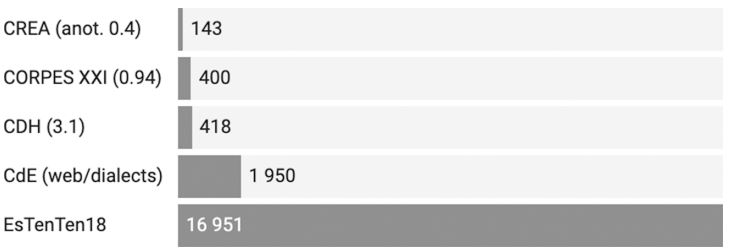


Figure 1: Spanish corpus size in millions of tokens.

The sizes shown in Figure 1 might give the impression that these resources are already beyond the minimum necessary for the exhaustive description of any question in Linguistics. Yet, the endless universe of the web and of social networks is still searched for new data, as if the big size of corpora were not enough for the description of some words' constructional or diachronic, stylistic or social variation profile. Equally paradoxically, small corpora are built more and more frequently to fill the gaps left by bigger, general corpora.

Computational Linguistics thus currently works on three fronts: the compilation of macro-corpora reference corpora, the annotation of highly specific small

Note: This work is part of the ALEA XVIII Project, funded by FEDER/Junta de Andalucía-Consejería de Transformación Económica, Industria, Conocimiento y Universidades/Reference Project P18.FR.695. It is also part of the ALEA XVIII. Oriental, financed by FEDER/Junta de Andalucía-Consejería de Transformación Económica, Industria, Conocimiento y Universidades, reference project A-HUM-116-UGR20.

https://doi.org/10.1515/9783110753523-008

corpora, and the improvement of traditional Corpus Linguistics by means of the analysis of massive internet and social network data.

This paper is an overview of Spanish Corpus Linguistics. Section 2 reviews the synchronic and diachronic corpora available and points out limitations imposed by source quality and by the interfaces used (in general, RAE corpora offer better data selection and achieve a higher representativeness, whereas non-RAE corpora use more flexible and powerful search engines, as shown in section 2.1). Based on the analysis of the Colombian Spanish token *parce*, section 2.2 shows that inaccurate search results are closely related to low quality samples and geographic metadata. Section 3 uses massive corpora, internet data and social network data for improved results on the little evidence of the quantifier *algotro* ('some other') available in RAE corpora. Finally, section 4 compares Big Data sources with two specific diachronic corpora: *Post Scriptum* (P.S.) and *Oralia diacrónica del español* (ODE).

2 Spanish Reference Corpora and Massive Corpora

General corpora or reference corpora are corpora intended for the attestation of general properties of a language at a given period of its history. For Spanish, a general or reference corpus must contain all types of texts, of all the periods into which the timeframe intended for research can be divided, and from all the countries where Spanish is spoken as a first language.

The *Corpus del Español del siglo XXI* (CORPES XXI) and the *Corpus del Español* (CdE web/dialects) are the two commonly acknowledged reference corpora of contemporary Spanish. The *Corpus del Diccionario Histórico de la Lengua Española* (CDH) and the historical subcorpus of the *CdE* (CdE hist) are diachronic Spanish reference corpora. The basic properties of all four corpora are outlined in Table 1 below.

The latest versions of the two RAE corpora, CORPES XXI and CDH, amount to ca. 400 million words. The former contains samples produced since 2001 and is intended to increase by ca. 25 million words per year. Transcripts of spoken samples amount to 1%, some linked to audio files. The intended variety proportion is ca. 30% European Spanish and 70% American Spanish.

The CDH includes the samples of the first RAE corpora, CREA and CORDE, after descriptive annotation (lemmatization and morphosyntactic labelling), similarly to CORPES XXI. Unlike the four major types of samples in CORPES XXI (fiction, non-fiction, press, spoken), the samples of the CDH are classified by topic

Table 1: The Spanish reference corpora.

	Tokens (by million)	Spain	America	Period	Fiction	Non-fiction	Press	Spoken
CORPES XXI (0.94)	400	35%	65%	2001–21	28%	21%	47%	1%
CDH (3.1)	418	71%	29%	12th c.-2005	classified by topic			
CdE (web/dialects)	1950	22%	78%	2013–14	Blog (53%) / General (47%)			
CdE (hist)	100	Data not available		13th c.-20th c.	20th c.: 25%	25%	25%	25%

(i.e. arts, social sciences, science and technology, leisure and everyday life, politics and economy, and health).

The CDH can be divided into three subcorpora, each of which can be accessed separately: i) the CDH core subcorpus (*CDH nuclear*) is a 63-million-word representative collection of samples taken from the CORDE and CREA; ii) the CDH XII-1975 subcorpus is a 230-million-word collection of most of the contents of the old CORDE corpus; and iii) the CDH 1975–2000 subcorpus is a 125-million-word collection of the CREA contents not included in the CDH core subcorpus. The proportion of European vs. American Spanish for the period from 1492 onwards in the CDH is 71% vs. 29% respectively.

The *CdE web/dialects* corpus is a reference macro-corpus (nearly 2,000 million words) of web samples of the period 2013 and 2014. It is arranged as two large sets (blogs vs. general) and is representative of the 21 Spanish-speaking countries.[1] The CdE's historical subcorpus contains samples from the 13th c. to the 20th c. Query results can be sorted by century and, for the samples of the 20th c., also by sample type (note that the 20 million words of the 20th c. are evenly distributed over the four sample types shown in Table 1).

At 16,951 million words, *EsTenTen18* is the biggest among the so-called massive corpora of Spanish. The samples were extracted automatically from internet sources and can be searched with Sketch Engine. Structured by subdomains (European Spanish domain.es, Mexican domain.mx, Chilean domain.cl, etc.), it allows to combine searches by descriptive and geographic data (see section 2.2).

1 21 countries including the United States, 22 including Equatorial Guinea.

2.1 Sample Quality vs. Interface Versatility

The main difference between the above corpora runs along the lines of Mair's (2006) contrast between 'big and messy' corpora vs. 'small and tidy corpora': the bigger the corpus, the lower the quality, the representativeness, and the accuracy of sample classification and annotation (both descriptive and presentational); by contrast, smaller corpora lend themselves to manual annotation and, therefore, achieve comparatively better sample selection and higher annotation accuracy.

RAE corpora are annotated and lemmatized remarkably accurately. Also, their samples are selected according to representativeness and are annotated with more accurate geographic, chronological and thematic metadata than non-RAE corpora (Rojo 2010). By contrast, non-RAE corpora rely on a more flexible and powerful search interface than RAE corpora, and count on bigger sizes: compared with CORPES XXI, CdE web/dialects is five times as big, and *EsTenTen18* is nearly fifty times as big.

While the quality of CORPES XXI's samples is praised on the CdE's website, it is also stated that '[. . .] it uses a fairly rudimentary web interface, which really limits what can be done with concordances, collocates, and frequency lists. In other words, the good textual data is "trapped" behind a poor interface, and is inaccessible to end users'.[2]

EsTenTen18 is praised for its size, for the collocate-based 'word sketches' and for the possibility to submit queries with CQP. By contrast, it is criticised for the poor lemmatization and for the amount of wrong or inaccurate annotation. Indeed, *EsTenTen18* becomes unbeatable for its size and for its powerful, user-friendly interface, when it comes to finding the combination profile (word sketch) of highly frequent words. Graphical representations of a given token's profile are easily generated, as in the adjective *severo* 'severe' of Figure 2. Remarkably, the same figure exposes one of the main shortcomings of this type of macro-corpora too, namely their poor morphosyntactic annotation: funny enough, the most frequent collocate for the adjective *severo* is Spanish Nobel prize winner's surname *Ochoa* (thus, *Severo Ochoa*).[3]

CdE web/dialects stands out for the possibility to research dialectal differences across the 21 Spanish-speaking countries. Thus, a single query for the adjectival suffix *-oso* returns Argentinian Spanish adjectives like *ochentoso* 'eighty-like', *noventoso* 'ninety-like', *criterioso* 'sensible', *modernoso* 'modern' or *culposo* 'guilty' vs. Euro-

2 https://www.corpusdelespanol.org/compare.asp (17–12-2021).

3 CdH yields the same wrong annotation. Wrong annotation can be revised only manually.

Figure 2: A graphical representation of the collocates of the adjective *severo* 'severe' in *EsTenTen18*.

pean Spanish adjectives like *lioso* 'messy', *cantoso* 'flagrant', *picajoso* 'fussy', *pasteloso* 'soppy' or *patoso* 'clumsy'.

The option *Chart* allows to obtain a very telling overview of well-attested general usage. Thus, the query "re _J*"[4] yields a chart comparing the normalized frequency of "*re*+adjective" in all the Spanish-speaking countries, and significant contrasts can be noticed: the highest frequencies occur in the varieties of Argentina (17.94 per million words), Chile (8.06 wpm) and Paraguay (5.58 wpm). Frequencies below 3.20 wpm (Mexico) are attested in the remaining varieties. The adverbial counterpart with *re-* (e.g. *rebién* 'very well', *remal* 'very bad', *retarde* 'very late', etc.) shows a similar distribution across varieties: Argentina attests 3.10 wpm, Chile 1.56 wpm and Uruguay 1.26 wpm. Guatemala attests a similar result as the south American countries: 1.07 wpm.

RAE corpora do not rely on search engines capable of rendering visual results as in Figure 2. CORPES XXI and CDH present quantitative results as absolute and relative frequencies by country. Surprisingly, the pie charts generated automatically only give results of absolute frequencies, and this may severely distort the picture. For example, the well-known American Spanish preference for *computadora* 'computer' vs. European Spanish *ordenador* 'computer', is confirmed by the

4 I.e. *re-* prefixed to an adjective for intensification, e.g. *rebueno* 'very good', *relindo* 'very nice', *reloco* 'very crazy', etc., NGLE 10.9j.

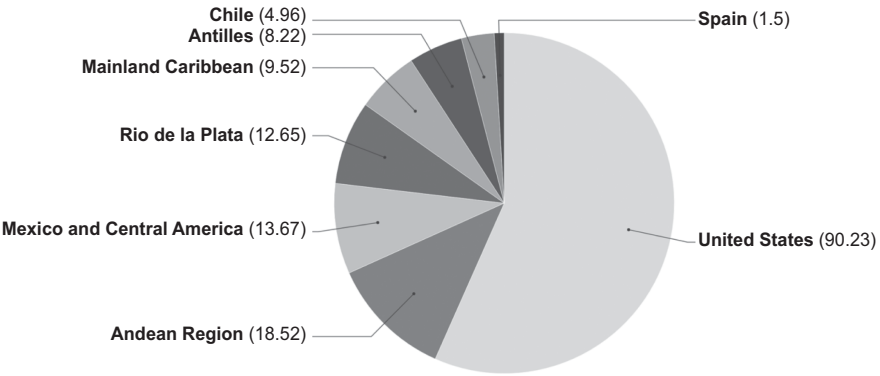


Figure 3: The frequency of *computadora* 'computer' in the CDH (wpm values).

CDH data: at 1.5 wpm, the relative frequency of *computadora* in European Spanish ranks lowest among the Spanish-speaking countries (cf. Figure 3, generated by the author, based on the CDH's wpm frequencies for this query).

Contrarily, based on absolute frequencies, the CDH's graphical representation (see Figure 4),[5] stands in sharp contrast with Figure 3 above, and is therefore misleading: as European Spanish amounts to 71% of the samples in the CDH, the absolute frequency of *computadora* for European Spanish (402 occurrences) is the highest in the corpus. This is a serious weakness of the concordancer's data management, and also one that could be easily overcome by linking pie chart generation to wpm frequencies instead of to absolute values.

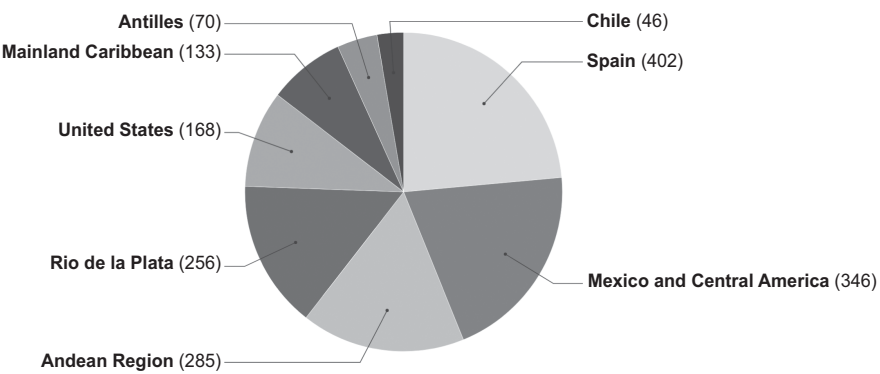


Figure 4: The frequency of *computadora* 'computer' in the CDH (absolute values).

5 This figure, generated by the author, is a copy of the figure generated by the CDH.

The lower flexibility of the interface is compensated in RAE corpora with a better sample selection and with more accurate descriptive (linguistic) and non-descriptive (presentational) annotation. CORPES XXI and CDH therefore allow a much more precise chronological and geographic identification of language data. Take, for example, the adjective *severo* again, originally used in Spanish to mean 'severe' (as in *castigo severo* 'severe punishment' or *crítica severa* 'severe criticism', etc.). A later development under the influence of English extended the semantic range to mean severity of illness (*depresión severa* 'severe depression', *discapacidad severa* 'severe handicap', *traumatismo severo* 'severe injury' etc.). Neither the CdE corpus nor the *EsTenTen18* corpus identify the earliest record of *depresión severa*, which CDH attests earliest in a sample of Venezuelan Spanish of 1976 (*sentimientos de culpabilidad y lo suficientemente severos* 'guilt feelings severe enough'). Something similar applies to the collocation *traumatismo severo*, first attested in Argentinian Spanish in 1988.

2.2 Precision and Recall. Corpus Evidence on Colombian Spanish *Parce*

'Precision and recall' are defined by Stefanowitsch (2020: 111–116) according to sample quality and annotation accuracy. Data retrieval is accurate whenever a query returns only exact matches. Thus, research on imperative verb forms ending in *-lde* (*dezilde* 'you tell him', *dalde* 'you give him', *enbialde* 'you send (to) him', etc.) in a non-annotated corpus of the 16th c. will retrieve both imperative forms and false positives, the latter as a result of the retrieval of nouns with the same ending, e.g. *alcalde* 'mayor', *balde* 'bucket' or *molde* 'cast'.

Exhaustive data retrieval ('recall') is achieved whenever every possible match is retrieved. This is especially difficult to attain in historical research, for the many orthographic variants that a token may display. Thus, the following variants are attested for the token *trébedes* 'trivet' in the ODE, some of which are quite unpredictable: *trevedes*, *trebedes*, *treuedes*, *treodes*, *trévedes*, *trebes*, *estrebes*, *esttreores*, *extrevedes*, *estrebedes*. These forms cannot be retrieved under the same query and are thus a major source of data loss during data retrieval (as 'false negatives').

This section assesses the degree of precision and recall of CORPES XXI, CdE and *esTenTen18* according to their samples and linguistic annotation. The source of the samples of RAE corpora is mainly publications, including revised editions. This reduces to a minimum the amount of typographical mistakes and inconsistencies, in contrast with corpora built with samples collected from blogs and non-institutional websites. This can be illustrated with the Colombian addressing term *parce* 'friend, pal'. A shortened form for *parcero*, it is used among the younger

speakers as an addressing term to express comradeship or conviviality. The term comes from Portuguese *parceiro* 'friend, pal'. It was allegedly used first in the lower class quarters of Medellín in the 1980s, and then spread over the rest of the country (Castañeda Naranjo 2005: 67).[6]

CORPES XXI contains enough evidence to describe the usage or the geographic distribution of *parce*: out of the 49 instances recorded in the corpus, only 8 are not from Colombian Spanish (4 are typographical mistakes[7] or foreign words[8] used in European Spanish samples; the other 4 are the vocative form used by Colombian Spanish speakers in literary works or journal articles).[9] The evidence available in this RAE corpus thus confirms that *parce* is associated with Colombian Spanish, and illustrates not just its combinatory possibilities and its meaning (1), but also its origin (1) and its chronological development (2):

(1) Parce!!! (de *parcero*, que en Colombia es amigo) Hermano!!! («Miguel Bosé se ofreció a mediar con las FARC al recibir nacionalidad colombiana». El Comercio. pe. Lima: elcomercio.pe, 2010-03-17).

(2) Es 1994, todavía son pocos los que dicen *parce* (Castro, Samuel: *A la velocidad del byte*. Medellín: Fondo Editorial Universidad EAFIT, 2008).

Parce is nearly always used as a vocative, before or after a pause (3–6). It is also recorded as a noun meaning 'friend, pal' ("se trataba de un parce de ellos"). It is often used with the pronoun *usted* ('you [formal]'), except for one example with *vos* ('you [informal]' 6) and another with *sumercé* 'you [formal]' (5).

(3) —Parces, ¿alguno de ustedes tiene algo para la cabeza? (Martínez, Fabio: «Los ensayistas del Parque del Perro». *El escritor y la bailarina*. Cali: Escuela de Estudios Literarios de la Universidad del Valle, 2012).

6 The earliest attestation in the CDH dates back to 1994: "Un ejemplo: ¿Entonces qué, parce, vientos o maletas? ¿Qué dijo? Dijo: Hola hijo de puta. Es un saludo de rufianes" (Vallejo, Fernando, La virgen de los sicarios [Colombia] [Santafé de Bogotá, Alfaguara, 1999).

7 *Parce* for *parece*: "parce que van dejando . . .".

8 The Latin formula "Parce nobis, Domine", or the French causal conjunction "parce que" 'because': "Hay una frase recurrente durante la película: parce que moi je rêve . . .".

9 The Mexican example is by a Colombian character in a play ("Cuántos años tenemos de parces, de amigos"). Two Ecuadorian examples are a news article about Colombian hit men ("acá lo cogemos, parce, y le damos paila"). The Bolivian example refers to Colombian singer Juanes' album *P.A.R.C.E.*

(4) – Si su mujer le puso los cuernos, parce, yo no tengo la culpa, la culpa la tiene usted (López, Andrés; Ferrand, Juan Camilo: *Las muñecas de los narcos*. Madrid: Aguilar, 2010).

(5) -Hum, parce, sumercé anda desactualizado (Álvarez, Juan: *C.M. no récord*. Bogotá: Alfaguara, 2011).

(6) —¿Querés, parce? (Franco, Jorge: *El cielo a tiros*. Bogotá: Penguin Random House Grupo Editorial, 2019).

The dialectal distribution of *parce* according to the chart based on CdE data runs against the data available from CORPES XXI, where the vocative is recorded in other varieties of Spanish too: Colombian (2.26 wpm), Salvadoran (2.03 wpm), Ecuadorian (0.88 wpm), Costa Rican (0.78 wpm) and Panamanian (0.67).

The quality of these varieties is, however, low. The use of the addressing term *parce* is well attested in the concordances of Colombian Spanish in the CdE,[10] even if it is fraught with false positives as a result of typographical mistakes. This is not always the case in the other subcorpora: all the occurrences in Salvadoran Spanish are a mistaken form for *parece* (“parce cada día más vacía”, “me parce muy interesante el comentario”, etc.); in Ecuadorian Spanish, 21 occurrences are for the name *Patricio Parces*; Panamanian Spanish contains 15 occurrences, 5 of which are typographical mistakes and the remaining 10 are vocatives but do not really evidence actual use in this variety: 2 occurrences come from a Colombian website (colombiatvglog.com), 4 are from a staged interview with a footballer from Barranquilla (Colombia), and the remaining 4 are comments on the Colombian TV series *El cartel de los sapos*.

The results available from *EsTenTen18* are unreliable too: at 4721 occurrences, *parce* has a frequency of 0.24 wpm, but most are typographical mistakes. Even more, only 66 occurrences of *parce* out of 217 in the Colombian section (.co) are vocatives. This means that, as the nominal form *parce* is virtually confined to Colombian Spanish, the true positives out of the original 4721 occurrences in the corpus must amount to slightly over 66.

Typographical mistakes mislead annotation and lemmatization to the extent that a high degree of inconsistency can be noticed: *parce* ‘parece’ is sometimes annotated rightly as *VMIP3S0* (i.e. the third person singular of the present tense, indicative mood) but is wrongly ascribed to the lemma *parce*; the opposite, i.e.

10 E.g. “parce, vos tenés que callarte”; “Buenos días, parce, hágame un favor”; “mis parces no se pierden ni un capítulo”; “quiubo, parce”; “vamos palante, parce, sintetiza el taxista”; “¿Parce, y la pasaste bien? Sí, güevón, super chimba”.

parce 'friend, pal' annotated as *VMIP3S0* ("decir parce, en vez de parcero"), is also recorded; some other times, both *parce* 'parece' and *Parce* 'friend, pal' are annotated as *NP* (proper noun), especially if the initial is upper case.[11] The precision of the corpus is, thus, remarkably low and, while it does not make it impossible to research specific cases in detail, data processing becomes significantly more demanding.

False negatives or misses (i.e. "fail[ure] to include instances of our phenomenon" Stefanowitsch 2020: 111) are a different case. In the example under study here, data may be missed, if the spelling associated with the realization of /θ/ in *parce* as /s/ (so-called *seseo*) were not considered. Lemmatization of the vocative does not attest such spelling, so additional queries are necessary for the form *parse* and its plural *parses*.

As in other examples described above, most of the instances retrieved are false positives: the technical term *parse* (meaning 'syntactic analysis') prevails in *EsTenTen18*,[12] and *parse* as a typographical mistake for *parte* 'part' ("parse integrante") distorts the frequency in the Puerto Rican subcorpus of CdE. The only relevant occurrences for this query are ca. 20 concordances taken from a blog about rock music where the author imitates colloquial speech ("eyos escuchan salsa y esa muciquita de regetoneros, parse, que paila que no aya tenido padres metaleros" (rockombia.com, CdE).

The above is intended to show how low data quality may lead to low quality query results and the latter, in turn, to wrong conclusions, e.g. if automatically-generated charts are taken at their face value, i.e. without concordance analysis. Awareness of the strengths and weaknesses of each corpus, i.e. of "the nature and composition of the corpus used" and "the kinds of linguistic information provided by automatic tools" is thus essential (Egbert, Larsson and Biber 2020: 1).

11 *Parce* is annotated as *NP* (*Nombre Propio* 'proper noun') in both "Parce, si usted puede" and "Parce 'parece' el problema de Linux".

12 This is clearly as a result of automatic data collection from computing blogs, which are of little interest for a general corpus of Spanish; even so, some useful concordances can be retrieved: "–No se me ahogue más en alcohol, parse; ya deje de chupar" (foroactivo.com, *EsTenTen18*); "así que pues le dejo ese consejito, parse alivien no se ponga a hacer afirmaciones tan absurdas" (prometec.net, *EsTenTen18*).

3 Beyond Corpora: The Web and the Social Networks

Octavio de Toledo y Huerta (2016) relied on systematic data gathering from online resources (Google Books, Google Scholar, and Google's search engine) to complete the insufficient lexicographical data and the little evidence of *algotro* 'some other' (from 'algún otro') available in RAE corpora. Additionally, he relied on the general archive of the *Real Academia de la Lengua Española* and on oral corpora (COSER). The data collected allowed to attest the origin of the abovementioned indefinite quantifier in Extremadura rather than in Andalusia. The data also allowed to identify the current distribution areas, namely El Salvador, Colombia, Mexico, Honduras, Guatemala, Argentina, Chile, Ecuador, Panama, Costa Rica and Peru (in order of decreasing frequency).

This section reviews the data collected by Octavio de Toledo on the reliability of CdE and *EsTenTen18* as regards research on low-frequency lemmas in RAE corpora. The value of additional evidence of *algotro* gathered from Twitter is then pondered as a qualification of the abovementioned corpus data.

The number of occurrences of *algotro* in RAE corpora is low but representative: 9 occurrences in the CDH between 1896 and 1954,[13] and 2 occurrences of Salvadoran Spanish in CORPES XXI. Figure 5 shows the wpm frequency of *algotro* in the *CdE web/dialects*. According to this figure, the quantifier's distribution by vari-

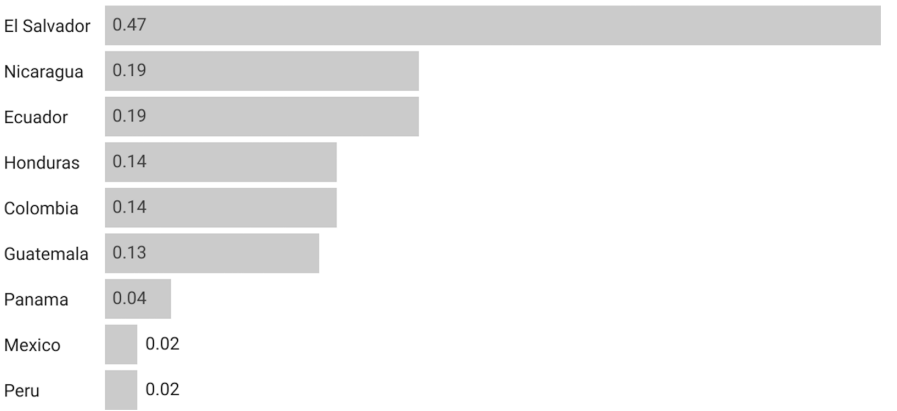


Figure 5: Wpm frequency of *algotro* in the CdE.

13 Of these, 4 are from Colombia, 2 from Honduras, 2 from Guatemala and 1 from Spain, specifically from Felipe Trigo's novel *Jarrapellejos* (1914), set in a village in Extremadura.

ety is close to Octavio de Toledo's claim, i.e. it is used mainly in Central America (El Salvador, Nicaragua, Honduras, Guatemala, and Panama), Colombia, Ecuador and, less frequently, in Mexico and Peru.

EsTenTen contains 132 occurrences of *algotro*, and the wpm frequency is therefore very low: 0.01. 50 concordances of *algotro* can be referred to American sources, 9 to Spanish, and the remaining 73 come from generic websites that cannot be ascribed to a specific variety. Among the above, the 50 American occurrences are distributed very much as described in the former paragraph:

– 5 from El Salvador (0.27 wpm),
– 1 from Guatemala (0.19 wpm),
– 2 from Honduras (0.15 wpm),
– 15 from México (0.1 wpm),
– 9 from Argentina (0.1 wpm),
– 8 from Chile (0.1 wpm),
– 6 from Colombia (0.1 wpm), and
– 4 from Nicaragua (0.1 wpm).

Twitter data reveal facts about *algotro* that are not evidenced by the above sources. The first fifty tweets containing the lemma *algotro* disclose the following distribution by country:

– Honduras (19 occurrences),
– El Salvador (10 occurrences),
– Colombia (8 occurrences),
– Mexico (6 occurrences),
– Guatemala (4 occurrences),
– España (2 occurrences), and
– Argentina (1 occurrence).

The most significant finding is that half the concordances are negative comments on the use of this quantifier. This is especially so in Hondurean Spanish, where 14 out of 19 tweets disapprove the use of this indefinite quantifier:

(7) Feliz día del idioma español . . . menos a los que dicen "haiga" "algotro" "embeces" . . . etc . . . no a ellos no! (Honduras).

(8) En una clase de la U un compañero exponiendo comete el terrible horror de decir algotro y la catedrática le hizo unos ojos que lo quemó y a todos nos quitó puntos por ese error, vieja cabrona (Honduras).

(9) ¿Qué flores se le compra a una dama que dice "haiga" y "algotro"? –Cilantro (Honduras).

(10) Valoro la creatividad de unir “algún otro” en “algotro”, pero no. No, por favor (El Salvador).

(11) Le dice mi esposa a mi hija de 3 años: –Me sorprende oírte decir la palabra algotro ya que es una palabra que cayó en desuso (México).

Twitter evidence stands out in other respects too. Remarkably, one of the examples of European Spanish confirms the current usage of this lemma in Extremadura (“Nos encanta “algotro”, que aún se usa en Extremadura, aunque el DLE no lo diga”). Otherwise, a tweet by a Mexican speaker illustrates the diastratic popular mark of *algotro* in Mexico:

(12) “Algotro lo tengo registrado en una de las entrevistas de mi tesis, de una mujer, de 20 y tantos, con estudios básicos, nacida y crecida en San Felipe, Guanajuato, México”.

Overall, the data available for *algotro* reveal the need for exhaustive procedures in low-frequency lemmas: RAE corpora are a reasonable starting point in that they supply fairly reliable chronological and geographic data. Three further sources may be used for additional evidence: i) CdE and *EsTenTen18* data, ii) Google searches, and iii) Twitter data. The resulting body of data allows the identification of the dialectal, combinatory and sociolinguistic profile of low-frequency lemmas.

4 Small Specific Corpora in the World of Big Data

Besides Big Data sources, small, specific corpora may widen the research data sources quite substantially. Specific corpora or complementary corpora are compiled according to a sample selection whereby the sources must share a given property that is relevant to the research objectives (Rojo 2021: 75). Thus, the sample may be by a given author, of a given literary or musical genre, of a given field of science, of a given period, etc.

Various specific corpora of Spanish are currently being compiled: diachronic corpora, like *Biblia Medieval*, CHARTA, CORDIAM, COREECOM, *CorLexIn*, etc., and spoken corpora, like COSER, ESLORA or PRESEEA. This section reviews two historical corpora managed with TEITOK both for language processing and for data selection and retrieval: *Post Scriptum* (Vaamonde 2017, 2018; Janssen and Vaamonde 2020) and *Oralia diacrónica del español*, ODE (Calderón Campos & Vaamonde 2020).

P.S. is a corpus of private correspondence of the Modern Period (1500–1833). It contains two million words distributed over two surcorpora: one for Portu-

guese and one for mainland Spanish. ODE is a corpus of handwritten documents of the 16th c. to the 19th c. Unlike the P.S. corpus, compilation of the ODE corpus is currently underway. It covers two sample types: i) witness statements at trials, and ii) inventories of personal belongings. The target size of the ODE corpus is one million words, and the original scope of sources has been widened from the historical kingdom of Granada (today's provinces of Granada, Almeria, and Malaga) to the rest of Andalusia plus Extremadura and Madrid. The two corpora allow simple retrieval as facsimiles, as palaeographic samples, and as modern text. CQL searches and result mapping are also available.

At one and two million words respectively, these specific corpora are intended to overcome the dialectal and/or stylistic limitations of the bigger reference historical corpora available of over 400 million words. Their purpose is, therefore, to supply corpus evidence for research on historical dialectology or pragmatics that is otherwise unavailable from larger reference corpora (Calderón Campos & Díaz Bravo 2021).

Regarding dialectal variation, reference corpora limit themselves to the 21 or 22 Spanish-speaking countries (cf. note 1). These corpora allow retrieval of specific usage in European Spanish (e.g. *mogollón* 'a lot', *comerse un marrón* 'to own up to something', *pasteloso* 'cheesy', etc.), Chilean Spanish (*fome* 'boring', *pololo* 'boyfriend', *erís* '[you.SG] are', etc.), or Colombian Spanish (*sumercé* 'you [formal]', *chimba* 'cool, nice', *parce* 'pal', etc.), but not within their regional or local varieties.

Regarding diaphasic or stylistic variation, reference historical corpora rely mainly on formal language, e.g. literature, historical prose, essays, and scientific and legal texts. Informal spoken language is barely represented in the corpora, especially for the period before the 19th c. As a way of example, *vos* 'you' is recorded 668 times in the CDH core subcorpus (European Spanish, 19th c.), most of them from samples of historical novels. Occurrences can be found in other genres too, e.g. 17 occurrences in romance novels like *Eumenia o la madrileña*. Precisely, example 13 illustrates the literary style of this genre, pompous ("tributó lágrimas a los quebrantos de Eumenia") and archaic (as evidenced by the use of *vos* 'you' as an addressing form), but barely representative of informal Spanish of the 19th c.[14] and of addressing terms:

14 Except for what regards the author's *laísmo*, i.e. the use of the feminine form of the pronoun *la* 'her' for the masculine or neuter *lo* 'him' or 'it', or for the gender-unspecified form *le* 'to him/her/it'.

(13) Tributó esta muger amable algunas lágrimas a los quebrantos de Eumenia, diciéndola: "Vos habéis sufrido mil penas, hija mía; lloráis aún la ausencia de un esposo, pero ¿qué sería si os hubiera abandonado antes de serlo, después de sediciros y deshonraros?" (1805, Zavala y Zamora, Gaspar, *La Eumenia o la madrileña*, teatro moral).

By contrast, the samples collected for P.S. and ODE are substantially different from those of reference corpora: not only are they more representative of spoken language, but they have also been transcribed according to the original spelling and thus make available data that would have been missed, if the present-day counterpart of the original samples had been used.

(14) Dijo a uisto y reconozido a la persona de Manuel Rodriges vezino de este dicho lugar, la que reconozida, le hallo vna herida en el vrazo disquierdo en la parte alta de el molleo, echa con instrumento cortante y punzante, como nabaja o cuchillo, y por los accidentes que pueden acadezer, tiene peligro de muerte (ARCHGR, Serie de pleitos, 5233/022, 1753, Cúllar Vega, Granada, ODE).

Example 14 shows how intervocalic *d* was frequently lost in the Spanish spoken in Granada in the 18th c.: *molleo* (referred to an arm) actually meant *'el molledo o bíceps'* 'the lean muscle or biceps' after *-d-* elision. Later hypercorrection is even more significant, as *-d-* was inserted between vowels, as in *acadecer* (for *acaecer* 'happen'). Neither *molleo* nor *acadecer* are recorded in the CDH, whereas 105 occurrences of the full form *molledo* are attested.

The samples compiled for the ODE were selected according to their value as evidence of informal, spoken language, and for the best possible exemplification of the language spoken (and pronounced) in Granada in the Modern Period. Similarly, P.S. contains transcripts of private correspondence, so the language spoken in mainland Spain in the same period can be analyzed:

(15) thio mio con la ocasion de hallarnos muy apurados de dinero ni tener donde cobrar por aber puestole a Dn Balthasar un pleitto las monjas de la conzepzion y aberle enbargado todas las renttas donde abia de cobrar y asta que se concluya no poder cobrar nada cansamos a Vm pidiendole que por amor de Dios nos aga gusto de darnos quatro o cinco mill Rs (1702, P.S.).

Example 15, taken from P.S., is a passage of a letter sent by Catalina Señor to her uncle, Pedro Señor y Angulo. A mother of seven children, Catalina Señor requests funds for child maintenance in her letter. The tenor is thus respectful, with use of the abbreviation *V.M.*, which the corpus editors rightly do not spell out. As the

P.S. corpus contains 9 letters sent by Catalina Señor to her uncle, other letters of the same collection reveal the meaning of the abbreviation: “en casa todos estamos buenos para lo que *usted* nos quisiere mandar que le obedezeremos con la boluntad que Vm sabe”, and “de corazon reciui la de *vuersa merced* y siento mucho que mi tia aya malparido”.

These passages thus reveal that the full form *vuesa merced* ‘your honour’ was still in use in the early 18th c. alongside the formal pronoun form *usted* ‘you’, which by then had become fully grammaticalized.

The letters reveal significant properties of the scribes’ language and, by extension, of the lexical resources of that period. The image copies of the documents evidence two different handwritings: one by a scribe who used *seseo* (*resetado* for *recetado* ‘prescribed’) and *yeísmo*, i.e. the pronunciation of the digraph *ll* as the grapheme *y* (*áyome* for *hállome* ‘I am’, *ayarme* for *hallarme* ‘to be’, *aller* for *ayer* ‘yesterday’); another by a scribe who used *leísmo*, i.e. the use of gender-unspecified *le* ‘(to) him/her/it’ for masculine *lo* ‘(to) him’ or feminine *la* ‘(to) her’ (“si Vm tubiere un capote que no le sirva me *le* embiara; no canso mas a Vm si no es que me *le* gde Dios”) and *laísmo* (“por no dar*la* pesadumbre le digo que no lo se y se me haze escrupulo el que aquella alma pierda las oraziones o misas que *la* puedan dezir”).

These letters are also useful for attestation of everyday words that are barely recorded in general corpora. Thus, Catalina, anxious about the cold in Madrid, repeatedly requests from his uncle “2 cargas de *arrax* porque los frios por aca an entrado” (‘two loads of [arrax] because the cold set in here’), i.e. two loads of “carbón de huessos de azeituna con que se hace un fuego mui apacible y durable para los braseros” (*Aut*.) (‘brazier coal made of olive pits for a very comforting and lasting fire’). This variant form of *errax*, originally from Arabic, was rare as late as the 18th c. and is recorded once in the CDH.

All in all, the above shows that specific questions need both specific *ad hoc* corpora to fill the gaps of general corpora, and the ensuing data analysis and interpretation, which go beyond mere large-scale data collection.

5 Conclusions

Review of the strengths and weaknesses of RAE (CDH and CORPES XXI) and non-RAE corpora (CdE and *EsTenTen18*) reveal higher sample quality and more accurate descriptive and presentational annotation in the former, and bigger size and higher interface flexibility in the latter.

Automatic sample collection from various websites and blogs increases corpus size and is less time-consuming and requires less effort during corpus compilation. Still, there is a downside:

1. sample selection is less precise and, as a result, the resulting corpus is less representative;
2. samples are collected from internet sources with poor geographical metadata, so a large number of examples cannot be ascribed to any language variety or are ascribed wrongly; and
3. sample quality is lower as a result of typographical mistakes (*parce* for *parece*, *parse* for *parte*, etc.) and of inconsistencies (passages in other languages, *parce que*); this results in wrong annotation and lemmatization and, therefore, the degree of precision and recall decreases.

Despite the above, the resulting picture is good, especially if the user is fully aware of the properties of their corpus and, especially, if complementary corpora can be added. The review of *algotro* illustrates the use of this *collaborative procedure* that runs from RAE corpora, goes through CdE and *EsTenTen18*, and reaches internet websites and social networks.

Small, specific corpora can supply data to address research questions that Big Data resources leave unanswered for their lack of highly specific samples and data analysis qualitatively different from their large-scale data collection.

Bibliography

Aut. = RAE (1726-1739): *Diccionario de Autoridades*. <https://apps2.rae.es/DA.html> (14-12-2021).

Biblia Medieval = Enrique Arias, Andrés: *Corpus Biblia Medieval* <http://corpus.bibliamedieval.es/> (12-12-2021).

Calderón Campos, Miguel & Díaz-Bravo, Rocío (2021): "An online corpus for the study of historical dialectology: *Oralia diacrónica del español*", in *Digital Scholarship in the Humanities*, 36, pp. 30–48.

Calderón Campos, Miguel and Vaamonde, Gael (2020): "*Oralia Diacrónica del Español*. Un nuevo corpus de la Edad Moderna", in *Scriptum Digital*, 9, pp. 167–189.

Castañeda Naranjo, Luz Stella (2005): *Caracterización lexicológica y lexicográfica del parlache para la elaboración de un diccionario*. Tesis Doctoral. Universitat de Lleida.

CdE= Davies, Mark: *Corpus del español*. <https://www.corpusdelespanol.org/> (15-01-2022).

CDH = Real Academia Española: *Corpus del Diccionario Histórico de la Lengua Española*. <https://www.rae.es/banco-de-datos/cdh> (16-01-2022).

CHARTA = *Corpus Hispánico y Americano en la Red: Textos Antiguos*. <https://www.corpuscharta.es/> (22-12-2021).

CORDE = Real Academia Española: *Corpus diacrónico del español*. <https://www.rae.es/banco-de-datos/corde> (09-01-2022).

CORDIAM = *Corpus Diacrónico y Diatópico del español de América*. <https://www.cordiam.org/> (22-12-2021).

COREECOM = *Corpus Electrónico del Español Colonial Mexicano*. <https://www.iifilologicas.unam.mx/coreecom/index.php?page=inicio&men=1> (22-12-2021).

CorLexIn = *Corpus Léxico de Inventarios*.<https://corlexin.unileon.es/> (22-12-2021).

CORPES XXI = Real Academia Española: *Corpus del español del siglo XXI*. <https://www.rae.es/banco-de-datos/corpes-xxi> (11-01-2022).

COSER = *Corpus Oral y Sonoro del Español Rural*. <http://www.corpusrural.es/> (22-12-2021).

CREA = Real Academia Española: *Corpus de referencia del español actual*. <https://www.rae.es/banco-de-datos/crea> (09-01-2022).

Davies, Mark (2009): "Creating Useful Historical Corpora: a Comparison of CORDE, the *Corpus del español*, and the *Corpus do português*", in Andrés Enrique-Arias (ed.), *Diacronía de las lenguas iberorrománicas: nuevas aportaciones desde la lingüística de corpus*. Madrid/Frankfurt: Iberoamericana/Vervuert, pp. 137–166.

Egbert, Jesse, Larsson, Tove & Biber, Douglas (2020): *Doing Linguistics with a Corpus: Methodological Considerations for the Everyday User*. Cambridge: Cambridge University Press.

ESLORA = *Corpus para el Estudio del Español Oral*. <http://eslora.usc.es/> (22-12-2021).

EsTenTen18 = *Spanish corpus from the web*. <https://www.sketchengine.eu/estenten-spanish-corpus/> (31-12-2021).

González Sopeña, Inmaculada (in press): "Corpus de textos notariales extremeños (CORTENEX S. XVII). La edición de un corpus histórico-lingüístico en el ámbito de las humanidades digitales", in *Dialectologia*, n.º 31.

Jakubíček, Miloš, et al. (2013): "The TenTen corpus family", in *7th International Corpus Linguistics Conference CL*. Lancaster, pp. 125–127.

Janssen, Maarten & Vaamonde, Gael (2020): "Da edición dixital á análise lingüística. A creación de corpus históricos na plataforma TEITOK", in Rosario Álvarez & Ernesto González Seoane (eds.), *Calen barbas, falen cartas. A escrita en galego na Idade Moderna*, Santiago de Compostela, Consello da Cultura Galega (Ensaio& Investigación), pp. 271–292.

Mair, Christian (2006): "Tracking Ongoing Grammatical Change and Recent Diversification in Present-Day Standard English: The Complementary Role of Small and Large Corpora", in Antoinette Renouf and Andrew Kehoe (eds.), *The Changing Face of Corpus Linguistics*. Amsterdam: Rodopi, pp. 355–376.

NGLE = RAE and ASALE (2009): *Nueva Gramática de la Lengua Española*. Madrid: Espasa.

Octavio de Toledo y Huerta, Álvaro (2016): "Sin CORDE pero con red: *algotras* fuentes de datos", in *RILI*, XIV, 29, pp. 19–47.

ODE = Calderón Campos, Miguel & García-Godoy, M.Teresa (dirs.): *Oralia diacrónica del español* <http://corpora.ugr.es/ode/> (16-01-2022).

PRESEEA = PRESEEA (2014-): *Corpus del Proyecto para el estudio sociolingüístico del español de España y de América*. Alcalá de Henares: Universidad de Alcalá. <https://preseea.linguas.net/Corpus.aspx> (12-12-2021).

P.S. = *Post Scriptum. A Digital Archive of Ordinary Writing (Early Modern Portugal and Spain)*. <http://teitok.clul.ul.pt/postscriptum/> (05-01-2022).

Rojo, Guillermo (2010): "Sobre codificación y explotación de corpus textuales: otra comparación del *Corpus del español* con el CORDE y el CREA", in *Lingüística*, 24, pp. 11–50.

Rojo, Guillermo (2021): *Introducción a la lingüística de corpus en español*. London / New York: Routledge.

Stefanowitsch, Anatol (2020): *Corpus Linguistics. A guide to the methodology*. Berlin: Language Science Press.

TEITOK = Janssen, Maarten: *TEITOK, a Tokenized TEI environment*. <http://teitok.corpuswiki.org/> (12-12-2021).

Vaamonde, Gael (2017): *Userguide for Digital Edition of Texts in P. S. Post Scriptum*. <http://teitok.clul.ul.pt/postscriptum/> (15-12-2021).

Vaamonde, Gael (2018): "La multidisciplinariedad en la creación de corpus históricos: El caso de *Post Scriptum*", in *Artnodes*, 22, pp. 118–127.

Carolina Gainza C.

Literature and Algorithms: "Aesthesis" and "Mathesis" in Digital Humanities

Digital Humanities: A Field in Tension

Digital Humanities (DH) and digital literatures have shown an important growth in the last few decades. However, it is interesting to see how, along with their growth, they have also split into two different disciplinary fields. This separation is problematic, as it reproduces the disengagement between "mathesis" and "aesthesis", encouraged by the settlement of science and particularly computational thought as privileged forms of generating legitimate knowledge, and the disciplinary schism that has made dialogue between sciences, arts, and humanities difficult. On the other hand, the growing drift of humanities towards quantification corresponds to a colonizing movement which stems from its practice and institutionalization in Universities in the United States, affecting the diverse practice of situated digital humanities.

I consider myself part of the field of digital humanities and, in this article, I would like to explain why. In a general manner, digital humanities have been defined from the global north as a form of applied research within humanities which intensively analyzes data through digital tools. It is not concerned to the digitalization of humanities but is related to a computational turn that affect its methodologies, forms of creation, resource management and the generation of computational applications (del Río 2015). Given this heterogeneity it is hard to define the scope of applications of digital humanities, as pointed out by many authors in the Hispanic world (Rodríguez 2014; del Río 2015; Ortega and Gutiérrez 2014). However, it is relevant to highlight what Nuria Rodriguez remarked: "What defines, then, Digital Humanities in contrast to the set of humanistic disciplines that simply use technological tools is the search of new interpretative models, new disruptive paradigms for the comprehension of culture and the world" (2014: 14).[1]

Despite this difficulty, and in the search of a greater proximity to science, a quantitative perspective has prevailed in digital humanities along with a search for scientific/objective legitimation, which has often weakened the interpretative

1 All translations in this article are mine. "Lo que define, pues, las Humanidades Digitales frente al conjunto de disciplinas humanísticas que utilizan herramientas tecnológicas es la búsqueda de nuevos modelos interpretativos, nuevos paradigmas disruptivos en la compresión de la cultura y del mundo".

https://doi.org/10.1515/9783110753523-009

perspectives. Johanna Drucker, in her critique of the exclusion of aesthesis in the practice of digital humanities, observes:

> Digital humanities is an applied field as well as a theoretical one, and the task of applying these metaconsiderations puts humanist's assumptions to a different set of tests. It also raises the stakes with regard to outcomes. Theorethical insight is constitutes in this field in large part through encounters with application. The statistical analysis of texts, creation of structured data, and design of information architecture are the basic elements of digital humanities. Representation and display are integral aspects of these activities, but they are often premised on an approach influenced by engineering, grounded in a conviction that transparency or accuracy in the presentation of data is the best solution. Blindness to the rhetorical effects of design *as a form of mediation* (not of transmission or delivery) is an aspect of the cultural authority of mathesis that plagues the digital humanities community (2009: 6).

I believe there are two problems in digital humanities which urgently need to be addressed. The first one has to do with the separation between aesthesis and mathesis, which takes relevance away from the role of imagination, critique, creation, and subjectivity in digital culture, in favor of the use of data to give humanities a status of scientific objectivity. Personally, as I will argue here, I do not believe these two to be mutually exclusive activities. The second problem has to do with the institutionalization of North American and European digital humanities, whose definitions and practices, mainly related to this separation between mathesis and aesthesis, are imposed onto other regions of the world. This not only leads to a colonization of data, but it also overshadows the diversity in the exercise of digital humanities. As Gimena del Río points out: "a genealogy of the whats, whys and for whats of machine and software technology is perceived as an unpaid debt in Digital Humanities at a global (not just Hispanic) scale" (2016: 102).[2] Here, del Río remarks that within North American and European academia there is already a closure in respect to a definition of digital humanities, which excludes other perspectives, such as the Hispanic ones (del Río 2016).

As we can see, the definition of digital humanities is far from being a field free of tensions. Its diversity, size and reach are so large that at times it seems like everything nowadays can be considered digital humanities. However, I would like to point out that the same diversity – of languages, disciplines, geographies, and themes – is precisely the most important characteristic of digital humanities, which must be preserved. In fact, these different perspectives and applications make it so that each practice of digital humanities is something different. As Padmini Ray

2 "Una genealogía de los qué, los por qué y para qué sobre la tecnología de las máquinas y el software se percibe como la deuda pendiente de las Humanidades Digitales globales (no sólo las hispánicas)".

Murray stated "Your DH is not my DH—and that is a good thing" (Qtd. in Risam 2016). It is indeed a good thing, because it allows scholars to "glocalize" the digital humanities practice, attending to their local concerns. This is something that is also posed by Gimena del Río, an acclaimed voice in Latin American DHs, when she highlights the diversity that exists within digital humanities, as well as the need to ponder DHs from the global south.

Therefore, in this paper I am interested in defending a non-binary perspective capable of integrating worlds, where it is not necessary to distance ourselves from questions of aesthetics, creation, and qualitative perspectives in order to justify our humanistic practice in current digital times. Even though as a researcher I do not intensively use digital tools to analyze literary information – which would correspond to the dominant definition of digital humanities – the analysis of the uses, appropriations and resignifications of digital technologies, the elaboration of theories and concepts that allow as to approach and comprehend phenomena within digital culture, still constitute a fundamental part of the development of this area of humanistic studies. I believe that the use of software in digital humanities to analyze and visualize data is something positive and has opened a field of study that embraces artistic and cultural phenomena. This would have been impossible to tackle in past decades without the help of these technologies. However, we must be careful not to fall into a fetishism of data, which could make us lose sight of the need to develop critical and reflective perspectives in respect to how technologies affect humanity and how we relate with the phenomena of digital culture. This last part is what I believe the dominant forms of understanding digital humanities leave out. In this respect, Drucker points out:

> The attitude that objectivity -defined in many cases as anything that can be accommodated to formal logical processes- is a virtue, and the supposedly fuzzy quality of subjectivity implicitly a vice, pervades the computation community. As a result, I frequently saw the triumph of computer culture over humanistic values (7).

This dichotomy that exists within DH, objectivity/subjectivity – mathesis/aesthesis, is producing a growing disciplinary division, which affects what has been the task of humanities throughout its historical development. In this presentation I seek to plead for a dialogue between mathesis and aesthesis in digital humanities, my goal being to avoid the reproduction within humanities of the science/humanities dichotomy between science and humanities that prevailed during the 20th century.

Regarding visibility of Latin American DH, let us take as an example this geopolitical map of digital humanities (Image 1), created by José Pino-Díaz and Doménico Fiormonte (2018). The map is based on the participation of investigators in the congress of digital humanities that took place in Montreal in 2017. We can see

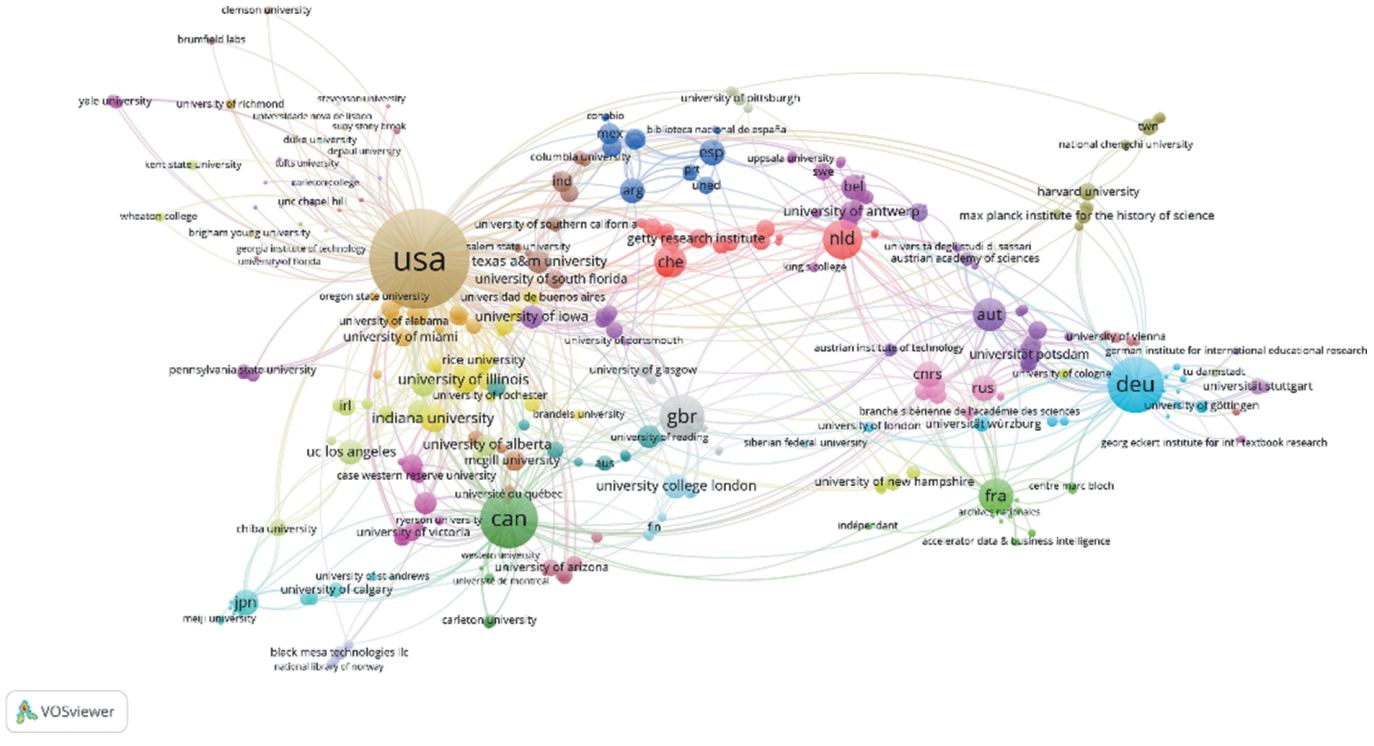


Image 1: Geopolitics of Digital Humanities (Pino- Díaz and Fiormonte 2018).

that Latin America, as well as other southern regions, have practically no representation. One could argue that this is because the congress was carried out in Canada, and that is one of the reasons, but the truth is that, in general, it looks like DH do not exist in Latin America and other regions of the world.

According to the authors, this map goes to show the hegemonic part that Digital Humanities from the global north play in respect to the definitions and practices of digital humanities. This, of course, does not mean that DH are not present in the global south, but it speaks to the invisibility of their practices in the face of the homogenizing tendency that stems from a greater institutionalization of DH in northern academia (Del Río 2016), which moreover, coincides with a predominance of mathesis in their practices. On the other hand, Élika Ortega and Silvia Gutiérrez generated a map of DHs in 2014, which considers Spanish, Latin American, and Portuguese academics. They compiled the information during 2013 through an online survey sent to researchers in digital topics in the humanities, and then spread it through social media. The results showed that in Spanish and Portuguese speaking countries, digital humanities were growing through literary studies, especially in the field of electronic and digital literature, and not so much in relation to matters of methodological innovation. In the analyses presented by these authors it is possible to observe within those years a growing link between DH and that which I call “digital aesthetics”, meaning the creative and aesthetic practices risen from the use of digital technologies in literary creation. In this sense, the penetration of the digital into humanities in Latin America had more to do with asking questions in respect to digital cultural phenomena than developing humanistic research involving the intensive use of digital tools.

The presence of digital humanities in Latin America has grown considerably in recent years, with the emergence of digital humanities organizations in Mexico, Colombia, Brazil and Argentina. It is also possible to identify a growing number of digital humanities researchers who are not on the maps. Gimena del Río (2019), together with other digital humanities researchers, have proposed the need to focus on the development of DH from the South, without waiting for validation from those developed from the North, as well as the need for Latin American governments to invest more in the development of these areas, both in infrastructure (databases and archives) and in research. For now, digital humanities in Latin America exist thanks to the effort and drive of researchers committed to the area, rather than concerns or policies that promote its development and institutionalization, as can be seen in Europe and the United States (del Río 2019).

Although digital humanities cover a wide range of topics, the homogenizing tendency of Northern DH has translated into a predominance of their mathematical dimension, data management and quantitative techniques for analyzing large

volumes of information, with the goal of identifying trends and patterns, elaborating maps, among others. As pointed out by R. Risam:

> In spite of ongoing work to rewrite the maps of global digital humanities, a troubling trend appears in digital humanities citations: erasure of local difference. Staci Stutsman's analysis of digital humanities syllabi has demonstrated that the same handful of theorists from these countries (Susan Hockey, Lev Manovich, Matthew Kirschenbaum, Dan Cohen, Franco Moretti, and Stephen Ramsay) is being taught repeatedly, with little variation. These are, indeed, the same names that recur in digital humanities scholarship in South Africa, Nigeria, India, and South Korea, regardless of their relevance to local context. There is an imperative here to move from a logic that centers the Global North—advanced industrial and high-income economies—in digital humanities toward embracing the diversity of practices around the world and the intersecting forces that shape them. (Risam 2016).

Personally, the predominance of a mathesis perspective concerns me. This is to say, the datafication and quantification that we currently observe in digital humanities trying to legitimize themselves as scientific, is something that happened before throughout the 20th century in the humanities. I think that one of the characteristics of digital humanities that should not be lost is their power to link humanities and science, that is, the possibility of generating transdisciplinary work. This would allow combining aesthetic dimensions, linked to interpretation and subjectivity, and mathematics, related to data and computational applications.

We can identify, therefore, two main ways of practicing digital humanities that must necessarily be in dialogue. The development of new methodologies is perhaps the dominant area, including the use of digital tools to analyze data, make visualizations and cross-references between large amounts of information, based on questions specific to the humanities. The other area has to do with research on cultural, literary, visual, and linguistic modalities in the digital era. I personally use digital tools to make maps and visualizations, but my main work has to do with what is proper to the humanities: the question of language, experience, aesthetics, literary creation, and how it accounts for the phenomena we are living. Methodological innovation, as a result of the incorporation of data analysis software in the humanities, has allowed the opening of transdisciplinary practices. But it is very important not to lose sight of the questions, concepts, and theories, which sometimes land in the back seat in relation to the fascination with data, or the pretension of making a science out of literature and the humanities in general.

In this sense, I am interested in arguing that digital humanities are meant to contribute in looking at the world differently, and not be reduced to data analysis in the humanities or visualization of literary and artistic databases. DH bring together research on how digital technologies affect our forms of creation and aesthetic experience, how we relate to them, and how they affect the status of humanity, along

with new methodologies for analyzing phenomena in the humanities. If DH are defined as transdisciplinary, it seems contradictory to me to introduce disciplinary binarisms within the very exercise of the humanities.

Algorithmic Imagination and the Poetics of Code

My intention, then, is to contribute to break down the humanities-sciences dichotomy that is reproduced in some of the dominant ways of practicing digital humanities. In the face of an increasing quantitative drift within digital humanities, in this article I am interested in showing how mathesis and aesthesis can establish a dialogue in the field of digital literature through the intensive use of programming languages in contemporary literary creation. As Johanna Drucker (2009) points out:

> From my very first encounters with digital media, I have been convinced that the powerful cultural authority exerted by computational media, grounded in claims to objectivity premise on formal logic, can be counterbalanced through aesthetic means in which subjectivity is central to the concept of knowledge as interpretation (xiii).

Digital literature illuminates questions regarding our digital environment and leads us to ask ourselves questions that are not only of literature, but the humanities in general. This type of literature has computer code at the basis of its definition. Following Katherine Hayles (2008), the digital code and the algorithms that compose it are a structural part of this type of creative practice, where code language participates in a direct or mediated way. In the first group, those where the literary piece is directly programmed, the works are based on a programming language that allows the generation of expanded, interactive and multimedia experiences that would not be possible to appreciate or access in a printed format. To this we can also add those literatures in which algorithms or artificial intelligences have been programmed to intervene in the creative process, whether it be in certain parts (as in generative literature) or in its entirety (as in some works generated by artificial intelligences). In the second, where the literary piece is not programmed directly but the code is part of the software or platform used, we find those literatures that experiment with digital media, where creators take advantage of the tools offered mainly by social networks. We find in this dimension the twitterature (Twitter), instapoetry (Instagram), transmedia and multimedia stories on various platforms, literatures on WhatsApp, among others. Literatures that experiment with the medium have been the least studied in the field of Latin American digital literature, and here is a field of research yet to be explored.

Code, the hidden language that seems intelligible to us, configures the poetics of digital writings: their interactivity, hypertextuality, hypermediality, and the possibility for these same algorithms to acquire an aesthetic sense. Digital humanities and digital literature fulfill the function of opening up spaces of legibility, both in the sea of literary and artistic data through methodologies of visualization and analysis in the former, and of algorithmic imagination in the latter. In digital literature digital codes are projected in verbal, visual and sound languages, interconnecting them and experimenting with algorithms to generate literatures that enable us to reflect on the language that surrounds us and passes through us, configuring an algorithmic culture. In this sense, digital literature exposes the problem of code, which currently affects our experiences and forms of perception.

Dene Grigar, in her introduction to the book "Electronic Literature as Digital Humanities" (2021) emphasizes that what we are really dealing with, therefore, are two complementary areas of study:

> In fact, we argue that Electronic literature is the logical object of study for digital humanities scholars who have, by the second decade of the twenty-first century, cut their teeth on video games, interactive media, mobile technology and social media networks (. . .) in sum, electronic literature is digital humanities because of our shared philosophy that a computer is not a tool or prosthesis that helps us to accomplish our work; rather, it is the medium in which we work (3).

The field of digital literature is primarily concerned with the study and criticism of literatures born in and for digital media. That is, it mostly deals with the aesthetic and programming language fields. Why should this be outside the digital humanities? In her claim against mathematization of DH, Drucker, paraphrasing Aristotle, declares "the role of aesthetics is to illuminate the ways in which the forms of knowledge provoke interpretation" (2009: xiii). Digital literary works are not only valuable because of their use of digital technologies and languages in artistic creation, but even more so because they intervene in the discourses installed in digital culture and demand new interpretative models from literary and humanistic criticism. I believe that to leave digital literature outside the field of the humanities, as I have personally experienced in some recent conferences or in conversations with researchers in the digital humanities, is to renounce the possibility that humanities constitute a bridge in the contemporary that allows the critical tradition of the humanities to dialog with the sciences and attend to their mutual influences.

Data and Aesthetics Can (and must) Dialogue Within Digital Humanities

In my recent research, developed alongside academic Carolina Zúñiga at the Digital Laboratory of the Universidad Diego Portales, we published a *Cartography of Latin American Digital Literature*, an information design project that visualizes information of Latin American digital literature. The website allows a cohesive visualization of the trajectory in this area, relating the works produced according to their place and year of production, literary genre, format, and techniques used. It is a multidimensional, comprehensive, and interactive research tool for researchers of literature and digital culture, as well as for the general public to learn about the works being developed in this field. The project records the development of a growing creative community from a territorial perspective. On the other hand, it explores interdisciplinary archival science, humanities, and design, taking advantage of the strengths of cartography in its practical and reflexive dimension. It contains and represents information, while trying to propose a critical view of the mapping exercise.

This description speaks of a project very typical to the digital humanities. However, beyond information design and visualization, the main objective is to condense these works in a single place in order to analyze them and characterize Latin American digital literary creation. What digital literature does is question the monopoly of verbal language, the definition of literary genres and territorial boundaries. On the other hand, digital literature also leads us to ask ourselves about the poetics of code, its creative possibilities and how they affect subjectivity, or the emergence of what I have called a “digital condition”.

Alexander Galloway (2004) pointed out several years ago how little attention literary criticism paid to data and algorithms. Digital literature exists in data, or in Yuk Hui’s (2016) words, writing is datafied and data objectified, that is, turned into digital objects. These objects, like the works present in cartography, are read by us, but also by machines, in order to be executed and have their various characteristics appear in front of our eyes. For these readings to occur, between digital objects, and between digital objects and humans, mediations and interfaces are necessary. For the machine to read code language, it must recognize it as an object, and this happens through metadata: “in the age of hypertext, online objects are only meaningful to humans, not to machines. However, in the age of metadata, online objects are considered meaningful to both, machines and humans. Machines understand the semantic meaning of objects via the structures given to the metadata” (Hui 2016: 52). These translations, between human and

non-human languages, exist because of the transmission of information through interfaces that allow reading and translating one language into another.

In our mapping, metadata is very important. We used metadata protocols (Dublin Core) that could be recognized by machines and, consequently, form networks with other digital literature archives. In other words, the digital system recognizes them as digital objects and can establish relationships with other digital objects. This recognition of the set of data as a digital object by the machine, is made possible by a mediation, a reading of data that is made possible by algorithms that function as interfaces, as mediators. In the absence of this interface corresponding to metadata protocols, the file or the literary piece remains isolated and does not connect with the existing digital medium. In this sense, interfaces allow the connection between different elements, which is fundamental for the digital object to be constituted as such. Subsequently, there are other mediations that make the interaction between these digital objects and humans possible. If we were to stick to the quantitative tendency of the digital humanities, all this theoretical-conceptual analysis on the objects contained in the cartography, of a second order if you will, would not have been considered.

From the proliferation of these literatures during the first decade of the 21st century, literary criticism established dialogues with computer science, media studies and philosophy of technology, among other disciplines, from which concepts and theories have emerged for addressing specific aspects of digital literature. Among these are Code Studies, Game Studies and Electronic/Digital Literature itself as a field of study on its own. Although it could be said that these areas do not innovate methodologically, their object of study is a digital object, which is problematized, analyzed and criticized from the field of digital literature studies by creating new concepts and theories that account for the specificity of digital literary processes and practices.

In the cartography we can find an important number of works that take advantage of both the methodologies of digital humanities and the conceptual-theoretical frameworks of digital literature to be analyzed. In this regard, these are works that require conversation between these two areas, a matter that respond to a local perspective of the development of these disciplinary fields in the region and not the trends of the global north that separate them. As Claudia Kozak (2020) points out, it is possible to make these areas converge if we adopt a decolonial perspective. The generative poetry of Milton Läufer, the hypermedia of Jaime Alejandro Rodríguez, the narrative and hypertextual poetry of Carlos Labbé and Carlos Cociña respectively, meme and gif poetry by Canek Zapata, or the interactive and multimedia poetry of Michael Hurtado or Karen Villeda require a new theoretical-conceptual apparatus in order to be approached. For example, in Unicode by the Peruvian poet Michael Hurtado, we find a sea of codes

apparently without center, illegible under the literary canons or apprehended meanings. In this poem, different codes are presented and engaged in dialogue: the numerical code, the genetic code, the poetic code and the unicode. When moving through the page, the codes move and recombine until the user stops the mouse. In this sense, it is a poetic work that, as is characteristic of digital literature, breaks with our horizon of expectations and requires new conceptual apparatuses that allow us to account for this aesthetic experience.

On the other hand, in *Do bots worry about writer’s block* by Argentine writer Milton Läufer we find a generative piece in which an algorithm co-writes with the writer. This piece is part of the “writer’s tools” series created by Läufer with the aim of “helping blocked authors”. From a word entered by the user, the algorithm returns a text related to that word and, as we advance with the mouse, the text increases its content. Under the verbal language that the interface shows us, algorithmic operations are performed which affect the text and, along with that, our subjective experience of literature, reading and writing is also affected.

Another interesting example is literary creation with artificial intelligence. Not only because of the type of writings that are generated, but also because it leads us to ponder the possibility of an aesthetic of algorithms, one that exists beyond the human. We find this in the case of *Mexica*, by Mexican author Rafael Pérez y Pérez, which contains a collection of stories generated by an artificial intelligence that has been perfected by its author since the late 1990s. The stories generated by these intelligent algorithms come from traditional Aztec tales. The algorithm is fed with a database of stories, from which it learns the grammatical structure, certain literary forms and logical sequences of actions. For example, if there is a conflict in the story it is resolved by a fight that may end in the resolution of the conflict or the death of the protagonist or another character. In this way, the algorithm, in an exercise of trial and error, learns to generate logical sequences, giving way to short stories with motifs learned from those stories with which the Mexica algorithm was fed.

In these literatures we can see that there is a poetic function of language, but not only in the verbal language used, but also in the code with which the piece is programmed. From this observation, it is possible to analyze the human-algorithm relationship, digital aesthetics, subjectivities, forms of imagination, writing and reading practices in digital, among many other aspects that today are addressed by the field of digital literature. An algorithm is made of numbers, it is pure mathesis. And yet, in digital literature it acquires an aesthetic dimension that must be studied through new concepts that lead us to embrace the aesthetic particularity of this phenomenon.

Tearing Down Disciplinary Frontiers: In Search of a Decolonized Practice of DH

According to what has been discussed so far, digital literature is a field that innovates not only in writing practices, but in turn, these demand a conceptual and theoretical creativity. In this sense, digital humanities cannot be reduced to visualization methodologies, big data and use of technologies, but it is also necessary to create new concepts and theories to address the digital phenomena that surround us. In this manner, digital humanities can and should converse with the field of digital literature, and that is something that has, in a way, already happened in Latin America. It is important that, from a decolonial point of view, we do not allow the trajectories that these areas have followed in other parts of the world to be imposed on us (del Río 2016; Kozak 2020). The fundamental question we all ask ourselves today is what is human and how we relate to non-human existences. And the humanities are, above all, asking questions about existence, subjectivity, and experience.

Considering digital literature as part of the digital humanities, which does not mean failing to recognize its singularity, aims towards generating a dialogue that avoids the reproduction of the dichotomy science/objectivity/data and humanities/subjectivity/interpretation. Based on the material gathered in the cartography project, we are working on a critique of digital literature that allows us to identify a certain poetic, which ultimately is what enable us to speak of literature, where the interaction between algorithmic languages, humans and the mediations that interfaces allow are taken into account. Just as we cannot speak of literature without referring to the poetic work with language, its structure and meanings, the ways of reading and the book, its material and conditions of production, digital literature pays attention to algorithmic languages, digital interfaces and their materialities, as part of what allows us to speak of literature in the digital context. The cartography of Latin American Digital Literature that we constructed not only groups and makes works and authors visible, but also raises questions regarding forms of appropriation, the existence of diverse relationships with technology, how it is re-signified and how it permeates the construction of Latin American subjectivities in the current global context. These literatures account, somehow, for what Yuk Hui calls “cosmotechnics” (2020), that is, they question the particular ways in which algorithmic languages are used, thematized and appropriated, configuring aesthetics that situate technologies in specific cultural contexts. This breaks away from the idea that, from the global south, we are “consumers” of technologies and literatures, as in these literary pieces we can observe exercises of decolonization and production of thought regarding how we incorporate and transform digital technologies.

To conclude, I have tried to manifest against the imposition of a dominant practice of digital humanities that reproduces the binary exclusion of science/humanities, objective/subjective, mathesis/aesthesis. Digital humanities, as we can observe in Latin America, should not exclude fields devoted to interpretation and aesthetics, such as digital culture studies, digital literature o media arts. These fields allow us to speculate and imagine possible and alternative futures. As I argued, the quantitative dominant practice of DH is associated with new forms of colonization of knowledge as well. So, I defend that the research practice we conduct in the field of Latin American Digital Literature can be defined as related and belonging to Digital Humanities. Particularly, my goal is to understand how digital culture challenges and disputes the homogenic forms of knowledge operating in computational logic and datafication. I might or might not use computational methodologies to organize information and to analyze it. But my research work, as humanist, is related mainly to proposing forms of reading and interpreting our contemporary digital culture through innovative creative practices.

Bibliography

Del Río Riande, Gimena (2015): “Humanidades Digitales. Mito, actualidad y condiciones de posibilidad en España y América Latina”, in *ArtyHum. Revista de Artes y Humanidades*, 1, pp. 7–19.

Del Río Riande, Gimena (2016): “De todo lo visible y lo invisible o volver a pensar la investigación en humanidades digitales”, in *UNED Revista Signa*, 25, pp. 95–108.

Del Río Riande, Gimena (2019): “Humanidades Digitales: infraestructuras visibles e invisibles”, en Humanidades Digitales LAB. <http://hdlab.space/HD-infraestructuras/#fnref:1> (04-01-2022).

Drucker, Johanna (2009): *SpecLab: Digital Aesthetics and Projects in Speculative Computing*. Chicago: University of Chicago Press.

Grigar, Dene and O’Sullivan, James (eds.) (2021): *Electronic Literature as Digital Humanities. Contexts, Forms, and Practices*. New York: Blooomsbury.

Hayles, Katherine (2008): *Electronic Literature. New Horizons for the Literary*. Indiana: University of Notre Dame.

Hui, Yuk (2016): *On the Existence of Digital Objects*. Minneapolis: University of Minnesota Press.

Hui, Yuk (2020): *Fragmentar el futuro. Ensayos sobre tecnodiversidad*. Buenos Aires: Caja Negra.

Hurtado, Michael (2019): *Unicode.* <https://www.cartografiadigital.cl/record/CD0000176≥ (15-12-2021).

Kozak, Claudia (2020): “Experimental Electronic Literature from the Souths. A Political Contribution to Critical and Creative Digital Humanities”, in *Electronic Book Review*. <https://doi.org/10.7273/zd5g-zk30>; (5-12-2021)

Läufer, Milton (2016): *Do Bots Worry About Writer’s Block?*. <https://miltonlaufer.com.ar/digresiones/> (15-12-2021)

Ortega, Élika & Gutiérrez, Silvia E. (2014): “MapaHD. Una exploración de las Humanidades Digitales en Español y Portugués”, in Esteban Romero & María Sánchez (eds.), *Ciencias Sociales y Humanidades Digitales. Técnicas, herramientas y experiencias de e-Research e investigación en colaboración*, CAC, Cuadernos Artesanos de Comunicación, 61, pp. 101–128.

Pérez y Pérez, Rafael (2017): *Mexica.* <https://www.cartografiadigital.cl/record/CD0000022> (30-11-2021)

Pino-Díaz, José and Fiormonte, Doménico (2018): "La geopolítica de las humanidades digitales: un caso de estudio de DH2017 Montreal", in J. Girón Palau and I. Galina Rusell, (eds.), *Digital Humanities 2018: Puentes-Bridges*, pp. 576–581.<https://riuma.uma.es/xmlui/handle/10630/16329?show=full> (04-01-2022).

Risam, Roopika (2016): "Navigating the Global Digital Humanities: Insights from Black Feminism". In Matthew K. Gold y Lauren F. Klein (eds.), *Debates in Digital Humanities 2016*. Minneapolis: University of Minnesota Press. <https://dhdebates.gc.cuny.edu/read/65be1a40-6473-4d9e-ba75-6380e5a72138/section/4316ff92-bad0-45e8-8f09-90f493c6f564#ch29> (04-01-2022).

Rodríguez, Nuria (2014): "Prólogo: Humanidades Digitales y pensamiento crítico", in Esteban Romero and María Sánchez (eds.), *Ciencias Sociales y Humanidades Digitales. Técnicas, herramientas y experiencias de e-Research e investigación en colaboración*, CAC, Cuadernos Artesanos de Comunicación, 61, pp. 13–18.

3 Practical Applications

Carolina Ferrer

The Literary System of the Iberian Worlds Through the Lens of Criticometrics

According to the historians José Javier Ruiz Ibáñez and Óscar Mazín Gómez:

> The Iberian worlds are constituted by the group of people whose culture and history were forged when the matrix of the Mediterranean world was projected through the Portuguese and Spanish expansion. The populations that are protagonists of this process (Europeans, Africans, Americans, and Asians) and their descendants share a common experience and a way of conceiving the world (Ruiz Ibáñez y Mazín Gómez 2021:2).[1]

The purpose of our study is to map the literatures that belong to the above defined Iberian worlds, in order to reveal the complex relations that the national literatures that constitute these worlds have developed through time.

To this effect, this research is located at the convergence of two phenomena: the conceptual turn from comparative literature into world literature (Damrosch 2014; Gupta 2009; Saussy 2006), and the emergence of big data in the humanities (Boyd and Crawford 2012; Mayer-Schönberger and Cukier 2013; Schreibman, Siemens, and Unsworth 2004, 2016). Specifically, in this paper, we will illustrate how this availability of massive amounts of information for the humanities – unimaginable not long ago – will allow us to analyze the configuration of the literary system of the Iberian worlds.

We would like to emphasize that, in this research, we modify the usual top-down viewpoint to introduce a bottom-up perspective. To achieve this, we use the methodological approach of criticometrics (Ferrer 2011), that we have developed based on the exploitation of digital databases and which makes it possible to articulate theoretical concepts with empirical research. Instead of imposing pre-established criteria, this approach stems from the observation of thousands of

1 "Los mundos ibéricos están constituidos por el conjunto de personas cuya cultura e historia se forjaron cuando la matriz del mundo mediterráneo se proyectó mediante la expansión portuguesa y española. Las poblaciones protagonistas de ese proceso (europeos, africanos, americanos y asiáticos) y sus descendientes comparten una experiencia común y una forma de concebir el mundo" (Ruiz Ibáñez y Mazín Gómez 2021: 2). We translate.

Note: This study draws on research funded by the Social Sciences and Humanities Research Council of Canada, granted to the CRSH 435-2018-1115 project « Les études littéraires et les nouveaux observables de l'ère numérique : le système de la littérature mondiale de l'après-guerre à nos jours ».

https://doi.org/10.1515/9783110753523-010

studies carried out by the international academic community and relies on the law of large numbers.

1 Systems Theory

In 2017, based on system analysis – introduced in the social sciences and the humanities by Niklas Luhmann (2000) and Immanuel Wallerstein (2004) –, and, more precisely, on the concept of *champ* developed by Pierre Bourdieu (1992; 1997) as well as on polysystem theory proposed by Itamar Even-Zohar (1990), we inaugurated a research project in order to map world literature (Ferrer 2018).

According to Pierre Bourdieu, "the social world can be decomposed into a multitude of microcosms, the fields or 'champs', each one featuring specific stakes, objects, and interests" (Bourdieu 1997: 119).[2] Moreover, the configuration of the fields is characterized by complex relations between their diverse components (Bourdieu et Wacquant 1997: 72).

Likewise, the interactions between national literatures also constitute complex phenomena. In order to understand these interactions, we use Itamar Even-Zohar's polysystem studies, that allowed him to formulate hypothesis about the functional relations between the components of the systems and subsystems under scrutiny. As stated by the Israelian semiotician, the literary system is defined as "the network of relations that is hypothesized to obtain a number of activities called 'literary,' and consequently these activities themselves observed via that network" (Even-Zohar 1990: 28).

One of our goals is to articulate theoretical concepts with empirical research. To achieve this, we developed criticometrics, a methodology that we have elaborated inspired by scientometrics.

2 Criticometrics

Initially developed by Derek De Solla Price (1963), scientometrics became possible due to the tools elaborated by Eugene Garfield, founder of the Institute for Scientific Information, that later became *Thomson ISI* and nowadays is known as *Clarivate*. The goal of scientometrics is to measure and to analyze scientific and

2 « Le monde social moderne se décompose en une multitude de microcosmes, les *champs*, dont chacun possède des enjeux, des objets et des intérêts spécifiques » (Bourdieu, 1997 : 119). We translate.

technological activity. By analogy, we created criticometrics,[3] with the goal to measure and to analyze critical activity in the arts, particularly in literary studies. This meant to adapt scientometric indicators to the reality of the existing databases dedicated to the humanities and the arts, with the aim of exploiting the metadata contained in these bibliographies.

Following our analogy, we present the main scientometric indicators. There are two categories of indicators. Those in the first category are called descriptive indicators or activity indicators. Those in the second category are relational indicators. They can all be measured at several levels of aggregation: researchers, research group, institute, country, discipline. The simplest descriptive indicator is the count of publications or patents. Another descriptive indicator is citation counts (Leydesdorff 1998). This corresponds to the number of times a text is cited in another publication. Supposedly, this indicator signals the quality of a publication. However, this argument has been extensively debated, and the consensus has transformed it into an indicator of visibility (Cozzens 1985). In the case of criticometrics, we follow the viewpoint stated by Kees van Rees, who considers that "A reliable indicator of the quality attributed to a work of art is permanent and intensive attention – in the form of (spoken or written) discourses" (van Rees 1997: 93).

In turn, the simplest of relational indicators is the co-signing of publications. One of the problems with this indicator is the difference of publication habits in the different disciplines. To overcome the shortcomings of the citations indicator, Henri Small (1973) created that of cocitations: to count the number of times that two citations appear simultaneously in a publication. This coincidence of references would indicate a closer link between the documents that contain them. Another method for solving the problems of the citations indicator was created by Callon, Courtial, and Penan (1993). Instead of considering the number of references, they proposed to count the cooccurrences of words in the documents. "The higher the frequency of the cooccurrence of words in different texts, the more reinforced are the research problems and the connexions between these problems" (Callon, Courtial et Penan 1993: 81).[4]

In the case of criticometrics, we measure the attention received by works and writers in the form of discourse, to use van Rees's terms. Consequently, we favor three types of indicators: firstly, the cooccurrences of words, such as the national literature label; secondly, the number of citations, for instance, the number of

3 This is a summarized version of criticometrics. For a more detailed presentation, see Ferrer 2022.

4 « Plus les mots co-occurrent fréquemment dans des textes différents et plus les problèmes de recherche et les connexions entre ces problèmes se renforcent » (Callon, Courtial et Penan 1993 : 81). We translate.

publications relating to a writer or a literary work; thirdly, the number of cocitations, which allows us to reveal the relationships between the writers, or literary works, who are the subject of comparative studies.

In this research, we interrogated the main literary database, the *Modern Language Association International Bibliography*, to which we will refer to as MLA. This database contains over 3 million references and covers more than 170 years of texts published by the international academic community. Several types of documents are listed: articles, books, book chapters, editorials, theses.[5]

Firstly, we compiled the critical references for each national literature, by querying the terms "Portuguese literature", "Spanish literature" and so on, thus, using a data mining technique (Han, Kamber, and Pei 2012; Witten, Frank, and Hall 2011) based on the cooccurrences of these terms (Callon, Courtial, and Penan 1993). Secondly, the total sample was created by forming an ensemble of the national literatures. Given that some references include more than one national literature, the conglomerate of references of the literary system of the Iberian worlds is not the simple addition of the national literatures.

3 The Literary System of the Iberian Worlds

The system of the Iberian worlds is composed of the countries whose territory is in the Iberian Peninsula or, at some point, were colonized by the Spanish Empire or by the Portuguese Empire. Since all the territories previously colonized by these empires obtained their independence several decades ago, and in some cases more than two centuries ago, we added an additional condition for the constitution of the system: at least one of the official languages of the selected countries must be Spanish (Castilian) or Portuguese. According to these criteria, the sample obtained from MLA contains 210,021 references and includes 2 European literatures, 20 literatures of the Americas, 6 African literatures, and 2 Asian literatures. The first reference in MLA was registered in 1887. The end date for our sample was set at 2018.

Based on the number of references by national literature, Figure 1 corresponds to the map of the literary system of the Iberian worlds.

In Figure 2, we observe the distribution of the publications by national literature. There is a huge disparity in terms of the critical bibliography dedicated to

5 However, in this study, we omitted dissertations, as they are essentially restricted to those defended in the United States, which would introduce a bias the data. See *Dissertation Abstracts International*.

each of them. For instance, Spain cumulates 531 times the number of publications compared to those dedicated to Equatorial Guinea, which is the only Spanish speaking national literature of the African continent.

From the continental viewpoint, Figure 3, we see that the European countries cumulate 54% of the publications, those belonging to the Americas 45%, and those from Africa less than 1% of the system. The publications about the Iberian literatures in Asia are extremely scarce.

The chronological evolution of the publications about the system is represented in Figure 4. Although the first reference goes back to 1887, we only observe a positive tendency of the series after World War Two, particularly since the mid-fifties. This slope remains positive until 2004, when the number of publications reaches its maximum of 5,839 references.

Table 1 shows the 25 most studied writers of the system of the Iberian worlds. Among them, 13 are European, and 12 from the Americas. Moreover, we observe that Spain concentrates 11 writers. There is only one woman in this list: Juana Inés de la Cruz.[6] Their life period cover from the 16th to the 21st centuries.

To diversify the list of writers, in Table 2, we present the 25 writers of the Iberian worlds whose critical bibliography in MLA concentrates the highest percentages of their national literature. We thus see several names from the Americas, Africa and even Asia. Among these, we find 4 of the 6 Spanish American writers who have received the Nobel Prize in literature. Moreover, we observe the presence of 3 women.

Table 3 corresponds to the 25 most studied literary works of the system. Again, Spain is at the top of the list, this time with 2 titles. We also observe the presence of a Portuguese title. The 22 other literary works are from the Americas, with 2 of them signed by women. We must point out that 5 titles belong to the Colombian writer Gabriel García Márquez. There also are 5 titles signed by Cuban authors.

In Figure 5, we represent the linguistic distribution of the publications by national literature and for the Iberian worlds system. At a systemic level, Spanish accumulates 57% of the sample, followed by English with 25% of the references. Portuguese is in third place with 10% of the publications. Next come French, 3%, followed by German, Italian and Catalan with 1% of the publications each. At the national level, we observe that, except for Guinea-Bissau, Timor-Leste, and Macau, where publications in English prevail, the highest percentage of publications corresponds to Spanish or Portuguese, both for the literatures of Spain

6 For a detailed analysis of the place of women writers in world literature, see Ferrer (2019).

and Portugal, as well as for the literatures of the countries that were formerly colonized by them.

Thus, we consider that it is extremely important to measure the linguistic diversification of the system. To this aim, we introduce a diversification indicator. It is inspired, on the one hand, by the Herfindahl-Hirschmann Index (Herfindahl 1950; Hirschmann 1945), used in economics to measure the degree of concentration of the markets, and, on the other hand, by the index developed in political sciences by Douglas W. Rae (1967), with the goal of measuring the fractionalization of political party systems. One of the advantages of this indicator is that it enables the comparison of systems that have different numbers of components.

The indicator of linguistic diversification that we propose is similar to the one elaborated by Rae:

$$\mathrm{LD} = 1 - \sum_{j=1}^{m} L_j^2$$

where L_j is the part of the total documents published in the language j.

The value of LD varies between 0 and 1. When one language concentrates an important proportion of the publications, the LD index is close to 0. On the opposite pole, when the publications are more equally shared by several languages, the index tends to 1.

Figure 6 represents the LD index of the literary system of the Iberian worlds and of the national literatures that compose it. Globally, its value is 0.6. Timor-Leste is the literature with the lowest index: only 0.26. It should be noted that, although Portuguese is one of the two official languages in that country, the critical bibliography in Portuguese corresponds to only 15% of the sample, since English accounts for 85%. São Tomé and Príncipe shows the highest index: 0.67. In this case, the bibliography is divided between Portuguese, 44%, English, 32%, French, 16%, and Spanish, 2%. In both cases, they are embryonic literary fields belonging to countries that have undergone great political changes in recent decades. Evidently, the percentage of publications in English, which fluctuates between 12%, in the case of Costa Rica, up to 85% in the case of Timor-Leste, would be an indication of a possible literary interference, according to the terms used by Even-Zohar.[7]

7 In our opinion, the interference exercised by the United States in numerous literatures of the Iberian worlds is undeniable. For instance, regarding the interference of the United States in Spanish American literature, see Ferrer (2014).

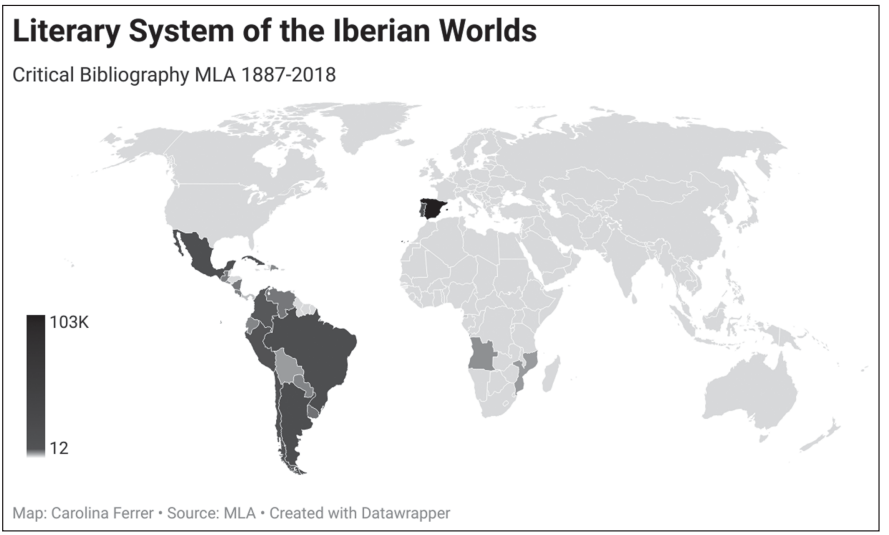
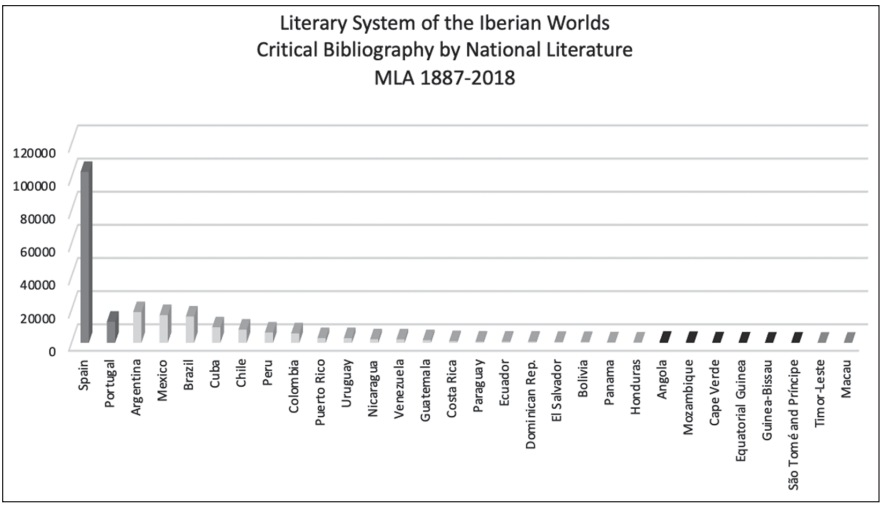


Figure 1: Map of the Critical Bibliography by National Literature (MLA 1887–2018).



Figrue 2: Critical Bibliography by National Literature (MLA 1887–2018).

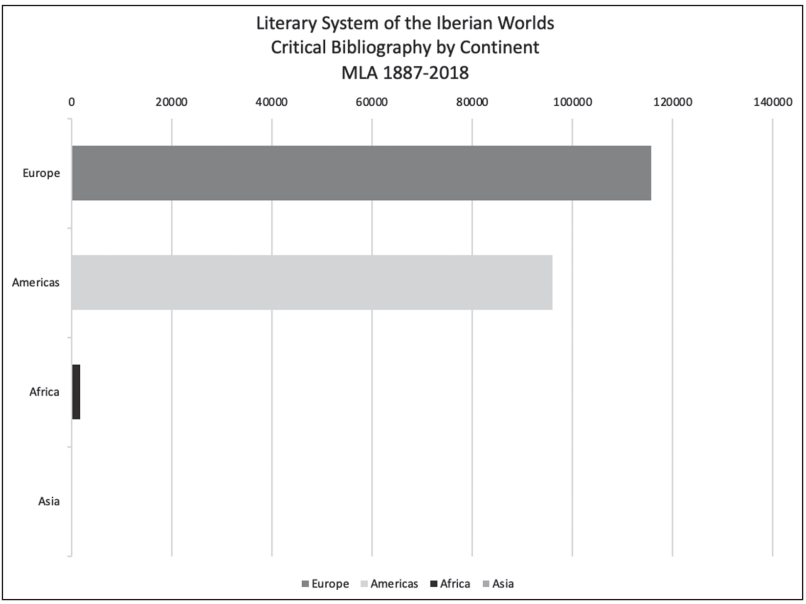


Figure 3: Critical Bibliography by Continent (MLA 1887–2018).

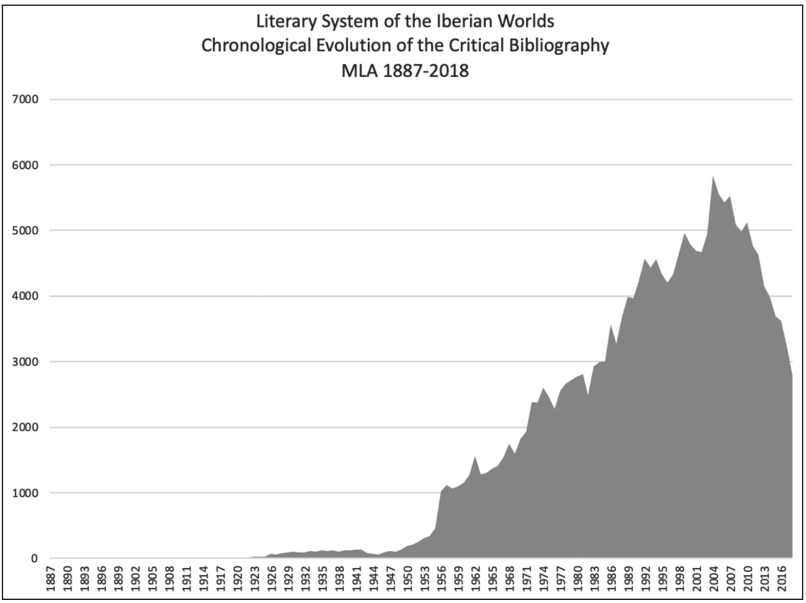


Figure 4: Chronological Evolution of the Critical Bibliography (MLA 1887–2018).

Table 1: The 25 Most Studied Writers of the Iberian Worlds (MLA 1887–2018).

National Literature	Writer	References	%LSIW
Spain	Cervantes Saavedra, Miguel de (1547–1616)	7211	3.4%
Argentina	Borges, Jorge Luis (1899–1986)	4372	2.1%
Spain	Vega Carpio, Lope Félix de (1562–1635)	2922	1.4%
Spain	Calderón de la Barca, Pedro (1600–1681)	2825	1.3%
Spain	García Lorca, Federico (1898–1936)	2498	1.2%
Spain	Pérez Galdós, Benito (1843–1920)	2359	1.1%
Spain	Unamuno y Jugo, Miguel de (1864–1936)	1878	0.9%
Colombia	García Márquez, Gabriel (1928–2014)	1822	0.9%
Argentina	Cortázar, Julio (1914–1984)	1723	0.8%
Spain	Quevedo y Villegas, Francisco Gómez de (1580–1645)	1550	0.7%
Spain	Valle-Inclán, Ramón María del (1866–1936)	1436	0.7%
Mexico	Fuentes, Carlos (1928–2012)	1275	0.6%
Nicaragua	Darío, Rubén (1867–1916)	1225	0.6%
Mexico	Paz, Octavio (1914–1998)	1183	0.6%
Chile	Neruda, Pablo (1904–1973)	1162	0.6%
Spain	Machado y Ruiz, Antonio (1875–1939)	1161	0.6%
Peru	Vargas Llosa, Mario (1936-)	1155	0.5%
Portugal	Pessoa, Fernando António Nogueira (1888–1935)	1089	0.5%
Cuba	Martí, José (1853–1895)	1088	0.5%
Spain	Ortega y Gasset, José (1883–1955)	1085	0.5%
Spain	Rojas, Fernando de (d. 1541)	1065	0.5%
Cuba	Carpentier, Alejo (1904–1980)	1022	0.5%
Mexico	Juana Inés de la Cruz (1648–1695)	1019	0.5%
Brazil	Assis, Joaquim Maria Machado de (1839–1908)	945	0.4%
Portugal	Camões, Luís Vaz de (1524/5–1580)	913	0.4%

Table 2: The 25 Writers with the highest percentage by National Literature (MLA 1887–2018).

National Literature	Writer	References	% NL	%LS IW
Paraguay	Roa Bastos, Augusto (1917–2005)	430	61.25%	0.20%
Nicaragua	Darío, Rubén (1867–1916)	1225	59.44%	0.58%
Timor-Leste	Cardoso, Luís (1958-)	4	33.33%	0.00%
Guinea-Bissau	Cabral, Amílcar (1921–1973)	21	32.81%	0.01%
Colombia	García Márquez, Gabriel (1928–2014)	1822	32.43%	0.87%
Mozambique	Couto, Mia (1955-)	145	32.37%	0.07%
Guatemala	Asturias, Miguel Ángel (1899–1974)	487	32.04%	0.23%
Equatorial Guinea	Ndongo-Bidyogo, Donato (1950-)	51	26.56%	0.02%
Argentina	Borges, Jorge Luis (1899–1986)	4372	24.05%	2.08%
Peru	Vargas Llosa, Mario (1936-)	1155	18.97%	0.55%
Uruguay	Onetti, Juan Carlos (1909–1994)	509	18.80%	0.24%
Panama	Jaramillo Levi, Enrique (1944-)	42	17.50%	0.02%
Dominican Rep.	Henríquez Ureña, Pedro (1884–1946)	110	16.52%	0.05%
Angola	Pepetela (1941-)	86	14.98%	0.04%
Chile	Neruda, Pablo (1904–1973)	1162	14.95%	0.55%
Honduras	Castillo, Roberto (1950–2008)	19	14.07%	0.01%
El Salvador	Castellanos Moya, Horacio (1957-)	73	13.88%	0.03%
Venezuela	Bello, Andrés (1781–1865)	277	13.59%	0.13%
São Tomé and Príncipe	Tenreiro, Francisco José (1921–1963)	5	12.82%	0.00%
Panama	Britton, Rosa María (1936-)	30	12.50%	0.01%
El Salvador	Alegría, Claribel (1924–2018)	64	12.17%	0.03%
Equatorial Guinea	Ávila Laurel, Juan-Tomás (1966-)	23	11.98%	0.01%
Cuba	Martí, José (1853–1895)	1088	11.80%	0.52%
Costa Rica	Vallbona, Rima de (1931-)	113	11.73%	0.05%
Peru	Vallejo, César Abraham (1892–1938)	691	11.35%	0.33%

Table 3: The 25 Most Studied Literary Works of the Iberian Worlds (MLA 1887–2018).

National Literature	Writer	Literary Work	References	%IW
Spain	Cervantes Saavedra, Miguel de	*Quijote* (1605, 1615)	4081	1.94%
Spain	Rojas, Fernando de	*La Celestina* (ca. 1499)	1072	0.51%
Colombia	García Márquez, Gabriel	*Cien años de soledad* (1967)	468	0.22%
Mexico	Rulfo, Juan	*Pedro Páramo* (1955)	304	0.14%
Portugal	Camões, Luís Vaz de	*Os Lusíadas* (1572)	267	0.13%
Argentina	Cortázar, Julio	*Rayuela* (1963)	215	0.10%
Peru	Garcilaso de la Vega, el Inca	*Comentarios reales de los incas* (1609)	190	0.09%
Paraguay	Roa Bastos, Augusto	*Yo el Supremo* (1974)	157	0.07%
Colombia	García Márquez, Gabriel	*Crónica de una muerte anunciada* (1981)	155	0.07%
Peru	Poma de Ayala, Felipe Huamán	*El primer nueva corónica y buen gobierno* (1615)	150	0.07%
Guatemala	Menchú, Rigoberta	*Me llamo Rigoberta Menchú* (1983)	145	0.07%
Cuba	Carpentier, Alejo	*Los pasos perdidos* (1953)	134	0.06%
Cuba	Carpentier, Alejo	*El reino de este mundo* (1949)	132	0.06%
Chile	Allende, Isabel	*La casa de los espíritus* (1982)	130	0.06%
Cuba	Lezama Lima, José	*Paradiso* (1966)	123	0.06%
Cuba	Cabrera Infante, Guillermo	*Tres tristes tigres* (1967)	118	0.06%
Guatemala	n.n.	*Popol Vuh*	111	0.05%
Colombia	Rivera, José Eustasio	*La vorágine* (1924)	109	0.05%
Chile	Bolaño, Roberto	*2666* (2004)	108	0.05%
Colombia	García Márquez, Gabriel	*El otoño del patriarca* (1975)	100	0.05%
Cuba	Villaverde, Cirilo	*Cecilia Valdés o la loma del ángel* (1839, 1882)	98	0.05%

Table 3 (continued)

National Literature	Writer	Literary Work	References	%IW
Peru	Arguedas, José María	*Los ríos profundos* (1958)	98	0.05%
Colombia	García Márquez, Gabriel	*El general en su laberinto* (1989)	92	0.04%
Colombia	García Márquez, Gabriel	*El amor en los tiempos del cólera* (1985)	88	0.04%
Colombia	Vallejo, Fernando	*La Virgen de los Sicarios* (1994)	88	0.04%

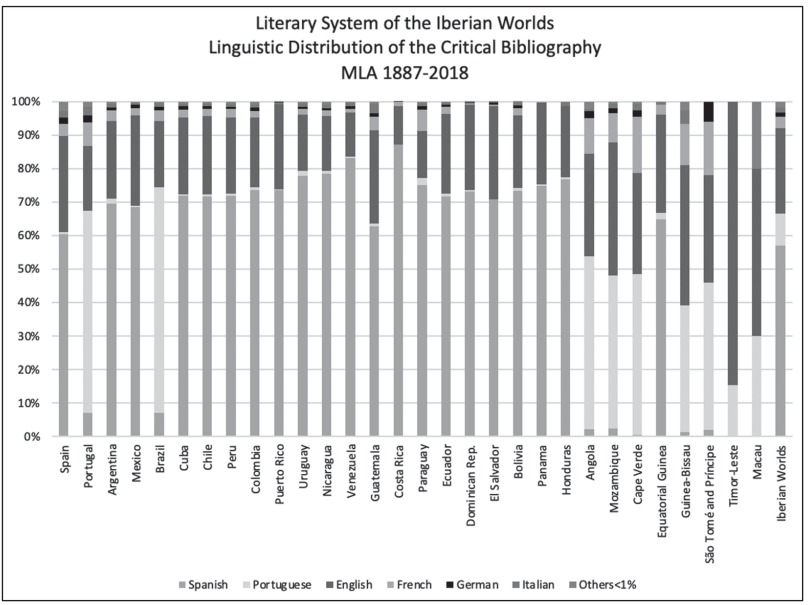


Figure 5: Linguistic Distribution of the Critical Bibliography (MLA 1887–2018).

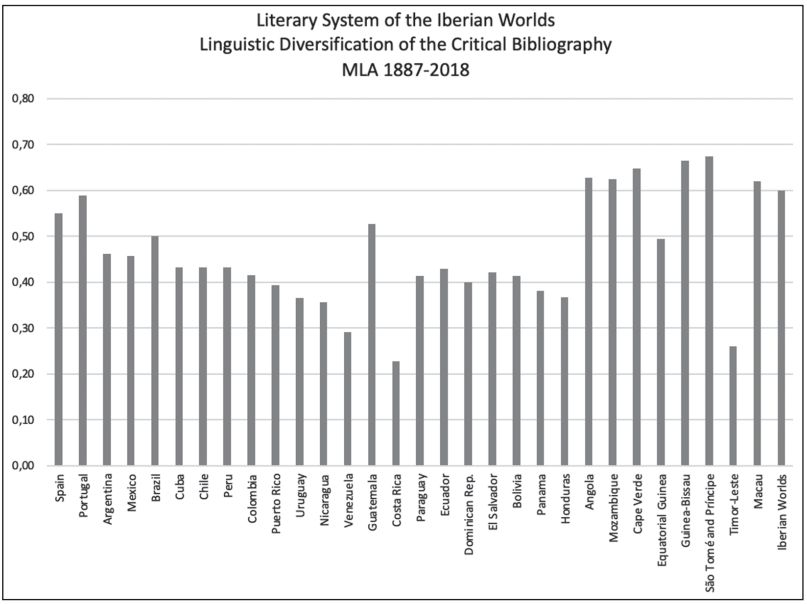


Figure 6: Linguistic Diversification of the Critical Bibliography (MLA 1887–2018).

4 Selected Writers

With the purpose of illustrating the relational analysis that can be carried out at the level of writers, we have selected a subset of them. For each continent, we chose the first man and the first woman whose language of expression is, respectively, Spanish and Portuguese. This way, we constituted a sample of 13 writers. Table 4 contains the essential metadata associated to them. There are 4 European writers, 4 from the Americas, 4 from Africa, and 1 from Asia. Their life period expands from the 16th to the 21st centuries. In fact, Cervantes, who lived in the 16th century, is the most ancient writer of the sample, followed by Juana Inés de la Cruz, who lived in the 17th century. Most of the other authors were born in the 19th or 20th centuries, with the writers from Africa and Asia being the youngest of the sample. Out of the 13 selected writers, 6 are women.

Figure 7 represents the number of references contained in MLA about each of these writers. We observe important differences in the number of references dedicated to each writer. Particularly, Cervantes and Borges concentrate very important volumes of publications. Also, in the cases of Spain and Portugal, there is

a striking disparity between men and women in the number of publications dedicated to them.

Figure 8 corresponds to the linguistic distribution of the publications about each writer. Except for the cases of Couto and Cardoso, whose bibliography is mostly in English, and Chiziane, whose bibliography is exactly distributed between Spanish and English, the main language of publication of the bibliography corresponds to the writing language of the authors. The bibliography that shows the greatest diversification is that related to Pessoa, and the least one is that on Cardoso.

The following 3 analysis refer to cocitations. For each writer, we selected the cocited authors that represent at least 1% of the bibliography about the analyzed writer. Figure 9 represents cocitations by gender. We immediately observe that women are rarely cocited. Likewise, we see that, except for Ndongo-Bidyogo and Couto, only women are cocited with other women. This indicator shows that women are underrepresented not only in terms of the number of publications about them, but they are also less frequently cocited.

In Figure 10, we have classified the cocited writers by continent. We note that Spanish writers are only cocited with other Europeans, whereas Pessoa and the writers of the Americas are cocited with writers from Europe and the Americas. Cardoso is co-cited with writers from Europe and Africa, while Ndongo-Bidyogo is cocited with writers from the four continents. Five writers are cocited with authors from Europe, the Americas, and Africa. Except for Cardoso, mainly cocited with Europeans, the most cocited writers belong to the same continent as the analyzed writer.

Figure 11 corresponds to the distribution of the writers cocited by national literature. We have classified the literatures into 3 categories: national literature, others of the literary system of the Iberian worlds, and extra systemic. Except for Cardoso, who is cocited exclusively with other writers from the Iberian worlds, each writer is cocited with at least one other author belonging to his/her own national literature. This percentage is particularly important in the cases of Cervantes, Pardo Bazán, Agustina Luís and Borges.

Thus, we observe that there are important differences in the patterns of cocitations of writers, depending on the level of development of their own national literature, as well as on the gender of the writer.

Table 4: Metadata about the Selected Writers (MLA 1887–2018).

Writer	National Literature	Life Period	Language	References	1st MLA Reference
Cervantes Saavedra, Miguel de	Spain	1547–1616	Spanish	7456	1888
Pardo Bazán, Emilia	Spain	1851–1921	Spanish	801	1926
Pessoa, Fernando	Portugal	1888–1935	Portuguese	1147	1955
Luís, Agustina Bessa	Portugal	1922–2019	Portuguese	72	1958
Borges, Jorge Luis	Argentina	1899–1986	Spanish	4667	1952
De la Cruz, Juana Inés	Mexico	1651?–1695	Spanish	1097	1926
Assis, Joaquim Maria Machado de	Brazil	1839–1908	Portuguese	958	1949
Lispector, Clarice	Brazil	1920–1977	Portuguese	630	1967
Ndongo-Bidyongo, Donato	Equatorial Guinea	1950-	Spanish	53	1992
Nsue Angüe, María	Equatorial Guinea	1945-	Spanish	20	1995
Couto, Mia	Mozambique	1955-	Portuguese	159	1987
Chiziane, Paulina	Mozambique	1955-	Portuguese	34	2000
Cardoso, Luís	Timor-Leste	1958-	Portuguese	6	2004

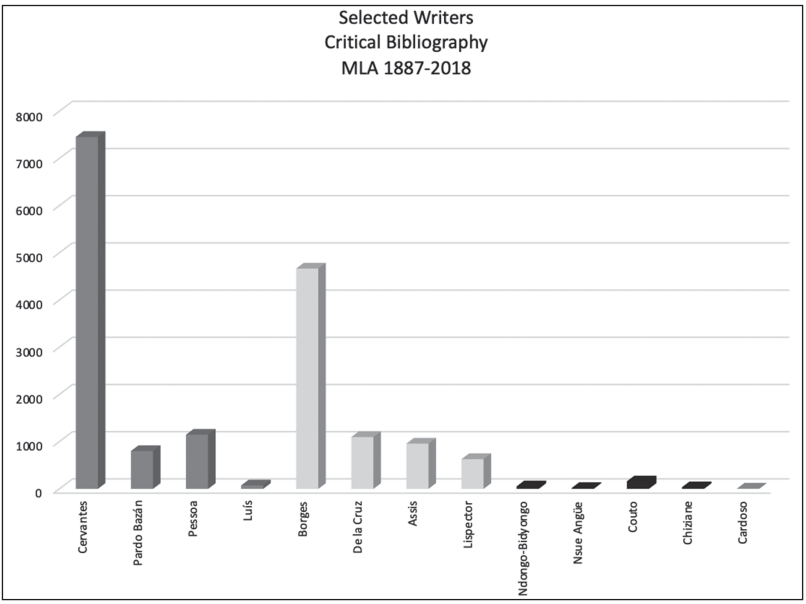
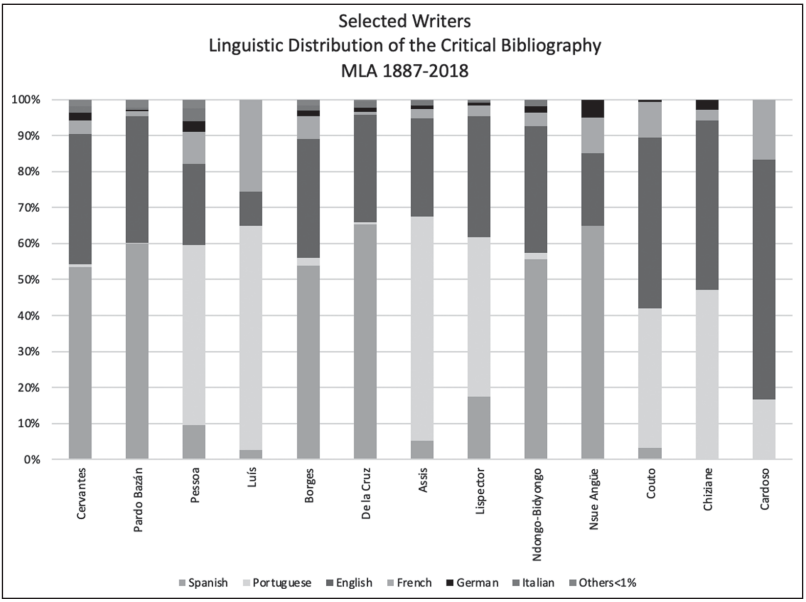


Figure 7: Critical Bibliography by Writer (MLA 1887–2018).



Figrue 8: Linguistic Distribution of the Critical Bibliography (MLA 1887–2018).

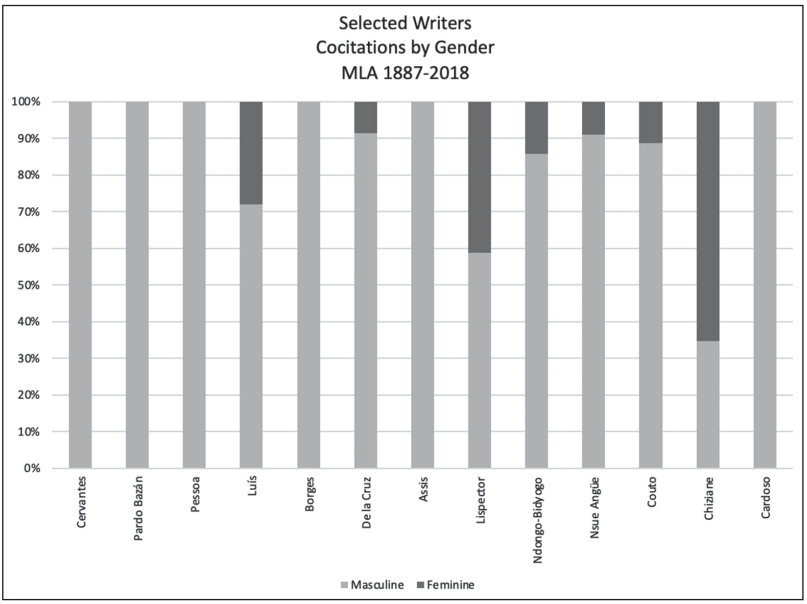


Figure 9: Cocitations by Gender (MLA 1887–2018).

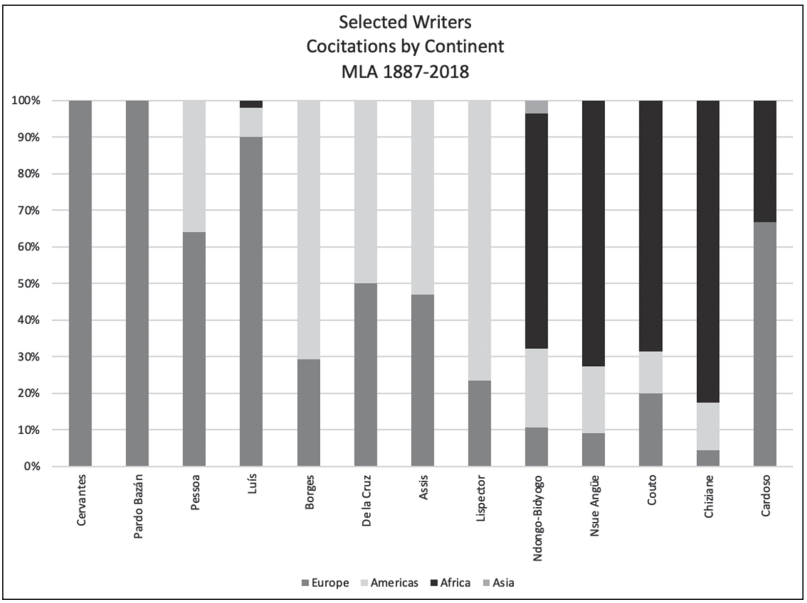


Figure 10: Cocitations by Continent (MLA 1887–2018).

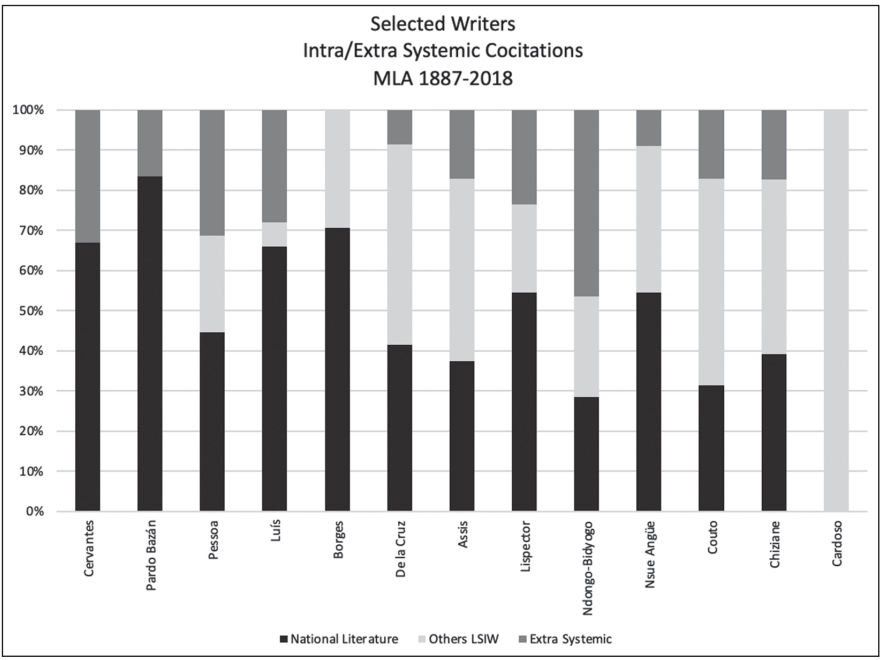


Figure 11: Cocitations by Region (MLA 1887–2018).

5 Conclusions

The analysis of the literary system of the Iberian worlds has revealed the complex relationships between 30 literatures located in 4 different continents that share Spanish or Portuguese languages as well as several historical events. In terms of national literatures, we have seen very significant disparities regarding the conentration of publications dedicated to them. We note that Spain occupies a central place in the system, despite the important evolution of other national literatures, such as those of Argentina, Mexico, Portugal, and Brazil. At the same time, we observe that some literary subsystems are still at embryonic levels of development. This is essentially the case of the Iberian national literatures found in Africa and Asia.

Regarding the writers, we were able to identify certain figures who occupy an important place at the level of the literary system of the Iberian worlds, while others are only recognized at the national level.

Additionally, we elaborated indicators to measure linguistic diversification and identified the presence of critical publications in languages that do not belong to the system of the Iberian worlds, mainly English. In some cases, where

English concentrates important percentages of the bibliography, this could be a sign of foreign and even extra systemic interference.

Furthermore, thanks to the analysis of a subset of writers, we observed that cocitations are much more important in critical bibliographies that refer to writers from the Americas and Africa. Several indicators also allowed us to corroborate the marginal place that academic literary criticism attributes to women writers.

Finally, we consider that this study has revealed that, through the exploitation of bibliographic databases and the introduction of quantitative methods in literary studies, such as criticometrics, it is possible to increase our knowledge regarding the configuration of world, continental, and regional literary systems, and the relationships between them.

Bibliography

Bourdieu, Pierre (1992): *Les règles de l'art. Genèse et structure du champ littéraire*. Paris: Seuil.

_____ (1997): *Méthodes pascaliennes*. Paris: Seuil.

Bourdieu, Pierre et Loïc JD Wacquant (1992): *Réponses. Pour une anthropologie réflexive*. Paris: Seuil.

Boyd, Danah, and Kate Crawford (2012): "Critical Questions for Big Data Provocations for a Cultural, Technological, and Scholarly Phenomenon", In *Information Communication & Society*, vol. 15/n 5, pp. 662–679.

Callon, Michel, Jean-Pierre Courtial et Hervé Penan (1993): *La Scientométrie*. Que Sais-Je? 2727. Paris: Presses Universitaires de France.

Clarivate. <https://clarivate.com>.

Cozzens, Susan E (1985): "Using the Archive – Price, Derek Theory of Differences among the Sciences", In *Scientometrics*, n 7, pp. 431–441.

Damrosch, David (ed.) (2014): *World Literature in Theory*. Malden: Wiley Blackwell.

Dissertation Abstracts International. <http://proquest.com>.

Even-Zohar, Itamar (1990): "Polysystem Theory", In *Poetics Today*, n°11, pp. 1–268.

Ferrer, Carolina (2011): "El boom hispanoamericano: del texto a la pantalla", In Santiago Juan-Navarro y Joan Torres-Pou (eds.), *Nuevas aproximaciones al cine hispánico: Migraciones temporales, textuales y étnicas en el bicentenario de las independencias iberoamericanas (1810–2010)*. Barcelona: Promociones y Publicaciones Universitarias, pp. 79–101.

_____ (2014): "Difusión de la bibliografía crítica sobre la literatura hispanoamericana a través de las revistas académicas 1896–2008", In Álvaro Baraibar (ed.), *Visibilidad y divulgación de la investigación desde las humanidades digitales. Experiencias y proyectos*. Pamplona: Servicio de Publicaciones de la Universidad de Navarra, pp. 289–305.

_____ (2018): "Les études littéraires à l'ère de la mondialisation : traces et trajets au prisme des nouveaux observables numériques", In *Zizanie*, vol. 2/n°1, pp. 76–101.

_____ (2019): "La place des écrivaines dans la littérature mondiale: cartographie métacritique de la marginalisation des femmes", In *Pontos de interrogação : revista de crítica cultural*, vol. 9/n°2, pp. 13–33.

_____ (2022): “De Confucius à Djebar, de Dante à Lispector. Ce que la criticométrie nous apprend sur la réception des écrivains et de leurs œuvres”, In Iona Galleron et Fatiha Idmhand (dirs.), *Dix ans de corpus d’auteurs*. Paris: Éditions des archives contemporaines, pp. 245–267.

Gupta, Suman (2009): *Globalization and Literature*. Cambridge: Polity.

Han, J., Kamber, M., & Pei, J. (2012). *Data mining. Concepts and techniques*. Waltham: Morgan Kaufmann.

Herfindahl, Orris C. (1950): *Concentration in the U.S. Steel Industry*, Unpublished doctoral dissertation, Columbia University.

Hirschman, Albert O. (1945): *National Power and the Structure of Foreign Trade*, Berkeley, University of California Press.

Leydesdorff, Loet (1998): “Theories of Citation?”, In *Scientometrics*, vol. 43/n 1, pp. 5–25.

Luhmann, Niklas (2000): *Art As A Social System*. Stanford: Stanford University Press.

Mayer-Schönberger, Viktor and Kenneth Cukier (2013): *Big data. A revolution that will transform how we live, work, and think*. Boston and New York: Houghton Mifflin Harcourt.

Modern Language Association International Bibliography. <www.mla.org>.

Price, Derek de Solla (1963): *Little Science, Big Science*. New York, Columbia University Press.

Rae, Douglas W. (1967): *The Political Consequences of Electoral Laws*, New Haven, Yale University Press.

Ruiz Ibáñez, José Javier y Óscar Mazín Gómez (2021): “Una presentación de *Los mundos ibéricos*”, In *Otros Diálogos del Colegio de México*, n°15, pp. 1–12, https://otrosdialogos.colmex.mx/una-presentacion-de-los-mundos-ibericos, consulted on 08-12-2021.

Saussy, Haun (dir.) (2006): *Comparative Literature in an Age of Globalization*. Baltimore: Johns Hopkins University Press.

Schreibman, Susan, Ray Siemens and John Unsworth (eds.) (2004): *A Companion to Digital Humanities*. Malden: Blackwell.

Schreibman, Susan, Ray Siemens and John Unsworth (eds.) (2016): *A New Companion to Digital Humanities*. Malden: Wiley and Sons.

Small, Henry (1973): “Co-citation in the Scientific Literature: A New Measure of the Relationship Between Two Documents”, In *Journal of the American Society for Information Science*, n 24, pp. 265–269.

Thomson Reuters Web of Knowledge. <http://www.isiwebofknowledge.com/>.

Van Rees, Kees (1997): “Modelling the Literary Field: From System-Theoretical Speculation to Empirical Testings”, In *Canadian Review of Comparative Literature/Revue Canadienne de Littérature Comparée*, March/mars, pp. 91–101.

Wallerstein, Immanuel (2004): *World Systems Analysis: An Introduction*. Durham: Duke University Press.

Witten, I. H., Frank, E., and Hall, M. A. (2011): *Data Mining. Practical Machine Learning Tools and Techniques*. Waltham: Morgan Kaufmann.

Diana Roig-Sanz, Alessio Cardillo & Ventsislav Ikoff

Global Translation Flows in Ibero-American Periodicals: A Network Science Perspective

1 Introduction

Literary, translation and art history, comparative literature, and intellectual history have been featured in the last two decades by a data-driven perspective and by the purpose of combining both qualitative and quantitative methodologies. Likewise, global studies have faced the idea of connectivity and movement as a real challenge. In this regard, researchers are still struggling to analyze the existence (or lack) of relations, flows, circulation and mobility, concepts that can shed light into processes of cultural, political, economic, or social transformation. The role of connectivity and the relevance of networks as the emerging form of social organization has been at the core of fundamental works (Castells 1996) and historical questions are increasingly analyzed in terms of network analysis, mathematical modeling and visualization techniques. Cultural phenomena related to the concepts of centrality and periphery also arise. Certainly, network science has put social relations and the study of cultural mediators at the center and important endeavors such as the platform Historical Network Research or the *Journal of Historical Network Research* emerged. They provide training, workshops, lectures, research bibliography and an open-access journal to a wide community that is now internationally oriented and work in fields and geographical areas less prone to this approach, including Ibero-America. Many researchers work in large-scale contexts and share this interest for the analysis of global connections and entangled histories (Middell and Naumann 2010; Berg 2013, Conrad 2016, Rotger, Roig-Sanz and Puxán 2019), but this is not yet widespread in all domains, academic traditions, and time periods (Liu 2018). For example, cultural analytics (Manovich 2015, 2020) and knowledge data discovery (Meyer and Schroeder 2015; Borgman 2015) have not been applied sufficiently in many non-European contexts to test assumptions on literary value, institutions, or the position of cultural producers in the cultural field, or to reassess the role of many actors. These shortcomings can be due to the lack of structure and digitalization of many sources and archives (Algee-Hewitt et al. 2016) in non-European contexts for a data mining approach. But also because of the fact that previous research on world literature has placed most of these actors in relation to their "peripheral" origins, or in a

Note: This research is framed within the ERC StG project Social Networks of the Past. Mapping Hispanic and Lusophone Literary Modernity, 1898-1959 (Grant Agreement: 803860) led by Diana Roig-Sanz.

https://doi.org/10.1515/9783110753523-011

subjugated relation to the center or to the empire. Therefore, the idea of the network as a mere metaphor to describe the existence of relationships among people or objects is no longer sufficient to address the complexity of the information encapsulated in big amounts of data (Carbó-Catalan and Roig-Sanz 2021).

Within this framework, this paper aims at offering some insights regarding methodological issues and practical applications when applying big data to disciplines in the humanities (Schäfer, and van Es 2017). Specifically, it aims at contributing to the analysis of circulation and global translation flows within a big translation history approach (Roig-Sanz and Fólica 2021; Fólica 2021) and a relational perspective (Ashrafi, Hashemi, and Akbari 2019). We advocate for global translation zones, which can be understood as a space of translation that is constituted upon the following criteria: a geographical scale (human and political, but also physical: the Andean mountains, Río de la Plata or the Caucasus), a time and a historical dimension (historical channels of translation), and in terms of agency and networks (Roig-Sanz and Kvirikashvili forth.). This means publishing zones (agreements between publishers, specific languages and literatures, and literary magazines) and circuits of soft power (the role of national or regional institutes in inter-peripheral translation flows or in the emergence of a translation policy). We claim that global translation zones must be explored in the *longue durée* and in the framework of a complex and multilingual history that cannot be overshadowed. In this respect, we aim at addressing the following research questions in the field of translation and literary periodicals in Ibero-America (Fólica, Roig-Sanz and Caristia 2020) in the first half of the twentieth century: 1) What is the level of internationalization of these journals? Which literatures and authors will be translated and circulated? What is the geographical distribution of authors and languages in relation to what we denote as global translation zones? 2) How do we analyze global translation flows through the lenses of network science?, and 3) What profiles can be found in relation to translators and writers and, specifically, to women translators or women writers if we apply a gender perspective?

To this aim, this paper will analyze the literary translations which were published in a corpus of literary magazines already digitized to unearth and restore a less-canonized translation history that is often overshadowed. The research is based on a dataset of contributions published in 42 modernist and avant-garde periodicals from Spain and Hispanic America between 1891 and 1936, cataloged and published by Ehrlicher (2020).[1] In the dataset, a contribution is considered any tex-

1 The full list of magazines cataloged for the dataset can be consulted in Ehrlicher (2020): Spanish-language Cultural Magazines from Modernismo to Avant-Garde: Processes of Modernization and Transnational Network Formation. Revistas culturales históricas en lengua española desde el modernismo hasta las vanguardias: procesos de modernización y formación de redes transnacionales. 0_Corpus-Overview.pdf. Accompanying Publications. DARIAH-DE. doi:10.20375/0000-

tual or graphical piece published in the journal. In this respect, our results are not exhaustive, yet they can highlight patterns of translation and circulation, as well as translation practices. While one of the principles of creating this corpus has been geographical variety and representation across Spain and Hispanic America, it has to be noted that more than half of the records are from Spanish periodicals, mostly published in Madrid. It is also worth reminding that the magazine as an object of research had a key role in the construction of a collectivity, as it offered a space for collaboration, exchange and collective projects. From a researcher's point of view, it offers the opportunity to trace networks of people and networks of translated literatures, writers, publishers, and translators. To this end, we apply a network science perspective (Barabási 2011) to explore and visualize metadata extracted from these magazines (Fólica, Ikoff, and Roig-Sanz 2018; Fólica, Roig-Sanz and Caristia 2020; Roig-Sanz and Fólica 2021; Lehmann and Ehrlicher 2022).

2 Data: Definition and Dimensions

The dataset contains about 31,500 data records (rows), each representing a contribution and one corresponding contributing author,[2] providing descriptive data such as the title and genre of the contribution, the contributor (author), date of publication, publication language, and for some entries, also the translator and/or the original language. Overall, the dataset is made of 31,500 records of contributions in 26 different languages, corresponding to 4,551 authors and 266 translators from 58 different countries. Table 1 (please see below) provides a summary of the composition of the dataset according to the place of publication, as well as what we have called global translation zones (please see the definition above). In the case of Ibero-America, we have highlighted the following global translation zones: the Iberian Peninsula, Río de la Plata, Andes, the Caribbean, Mexico, which has specific dynamics in relation to publishing, translation flows and the marketplace, and Other, which refers to Spanish-speaking journals published outside Ibero-America (for example, in France). As the magazines in the corpus differ from one another in terms of longevity, format, time of publication and volume, the properties of the corpus

000d-1d02-1. The ERC project Social Networks of the Past. Mapping Hispanic and Lusophone Literary Modernity, 1898–1959 has established a data sharing agreement with Prof. Hanno Ehrlicher from the University of Tübingen. It's also worth mentioning that translation is not at the core of his research, so this chapter aims at filling this gap.

2 This means that a contribution with multiple authors results in multiple records, one for each author, as explained by the authors of the dataset in the accompanying publications (Herzgsell 2020).

are quite heterogeneous. This means that the corpus is made by both periodicals with a high number of contributions, as well as periodicals with only a few of them. Moreover, it has to be said that although most magazines were fully cataloged, there are six for which this was done only partially, due to either the unavailability of the magazine as a digital object, or to limited resources for the cataloging task (Ehrlicher 2020).

Table 1: Number of journals, issues and records per global translation zone and city of publication.[3]

Region	Place of Publication	Magazines	Issues	Records	Records, %
Iberian Peninsula	Madrid	13	448	14276	45.32%
	Barcelona	4	39	1144	3.63%
	Sevilla	1	50	1063	3.37%
	A Coruña	1	11	348	1.10%
	Málaga	2	9	181	0.57%
	Santander	1	5	71	0.23%
TOTAL		**22**	**562**	**17083**	**54.24%**
Río de la Plata	Buenos Aires	4	174	5545	17.60%
	Montevideo	1	2	57	0.18%
TOTAL		**5**	**176**	**5602**	**17.79%**
Andes	Lima	1	32	1569	4.98%
	Santiago de Chile	2	47	998	3.17%
	Puno	2	35	420	1.33%
TOTAL		**5**	**114**	**2987**	**9.48%**
Caribbean	La Habana	3	90	2884	9.16%
TOTAL		**3**	**90**	**2884**	**9.16%**
Mexico	Mexico City	4	170	2382	7.56%
	Jalapa	1	10	488	1.55%
TOTAL		**5**	**180**	**2870**	**9.11%**
Other	Madrid-Paris	1	3	44	0.14%
	Paris	1	2	27	0.09%
TOTAL		**2**	**5**	**71**	**0.23%**
TOTAL corpus		**42**	**1127**	**31497**	**100.00%**

3 The dataset does not include magazines from Andean countries such as Ecuador or Bolivia, as well as Caribbean countries such as Puerto Rico and Venezuela. We hope to fill this gap in the future.

As we are interested in translation, i.e., textual contributions, we take into account all the records except for those classified as "image," leaving us with approximately 81% of all the records. Moreover, we also discard those records whose author is either unidentifiable (e.g., "R.A.C.") or does not correspond to a person (e.g., "Redacción"), as our research is mainly interested in authors/writers, translators or cultural mediators overall (Roig-Sanz and Meylaerts 2018). To this aim, we manually curated a list of 296 "bad authors' names" to exclude them from our corpus, leaving us with only 20,663 records to analyze, corresponding to the 65.51% of the raw data. The list of "bad authors' names" is available online[4]. On the other hand, as we are also mostly interested in translation, we have designed a criterion to discriminate those records corresponding to translations from the rest of them. According to such criterion, data record can be considered a translation either if the values of the "*original language*" and "*publication language*" fields are different (e.g., English and Spanish), or if the "*translator*" field is not empty (e.g., "Maseras, Alfons" or "Villaurrutia, Xavier"). By applying this criterion we split the dataset into a subset of "certified translations" (made of 981 records), and a subset containing all the other records (19,682 records).

3 The Internationalization of Ibero-American Literary Magazines Through Translation

Global translation flows and the international circulation of literature has played a historical role in the institutionalization of national cultures (Thiesse 2001; D'Hulst 2012), especially in contexts characterized by a significant backwardness compared to other spaces, or in contexts historically considered as "peripheral" despite being at the forefront of many innovative literary or translation projects. As we know, for translation in periodicals many important works were published first in journals and literary magazines, the latter being especially important in the Ibero-American context. Journals were the center of new movements, trends, and intellectual debates related to any issue in relation to culture. They also made visible the national (or international) recognition of an author, or a new literary genre being discussed in this media. At the same time, journals were an essential means by which literary and artistic groups staged public appearances, and connections among them reinforced their mediating role between the global, the regional, and the local scale of literary knowledge.

4 The complete list of "bad authors' names" is available at: https://cardillo.web.bifi.es/datasets/translation-list_bad_names.txt.

Within this general framework, this section seeks to analyze and measure the internationalization of Ibero-American literary magazines through translation with a twofold aim: 1) to analyze which kind of books, literatures, and authors were channeled through translation, as well as which is the geographical distribution of authors and languages in relation to what we understand as global translation zones, and 2) to find out how literary magazines worked on a global scale and set up a dialogue with other international periodicals. We can already confirm that Spanish-speaking literary magazines participated in a cosmopolitan community (Gramuglio 2013), and followed contemporary trends, which legitimated them as modern institutions in both Europe and the world. In Spain, literary journals also fostered regional identities (the Catalan *La Revista* or the Galician *Nos*) and also aimed to connect with the transatlantic world (the Galician and Uruguayan *Alfar*). Indeed, it has to be said that the international circulation of literatures in translation has pushed forward the emergence of the first public translation policies to standardize translation practices (Carbó-Catalan 2022). It has also helped to promote translation and cultural development (Pegenaute 2018) and to institutionalize the profession of the translator. Translation in periodicals also encouraged the inclusion of non-European writers in an international network of culture.

In Latin America, cosmopolitan purposes also coexisted with national cultural projects (Fólica 2022), indigenism, and revolutionary ideals (the Peruvian *Amauta*). In that respect, modernist journals contributed to enhance a newly stratified Ibero-American market by conferring prestige and symbolic capital to writers eager to reach an international audience. Apart from the cultural projection of their own countries abroad, writers with diplomatic careers also made a living as authors or translators, increasing their literary prestige. Translations in magazines were also crucial for the popularization and democratization of culture, and they acted as a shared space for both the dissemination of the various national literary and aesthetical projects, and for the increasing cosmopolitanism of Spanish-speaking literatures in the international literary field, (Willson 2004).

In this respect, our research shows that international contact, as well as the intensification of connections among various countries and global translation zones around the world throughout the period that some authors have called the first globalization (Rosenberg 2012) affected multiple realms in the social field, including the cultural one. These connections had close ties to the reinforcing of national identities. Indeed, while the national space provides a scale of reference, the analysis of translation flows in Ibero-American periodicals actually includes many scales: the local, national, regional, and global (Bender 2006). In this sense, translated literature constitutes a privileged object with which to view the selection of various authors that multiple editors made within a myriad of magazines. Simultaneously, they also cast light on the construction of a literary and transla-

tion canon – which was not homogenous across the region, despite sharing certain similarities in some cases. Thus, translation emerges as a way of contributing to the construction of the nation, but is also a way of positioning the nation within the world through the Other. Works by Thiesse (2001) and Wilfert (2020) have demonstrated the complexity of this relationship – and we may add that, through translation, a set of texts can become cultural heritage. In this sense, the magazines in our corpus show how these publications situate themselves within an international debate that cannot be taken separately from the many battles for legitimacy and consecration taking place on multiple scales.

In what follows, we offer some quantitative insights that help us give an overview of the cultural phenomena we just briefly described above. Within a distant reading approach (Moretti 1998, 2005, 2013), we have identified in our corpus the presence of literary translations and the first observation that comes out is that the fraction of contributions explicitly labelled as translations – identified via either the name of the translator or the original language – tends to be small, given that recognizing a contribution as translated or acknowledging the translator was not common practice. For instance, in our dataset of the 31,497 records, 10,864 belong to unidentifiable authors. Thus, translations in our dataset represent a small proportion of all publications, about 4.6% across the six global translation zones that we have defined. As displayed in Figure 1, the most frequent languages are Spanish, French, Catalan, English, German, Italian, Portuguese, and Galician.

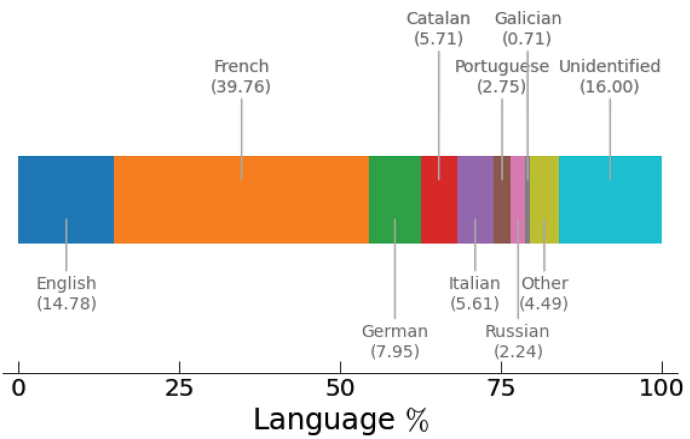


Figure 1: Fraction of records according to their original language. We report the eight most translated languages in our dataset. We also display the fraction of records whose language is not among the most translated ones (Other), or it is not available (Unidentified).

If we analyze the nationalities of authors' writing languages different from Spanish (Figure 1), we notice the presence of a non-negligible fraction of Italian, Catalan, and Galician authors. In fact, the role of the Galician, Catalan, and Italian migrant communities in the Río de la Plata's region is fundamental and multilingual practices appear in those magazines. In this respect, it is important to acknowledge that this research has only considered as translations those texts with information on the translator or the original language, but for the purpose of identifying translations which are not explicitly marked as such in the periodical, the idea that the authors' nationality or country of birth may be used to infer the language of their works does not always stand up to scrutiny. For instance, we are aware that some authors with foreign names were able to express themselves in languages other than their mother tongue. Thus, we are legitimate to believe that some of these authors wrote in Spanish as a consequence of people's transnational movements. This is the case of the French national Paul Groussac, director of *La Biblioteca*, and director of the Argentinean National Library between 1885 and 1929, or Israel Zeitlin, who also lived in Argentina and wrote in Spanish despite being from Russian origins. Hence, in our corpus the contributions of these authors are considered as normal contributions, not translations.

If we now examine the place of origin of all foreign authors published or translated in the global translation zones that we have proposed for this chapter (i.e., Andes, Caribbean, Río de la Plata, Mexico, Iberian Peninsula, and Others), the results suggest significant differences between them (see Table 2 below). For the Río de la Plata, about 62% of the translations were made from French, mostly from authors of French origin (e.g., Henri Barbusse or Romain Rolland). However, there are also French-Latin American authors such as Jules Supervielle (French Uruguayan), or other authors with East European origin. If we examine the number of articles by authors identified as foreign in a small sample of four Argentine magazines (*La Biblioteca*, *Proa*, *Martín Fierro*, and *Claridad*), we already see (Figure 2, see below) that most of them come from France, Russia, and Italy. Of the 131 translations identified, 80% of them belong to French, Russian, and Italian authors, relegating English-speaking authors (i.e., those from the United States, United Kingdom, Ireland, and Canada) and Portuguese language authors to marginal roles.

Similarly, literary magazines from the Iberian Peninsula and the Andes also contain many translations from French. About 50% in the case of the Iberian Peninsula with authors like Émile Zola, Paul Verlaine, or Guillaume Apollinaire being some of the most translated. In the Andes, translations from French (e.g., Panaït Istrati, or Paul Verlaine) just about edge those from English (e.g., the American writer Waldo Frank or the British Romantic poet P. B. Shelly), and Russian (e.g., Anatoli Lunacharsky or Isaac Babel) languages. Overall these three groups make

Table 2: Distribution of languages across global translation zones. For each translation zone, we report the percentage of records written in each of the three most common original languages in each region together with the same quantity computed for all other languages (Other languages).

Region	1st Language	2nd Language	3rd Language	Other Languages
Iberian Peninsula	French	English	German	
	50.28	12.01	9.19	28.52
Andes	French	English	Russian	
	26.09	21.74	19.57	32.61
Rio de la Plata	French	Italian	English	
	61.97	11.27	8.45	18.31
Caribbean	English	French	Catalan	
	47.12	26.92	9.62	16.35
Mexico	French	English	Italian	
	46.15	41.03	7.69	5.13
Other	French	–	–	
	100	0	0	0

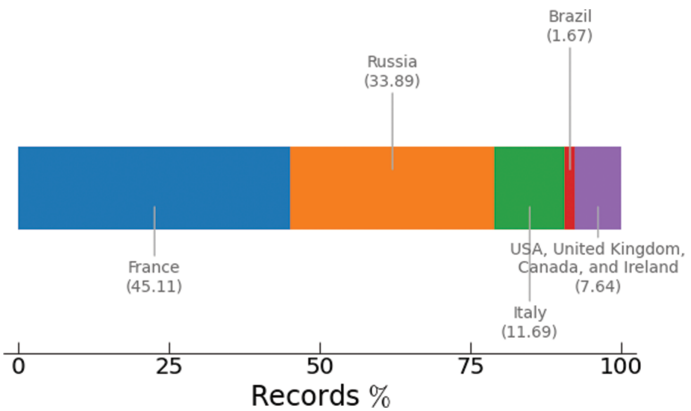


Figure 2: Percentage of records by authors belonging to certain foreign countries, published in four Argentine magazines (La Biblioteca, Proa, Martín Fierro, and Claridad).

up 67.4% of the whole translations, leaving a 32.6% of translations to other languages, thus suggesting a greater variety of translated literatures. These examples also show an interest in the Russian Revolution (e.g., the short stories by Babel or the works by Lunacharsky), the Soviet Union and communism (e.g., Barbusse, Frank), but also the fight against fascism and their interest in pacifism (e.g., Roll-

and). We can also identify some authors writing in socialist periodicals (e.g., Istrati) and Nobel Prize nominated or winning authors (e.g., Zola, Supervielle, or Rolland).

In the Caribbean region the most translated authors proceeded from English speaking countries (47%), mainly from the US (for example, the writer and journalist Christopher Morley) and the United Kingdom (e.g., Rudyard Kipling and the philosopher Bertrand Russell being the most translated). The second largest group, corresponding to 27% of the records, is made of a set of authors from France with no more than a couple of contributions each.

Finally, Mexican magazines published mainly translations from French (46%) and English (41%). In the case of French, we have identified translations from Guillaume Apollinaire, Jean Cocteau, and Paul Éluard, but also from Valéry Larbaud, Paul Morand, and the French Mexican author Ramon Fernández. Other writers such as the French philosopher in science and religion Emilio Boutrox, Louis Farigoule (better known as Jules Romains), Émile Salomon Wilhelm Herzog (better known as André Maurois), Alexis Leger (better known as Saint-John Perse), Maxime Leroy, Gaston Sevrette, and Adolphe Ferrière also appear. In the case of English, it is worth highlighting the translations of William Blake, David Herbert Lawrence, John Masefield, Alice Meynell, Nathan Asch, Aaron Copland, T.S. Eliot, John Gould Flecher, Waldo David Frank, Langston Hughes, Edgar Allan Poe, Dorothy Schons, Carl Van Doren, and Thornton Wilder. As it happened in other Ibero-American global translation zones, Mexican literary magazines were interested in Romantic writers such as Blake and Poe, avant-garde authors and, more specifically, cubist and surrealist writers such as Apollinaire, Cocteau, and Éluard. However, they were also interested in other modernist literary movements such as the Anglo-American imaginism of the poet John Gould Flecher, or the French imagist Paul Morand. The analysis of translation flows in Mexican literary magazines also gives a dominant place to poetry (e.g., Apollinaire, Cocteau, Éluard, Blake, Flecher, Eliot). Unlike magazines in the Andean mountains, Mexican periodicals also translated the writings of women who were close to the feminist mouvement. For example, Alice Meynell (a British writer, publisher, critic and suffragist) and Dorothy Schons, who authored the first English-language novel on the Mexican Sor Juana Inés de la Cruz placing her as one of the earliest American feminists. There was also an interest in the translation of works by authors close to the Communist party (e.g., Paul Éluard and Ramon Fernández). We can also stress some interest in philosophy (e.g., the French Émile Boutroux), education (e.g., the Swiss Adolphe Ferrière) and music (a translation into Spanish of an essay originally written in English by the American composer Aaron Copland on the Mexican composer Carlos Chávez).

Finally, we perform the analysis of the overall flows of translation occurring between countries, languages, and regions. To this aim, Figure 3 portrays the flows of translations via a so-called Sankey diagram[5] (Wilke 2019).

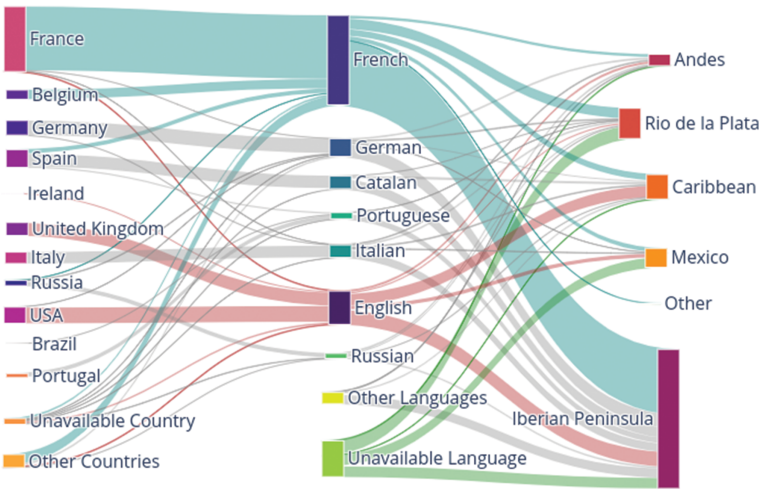


Figure 3: Overview of the flux of translations between countries, languages, and regions. This Sankey diagram displays the number of translation records between the authors' countries of origin (left column) and the most common original languages (middle column), as well as between the original languages and the translation regions (right column). The height of each box denotes its total number of records. In this diagram, each box corresponds to some feature, and a line connecting two boxes indicates the existence of a "flux" between them. We highlight the translation flows involving the English, French, and Unavailable languages.

The size of a box denotes the amount of elements with that feature, whereas the thickness of the line connecting two boxes denotes the amount of elements possessing both features. In our case, translated contributions can be classified according to three features: the author's country of origin, the contribution's original language, and the global translation's zone where the translation has been published. The boxes corresponding to each of the aforementioned features' groups are aligned horizontally, with the left column corresponding to the countries of origin, the middle column to the original languages, and the right column to the translation's regions. For simplicity, we display only the most important languages/countries and group

5 Sankey diagrams are named after Irish Captain Matthew Henry Phineas Riall Sankey, who used this type of diagram in 1898 in a classic figure showing the energy efficiency of a steam engine (see https://en.wikipedia.org/wiki/Sankey_diagram).

all the other languages/countries into the "Other Languages" and "Other Countries" categories. The "Unavailable Language" and "Unavailable Country" boxes correspond, instead, to those records for which information on the original language or author's country of origin are missing. Thus, a quick analysis of Figure 3 reveals that French and English are the two most translated languages, with the former constituting the bulk of our corpus. Also, French is the main translated language in the magazines of the Iberian Peninsula, albeit it plays an important role also in the region of Rio de la Plata. Authors writing in French come mainly from France but, we also observe the presence of other countries like Belgium, and a non-negligible fraction of authors from Spain and the "Other Countries." Regarding the English language, authors writing in such a language are mostly from the USA and the United Kingdom. Interestingly, the Sankey diagram highlights how English plays a significant role for translation only in journals of Caribbean and Mexico zones as we have also highlighted above with specific examples of translated authors, while it is more marginal in all the other regions (including the Iberian Peninsula).

Likewise, we want to highlight the case of contributions for which the information on the original language is missing (i.e., those corresponding to the "Unavailable Language" box). These contributions have been identified as translation due to the availability in their records of the information on the translator's name. They constitute a significant fraction of translations made within the Mexico and the Rio de la Plata regions and highlight the importance of identifying translations using a criterion based not exclusively on the analysis of the languages' information.

Finally, indigenous literary production was also present in Ibero-American journals. In the case of indigenous languages, we can highlight the work by Eustaquio Rodríguez Aweranka and Inocencio Mamani. *Tres Poemas* by Rodríguez Aweranka, Mamani and Manuel Zúñiga Camacho Allqa were published in Quechua and Aymara in issue 34 (1930) of *Boletín Titikaka* (1926–1930) to honor José Carlos Mariátegui, who had recently passed away. In the case of poetry in vernacular languages, we also find texts in the Quechua original, accompanied by a Spanish translation in the case of Mamani (editorial, *Boletín Titikaka*, num. 19, in Quechua and Spanish; also in num. 27 in Quechua and Spanish) and Aweranka (*Boletín Titikaka*, num. 32, in Quechua and Spanish). However, we should not consider them as translations, as Rodríguez Aweranka, Mamani and Zúñiga Camacho Allqa were authors themselves (poets and playwrights), and Spanish was their second mother tongue. Indeed, it is worth mentioning that the number of records for Quechua-Spanish translations is very low: the already mentioned 3 records over more than 31.000 entries. Yet, such types of cases are relevant as they highlight the presence of indigenous languages and multilingualism in Ibero-American literary magazines. Moreover, the specific case of *Boletín Titkika* is a great example to see how to combine the vernacular and the cosmopolitan.

4 Global Translation Flows in Ibero-America Through the Lenses of Network Science

One way to measure the role played by each magazine/region for translation is to analyze the data through the lenses of network science: an emerging discipline mixing together the methods of graph theory, mathematical modeling, physics, computer science, and statistics (Barabasi 2011). For this reason, we represent the data as *bipartite networks*, and analyze their properties.

A bipartite – or, more in general, *multipartite* – network/graph is a network having two (or $m > 2$) kinds of vertices/nodes, and in which an edge can exist only between vertices of different kinds (Latora 2017). For instance, in our case vertices can be either authors or magazines, and an edge connecting author i and magazine j, $e(i,j)$, encodes the fact that such an author has published a contribution in that magazine. Moreover, author vertices can be further distinguished according to the *gender* attribute (M, F, or NA). The *weight* attribute of an edge connecting vertices i and j, w_{ij}, can denote either the mere existence of a relationship between them (i.e., $w_{ij} = 1$) or its intensity (*i.e.*, $w_{ij} = \alpha \; with \; \alpha \geq 1 \in R^{+}$). In the former case the network is said to be *unweighted*, whereas in the latter case the network is said to be *weighted*. Here, the weight of an edge can denote, for instance, the number of contributions one author has published in a given magazine. Given a graph G with N vertices, its structure is mathematically encoded into the so-called *adjacency matrix* $\mathcal{A}$. Such a matrix is an $(N \times N)$ array whose elements a_{ij} are equal to one if an edge exists between vertices i and j, and zero otherwise. The weighted counterpart of the adjacency matrix $\mathcal{A}$ is called *weight matrix*, $\mathcal{W}$, and its elements w_{ij} are nothing but the weights of the edges. One indicator used to measure the importance of a vertex is its *degree*, which counts the number of edges incident with it. Using the information encoded in the adjacency matrix $\mathcal{A}$, the degree of a vertex i, k_i, can be written as:

$$k_i = \sum_{j=1}^{N} a_{ij}. \tag{1}$$

The weighted counterpart of the degree is called *strength*, s, and can be obtained from Eq. (1) by replacing a_{ij} with w_{ij}.

In our work, we have considered –basically– two kinds of bipartite networks: that of author-magazine relationships, and that of language-magazine ones. It is possible to build two networks for each kind of relationship: one extracted from the certified translation records G_{ct}, and another obtained from all the other records, G_{ot}. Eventually, we can also generate a network obtained by merging together the aforementioned networks, G_{merge}. Finally, magazine vertices can be

collapsed into "regions," thus enabling a macroscale level analysis of the system. Table 3 summarizes the main features of the author–magazine relationships' networks.

Table 3: Summary of the main features of the author–magazine networks. For each network, we report its number of vertices N, of edges K, the number of magazine N^{magz} and author N^{auth} vertices. We report the number of men, women, and "unavailable gender" authors N_m^{auth}, N_w^{auth}, and N_o^{auth}. We present also the average edge weight, $\langle w \rangle$, as well as the average values of the degree, $\langle k \rangle$, of magazine and author vertices, and of men and women vertices. The symbol $\langle x \rangle = \frac{1}{N}\sum_i^N x_i$ denotes the arithmetic average of quantity x.

	G_{ct}	G_{ot}	G_{merge}
K	598	5699	6146
N	485	4341	4593
N^{magz}	38	41	41
N^{auth}	447	4300	4552
N_m^{auth}	405	3516	3730
N_w^{auth}	19	159	177
N_o^{auth}	23	625	645
$\langle w \rangle$	1.59	3.46	3.36
$\langle k \rangle$	0.41	0.38	0.37
$\langle k \rangle^{magz}$	15.74	136.98	147.68
$\langle k \rangle^{auth}$	1.34	1.34	1.37
$\langle k \rangle_m^{auth}$	1.37	1.38	1.41
$\langle k \rangle_w^{auth}$	1.0	1.35	1.32

Figure 4 presents a visual representation of the G_{ct} network. As edges in bipartite networks can only connect vertices of different types, it is convenient to draw each group of vertices on one side of the figure. In our case, magazines are arranged on the figure's left side whereas authors are arranged on the right side. Moreover, we have used distinct shapes to denote vertices' types (squares for magazines and circles for authors), and colors to encode attributes like the magazine's region or the author's gender. A quick glance at the intricate web of connections between authors and magazines does not reveal any remarkable feature. However, a closer look highlights that the size of the magazine vertices is not the same for all of them. Such sizes' heterogeneity is due to the fact that the size of a magazine vertex is equal to its degree, k_i (i.e., the number of distinct authors who have published a contribution in such a magazine), and highlights the presence of magazines which

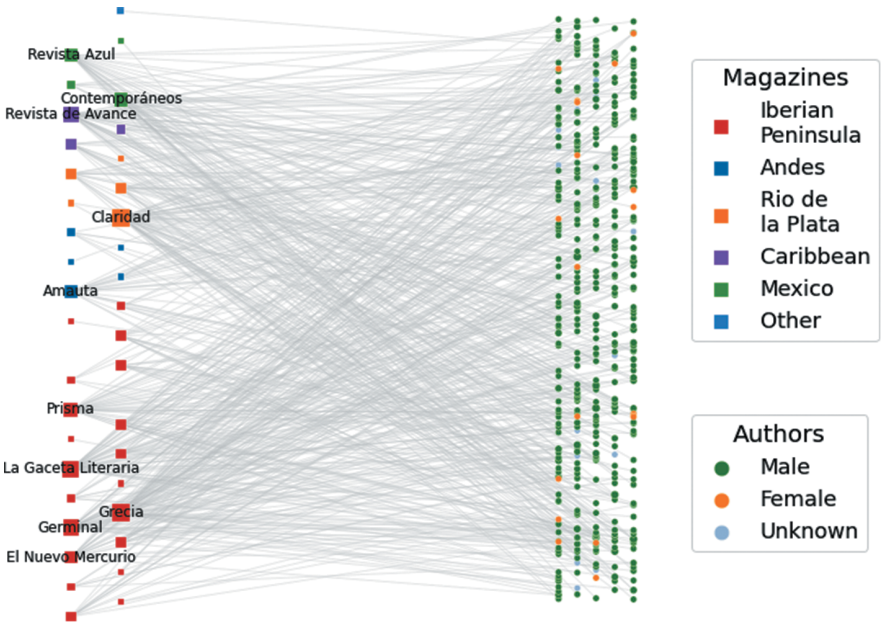


Figure 4: Visual representation of the bipartite network of magazines–author relationships obtained using only translations' data, G_{ct}. An edge between a magazine vertex (left side) and an author vertex (right side) exists if the latter has published a contribution on such a magazine. The size of magazine vertices' denotes their degree, whereas its color denotes the region to which the magazine belongs to. We display the names of those magazines whose degree is higher than the average. Authors' vertices, instead, are colored according to their gender attributes.

published the contributions of many authors (e.g., the Mexican *Contemporáneos*, the Argentinean *Claridad* and *Prisma*, and the Peruvian *Amauta*). On the other hand, the color of authors' vertices helps to grasp the huge disproportion existing between the amounts of men and women authors.

After analyzing the network of author–magazine relationships, one can ask how wide a magazine's languages portfolio is. To measure the level of "internationalization" of a magazine i, we consider two indicators. One is the degree of the magazine, k_i (see Eq. (1)), computed in the language–magazine network. The other indicator, η, accounts instead for multiple aspects simultaneously and is defined as:

$$\eta_i = \frac{1}{\Delta T_i}\left(\frac{k_i}{N_{\text{lang}}} + \frac{s_i}{W_{\text{TOT}}}\right), \tag{2}$$

where ΔT_i is the magazine's *age* (computed as the difference between the publication dates of the first and last issues available in our corpus). N_{lang} is the number of original languages available in the whole corpus, and W_{TOT} is the total weight of the network (i.e., the sum of the weights, w_{ij}, of all the edges in the network). In particular, the term $1/\Delta T_i$ in Eq. (2) accounts for the magazine's longevity by dividing the other quantities by the magazine's age ΔT_i (i.e., converts them into the average unit time counterparts). The term $k_i/N_{\text{lang}} \in [\varepsilon, 1]$, instead, accounts for the *language diversity* of contributions with $\varepsilon = 1/N_{\text{lang}}$ corresponding to the least diverse case (corresponding to translating contributions only of one original language), and 1 (i.e., $k_i = N_{\text{lang}}$) denoting the most diverse one. Finally, the term $s_i/W_{\text{TOT}} \in [\varphi, 1]$ accounts for the volume of contributions translated by a magazine, with $\varphi = 1/W_{\text{TOT}}$ denoting a magazine publishing a single translation, and 1 (i.e., $s_i = W_{\text{TOT}}$) the case of a magazine publishing all the translations in the whole corpus. Table 4 collects the values of both indicators for all the magazines computed in the magazine–language network built using only translation records. For each region, we highlight the magazine with the highest value of k and η.

Table 4: Estimating magazines' internationalization. For each magazine, we report its internationalization score computed either using Eq. (1) (column k), or Eq. (2) (column η). We highlight the row corresponding to the most international magazine according to each score. Scores indicated with "–" corresponds to those magazines which do not have any translation.

Global translation zone	Magazine Name	k	η
Iberian Peninsula	Alfar	6	0.039058
	Alma Española	2	0.01787
	Arte Joven	4	0.002832
	Carmen	2	0.023826
	El Nuevo Mercurio	4	0.02609
	Gente Vieja	5	0.006177
	Germinal	7	0.0219
	Grecia	8	0.024013
	Helios	3	0.011308
	Horizonte (Madrid)	1	0.035841
	La Gaceta Literaria	11	0.012342
	La Vida Literaria	7	0.083503
	Litoral	1	0.002306
	Luz	6	0.033159
	Prisma	7	0.072038
	Reflector	–	–
	Renacimiento	1	0.007942
	Revista Nueva (Spain)	6	0.047749

Table 4 (continued)

Global translation zone	Magazine Name	k	η
	Sur	–	–
	Ultra	2	0.011102
	Vida Americana	1	0.071479
	Vida Nueva	4	0.14306
Andes	Amauta	5	0.007634
	Boletín Titikaka	2	0.011001
	Editorial Titikaka	1	0.002978
	Instantáneas	1	0.014296
	Luz i Sombra	2	0.028653
Rio de la Plata	Claridad	7	0.009466
	La Bibliotéca	2	0.006216
	Martín Fierro	4	0.006827
	Proa	4	0.016858
	Vanguardia	0	0
Caribbean	Cuba contemporánea	3	0.001485
	La Habana Literaria	4	0.095407
	Revista de Avance	7	0.011993
Mexico	Contemporáneos	4	0.007007
	Horizonte (Jalapa)	4	0.023839
	Irradiador	–	–
	Revista Azul	0	0
	Revista Nueva (México)	1	0.071479
Other	Creación	–	–
	Favorables Paris Poema	1	0.023843

The results highlighted above can make us think about how to empirically quantify the international character of Ibero-American journals and, more generally, the circulation of translation in periodicals. If we take into account linguistic diversity (i.e., the number of translated languages and literatures), we are focusing on the degree of magazines (given by Eq. (1)), and rank them according to it (blue filled rows of Table 4). Specifically, we have, *La Gaceta Literaria* for the Iberian Peninsula, *Contemporáneos* in the case of Mexico, the Cuban *Revista de Avance* for the Caribbean, the Peruvian *Amauta* for the Andes, and the Argentinean *Claridad* for the Río de la Plata. All of them were avant-garde magazines, and tried to combine aesthetics and politics, socialism and cosmopolitanism. The translation of socialist authors, within the reference of the Russian Revolution, was also remarkable and these magazines channeled through translation the re-

lationship between the artistic avant-garde of the interwar period, and the proletariat movement. Without going too much in depth, let us only remind that the translation of socialist authors in the Mexican magazines *Contemporáneos* also contributed to the consecration of the novel of the Mexican Revolution, which also circulated internationally[6]. If we, instead, rank the entries of Table 4 using Eq. (2) (which accounts simultaneously for linguistic diversity, volume of translations, and the longevity of the magazine) the most international magazine of each region (corresponding to the rows of Table 4 highlighted in yellow) changes. Specifically, we notice the presence of a group of magazines established at the end of the nineteenth century: *La Habana Literaria* (1891–1893), the Spanish *Vida Nueva* (1898), or the Chilean *Luz i Sombra* (1900). All of them are characterized by a shorter longevity compared to the group of magazines mentioned previously, albeit they possess a similar linguistic diversity and international scope. For the twentieth century, this is also the case of the Mexican *Horizonte*, which appears to be more international than *Contemporáneos*, the latter having a four times longer lifespan, despite having a slightly larger number of translations. We observe a similar phenomenon for the case of the Argentinean *Martín Fierro* and *Proa*. Whereas the first one lasts longer and has more contributions (translated texts), the internationalization of the second one according to Eq. (2) is higher.

5 Global Translation Flows in Ibero-America within a Gender Perspective

The use of big data and network science methods can unveil under-studied cultural mediators, less-studied geographical scales, and world literary fields. The latter is a much-needed methodology to explore how particular cultural developments were undertaken by less well-known agents in less-studied geographical settings and, at the same time, avoid simplifications like that a whole region can be represented in translation by single authors (for example, Neruda or Borges in the case of Latin America). Hispanic cultural mediators contributed to foreign journals, and non-Spanish speaking critics and writers also published in Ibero-American periodicals. Likewise, Hispanic periodicals advertised other journals, allowing us to examine the networks of their alliance or rivalry relationships. Adding a gender perspective to the analysis of

6 For example, the novel *Los de Abajo*, by Mariano Azuela, was translated into English by Enrique Munguía (*The Underdogs*, 1929), into French by Joaquín Maurín (*L'Ouragan*) and into German by Hans Dietrich Diesselhoff (*Die Rotte*, 1930).

global translation flows could help us in shedding light on a shadowed canon which has suffered little recognition despite being popular at the time, and not very acknowledged today or even completely forgotten. In this respect, the focus on Ibero-American women (both writers and translators) is twice as rebellious as they have been considered as the "periphery of the periphery." Within this framework, a network science approach can verify the cultural mediator's profile and corroborate or question prevailing conceptions regarding ethnicity, class, and the contribution of women in intercultural networks. For example, it can unearth cultural mediators who were overshadowed by mainstream history and may appear in the center of the network showing a more significant role. Or, quite the contrary, a network science approach can show the peripheral position of well-known authors in cultural mediating processes. Network analysis can also restore the presence of women, contest their upper class and white ethnicity, and address how women joined forces on a transatlantic scale through their professional and personal relationships, as well as their travels and stays abroad, showing how they not only contributed to the building of the Ibero-American modernity, but also to a modern treatment of gender issues, (Roig-Sanz 2023).

In fact, the turn of the 20th century brought in the Ibero-American field important transformations in the traditional relation of women in the private sphere. The emergence of women's rights movements brought to light the collective identity of women in many fields, and women writers, and feminist periodicals gave voice to a wide range of concerns. However, most Spanish, Portuguese, and Latin American literary histories disregard women in Ibero-American modernisms, and even major figures are rarely included. Thus, we lack a clear understanding of their public and mediating role beyond national borders, calling us to examine how Ibero-American women contributed to the shift of women's roles in the modern world through their work as diplomats, journalists, editors, cultural animators, radio speakers and, ultimately, translators.

Concerning translation, we have extracted from our data a total of 266 translators. Beside their names, no other meta information is available on these translators, so we do not have data on their gender, unless a translator is also considered as an author in our database. In such a case, information about the translator's gender could be available. Although the aim of this paper is not to retrieve the gender of all the translators, it is worth noting that data on translators is often lacking or hard to find. At present time, our preliminary results show the following regarding the presence of women translated authors. Mexican literary magazines are those translating in proportion more English speaking women writers. In the Caribbean and Iberian Peninsula, women authors represent 7.69% and 8.82%. For French speaking authors, only the Iberian Peninsula, Río de la Plata, and Caribbean magazines translated a small amount (< 4%) of women. Surprisingly, as Figure 5 attests (see below), there are no translated women writers in the Andean mountains and Other regions.

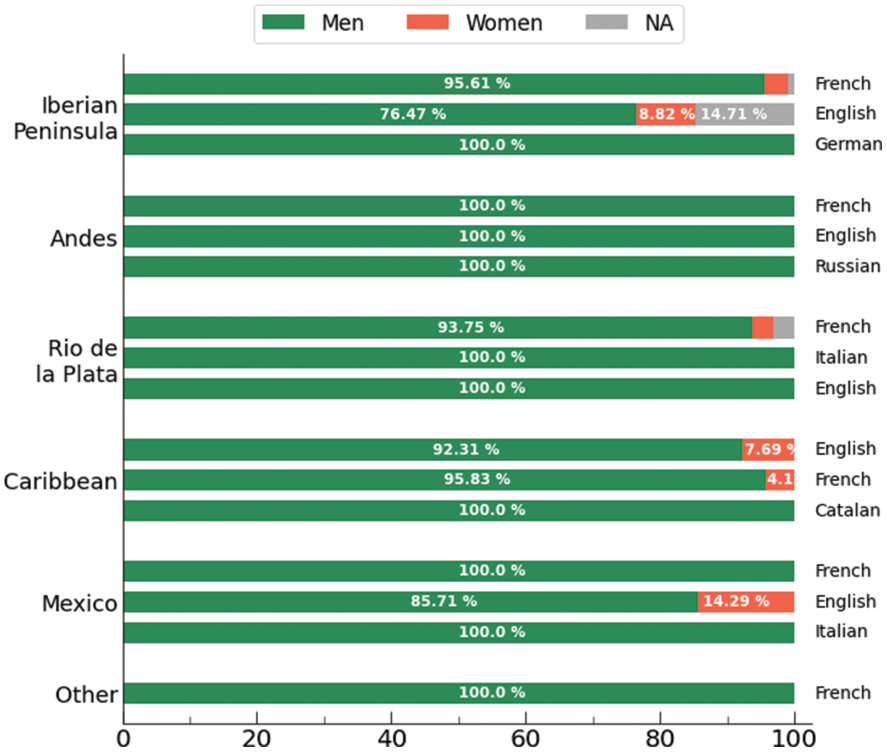


Figure 5: Percentage of authors per gender group (men, women, and "unavailable gender information" (NA)) across global translation regions. In each region, we compute the percentage of authors per group for each of the three most translated languages.

Among the 266 translators that appear in our dataset, we can only find 14 different names of women translators (13 as two of them are the same person) for the six translation zones of our choice (Caribbean, Andes, Mexico, Iberian Peninsula, Río de la Plata, and Other). The list is the following: for the Río de la Plata, the Argentinean Adelina del Carril, Luisa Díaz Sáenz-Valiente, and Gràcia B. Llorens, pseudonym of the Catalan poet, journalist and translator Maria Gràcia Bassa i Rocas. Adelina del Carril, Ricardo Güiralde's wife, was a translator for *Proa*; Díaz Sáenz-Valiente was the translator of Pierre Reverdy in the Spanish literary magazine *La Gaceta Literaria*, and Gràcia B. Llorens translates from Catalan into Spanish in the Argentinean journal *Claridad*. For the region Mexico, Antonieta Rivas Mercado, feminist writer, translator and artist, and Luz Murguía de Ramírez, who founded the journal *Violetas* with Mateana Murguía; for the the Caribbean zone the Cubans Emilia Bernal, Esther Lucila Vázquez, Mary Antiga Caballero, Mary Caballero de Ichaso, and Aurelia Castillo de González, and Carmela Eulate Sanjurjo from Puerto Rico. Emilia Ber-

nal, Mary Caballero Antiga, and Mary Caballero de Ichaso are translators in the Cuban journal *Revista de Avance*, whereas Aurelia Castillo González, in *Cuba Contemporánea* and *La Habana Literaria*. Caballero Antiga and Caballero de Ichaso were the same person, signing the latter one with the family name of his husband, the well-known Cuban intellectual Francisco Ichaso.

For the Iberian Peninsula, Zenobia de Camprubí, Juan Ramón Jimenez's wife and the first translator of Rabindranath Tagore into Spanish, María Teresa León, who translates in our corpus with her husband Rafael Alberti, Carmina Colomé, and Tatiana Enco de Valero, an exiled translator of Russian origin living in Madrid who translates from Russian into Spanish a short-story by Ievgueni Zamatin in *La Gaceta Literaria*. We have not included in this list the case of Kaethe Lewy. Even though there is no reference to the translator, she probably translated from German into Spanish some excerpts of a long interview with Jacques Maritain published in *La Gaceta Literaria* ("Catolicismo en el extranjero. Francia. Neotomismo. Conversación con Jacques Maritain"), which was originally translated from French into German by Käthe Lewy and published in the German literary magazine *Die Literarische Welt*. The name of Käthe Lewy might be one potential variation of the name Ketty Levy, Enriqueta Levy de Rodríguez, a Spanish translator of German language. If this holds true, then we could analyze an interesting case of triangular translation between French, German and Spanish. Among Spanish women translators, we must also highlight the fundamental role of Zenobia Camprubí, who was the translator of some prose by Rabindranath Tagore in the literary magazine *Grecia*. She translates from English into Spanish and it is well-known that despite signing many of these translations together with her husband Juan Ramón Jiménez she was the main translator. Unfortunately, her name does not appear in the sole dictionary of translation for Spain (*Diccionario histórico de la traducción en España*, 2009), it does the name of her husband. A great interest in the translation of poetry and in the translation of other women's work is also noticeable. For example, Carmina Colomé translates some poems by Cora Laparcerie, a French poet and actress, in the magazine *Grecia*, whereas the Cuban Aurelia Castillo de González translates poetry from Gregh Fernand and Alphonse Lamartine in *Cuba contemporánea* and *La Habana literaria*, the Mexican Luz Murguía de Ramírez translated poems by Victor Hugo in *Revista Azul*, or the Argentinean Díaz Sáenz-Valiente translates prose poetry by Revery.

Regarding the socio-biographical profile of these women translators, we can expect high society and educated women such as Zenobia Camprubí, but also feminist writers such as the Catalan Llorens, the Mexican Rivas Mercado, or the Cuban Caballero de Ichaso, who was involved in the foundation of the Lyceum Club in La Habana, and Castillo González, who was also concerned by the situation of black and mulatto Cuban women. There were also women translators who

were close to communist ideology (for example, Enco de Valero) and women translators who also had fundamental and transforming experiences of traveling (for example, Zenobia de Camprubí or Llorens) or exile (for example, Enco de Valero). Some of them were also Jews (for example, the above-mentioned potential translator Lewy). The variety of languages they translate is also broad, being French the most common. Beyond French and English, we can also highlight German and Russian. As said above, there are no women translators in the journals published in the Andean mountains and Other regions. In our dataset there are other women in the role of authors who were translators too. This is the case of the French Mathilde Pomès.

6 Conclusions

In this chapter, we have analyzed a large scale collection of scattered translations and circulation of world literature in the Ibero-American literary press using methods of network and data science and a perspective of big translation history (BTH) (Roig-Sanz and Fólica 2021). Specifically, we have selected a sample of journals published in what we have called global translation zones: the Iberian Peninsula, the Andean Mountains (Chile and Perú), the Río de la Plata (Argentina and Uruguay), the Caribbean (Cuba), Mexico, and Other (France). By focusing on these zones, we show that by examining units of analysis other than nation states, we escape from central languages (English or French) and more common disciplinary approaches, and succeed in locating cultural transfers in other spaces, such as a wider region or among minorities and small languages. Thus, we have provided some examples to compare how looking at the data at different scales can change our perspective and highlight similarities and differences between translation zones. In the case of a vast region like Ibero-America, the information available within the corpus analyzed here can be leveraged to shed light on the role played by apparently less prominent localities for translation or intercultural exchange, and not only in Madrid, Buenos Aires, Mexico City, or São Paulo. This is the case of Puno, in Peru, where *Boletín Titikaka* was published, or Barcelona, in Catalonia, Spain, for *Luz* and *Prisma*. These cities hosted many seminars, lectures, poetry readings, and a wide range of other cultural productions with the purpose of breaking with the elitist idea of culture established, for instance, in Lima; or with both cosmopolitan and nationalistic goals as in Barcelona. Thus, both Puno and Barcelona were vibrant places of intellectual discussion, cultural, political and artistic renewal and global translation flows.

Likewise, by understanding translation as an essential element of literary geography, we may also seek to investigate relevant differences between translation practices established in port cities and the mountain capitals, as well as specific challenges when analyzing cultural translation in periodical publications. In the journals of our dataset, with different periodicity and published at different time periods, we have identified their level of internationalization through the analysis of translations, and literary excerpts of world literature mainly published in Spanish, but also in other languages. By analyzing the publications' records we pinpoint those contributions associated with translations and, then, build bipartite networks of relationships between either authors and magazines or between magazines and languages. Journals and literary magazines tried to increase their prestige and relevance through the connections with international: through the publication of a wide range of translations in their pages, or via a wide network of international relationships. In particular, the analysis of the structural properties of the network of magazines–languages relationships has proved to be useful to quantify the degree of internationalization of those magazines. Indeed, the data on translations and their original languages allowed us to calculate an index to measure the international character of a literary magazine. Such an index can be used as a comparison metric, particularly when studying magazines participating in international literary networks. While at first glance the overall results might mismatch with the expectations of literary historians, a closer inspection suggests that such a measurement may be promising. Further development of Eq. (2) could include a more refined calculation accounting for other variables like the number of issues or contributions over time, or the ratio of translations vs. original works, and also be fine-tuned by including factors to weight the contribution of each term.

In future endeavors, we would like to push forward other potential lines of interest. These new research avenues include i) to amplify and balance the dataset by adding more magazines from Ibero-America (including the Lusophone area) and from different time periods within the general time frame of our interest (1898–1959), ii) to provide comparisons on a lower level, e.g., between magazines or between cities. The current approach assumes translation zones as homogenic, while a detailed analysis of the magazines could confirm or disprove this idea. And iii) to consider the chronological factor, which means, on the one hand, to compare magazines, cities, regions horizontally in discrete time periods, and consider the specificities of each period, and, on the other, to compare the evolution within a magazine, city, region longitudinally, i.e., across different discrete periods of time. Finally, we would like to delve more into the gender perspective, and offer new knowledge on the socio-biographical profile of women translators and authors. In this regard, we envision two potential paths: Named

Entity Recognition (NER) using machine learning algorithms or a combination of techniques including NER and complementary cross-validation of the information via platforms like VIAF, or customized analysis of the contribution's type. We can also identify Ibero-American women translators in projects such as the WikiProject Women in Red.

Bibliography

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

Ashrafi, Nasrin, Mohammad Reza Hashemi & Hossein Akbari (2019): "Towards a new methodological approach to social historiography of translation. A social network analysis perspective", in *Translation Spaces*, 8, pp. 231–56. doi: 10.1086/599247.

Barabási, Albert-László (2011): "The network takeover", *Nature Physics*, 8, pp. 14–16. doi:10.1038/nphys2188.

Bender, Thomas (2006): *A Nation Among Nations. America's Place in World History*. New York: Hill and Wang.

Berg, Maxime (2013): *Writing the History of the Global. Challenges for the 21st Century*. Oxford: Oxford University Press.

Borgman, Christine (2015): *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge, MA: MIT Press.

Carbó-Catalan, Elisabet & Diana Roig-Sanz (2021): "Las instituciones culturales y la cooperación. La creación de redes transnacionales en la primera modernidad española", in Jordi Gracia & Domingo Ródenas (eds.), *Las dos modernidades. Edad de Plata y Transición Cultural en España*. Madrid: Visor, pp. 193–229.

Carbó-Catalan, Elisabet & Reine Meylaerts (2022): "Translation Policies in the Longue Durée: From the International Institute of Intellectual Cooperation to UNESCO". In Diana Roig-Sanz & Neus Rotger (eds.), *Global Literary Studies: Key Concepts*, Berlin: De Gruyter.

Castells, Manuel (1996): *La era de la información. Economía, sociedad y cultura*, vol. 1. México: Siglo XXI.

Conrad, Sebastian (2016): *What Is Global History?* Princeton: Princeton University Press.

Cussel, Mattea (2021): "Methodological nationalism in translation studies. A critique", in *Translation and Interpreting Studies*, 16, pp. 1–18.

D'Hulst, Lieven (2012): "(Re)locating translation history: From assumed translation to assumed transfer", in *Translation Studies*, 5, pp. 139–55.

Ehrlicher, Hanno (2020): "Revistas culturales históricas en lengua española desde el modernismo hasta las vanguardias: procesos de modernización y formación de redes transnacionales – Spanish-language Cultural Magazines from Modernismo to Avant-Garde: Processes of Modernization and Transnational Network Formation", in DARIAH-DE. doi:10.20375/0000-000d-1cff-6.

Fólica, Laura (2021): "Digital Humanities and Big Translation History in the Global South: the Spanish-Speaking Case", in *World Literature Studies*, 3, pp. 104–116.

Fólica, Laura (2022): "Cosmopolitanism Against the Grain: Literary Translation as a Disrupting Practice in Latin-American Periodicals (*Revista Nosotros*, 1907–1945)", in Diana Roig-Sanz and Neus Rotger (eds.), *Global Literary Studies: Key Concepts*, Berlin: De Gruyter.

Fólica, Laura, Ventsislav Ikoff & Diana Roig-Sanz (2018): "Translations and Reviews in Iberoamerican Modernist Periodicals (dataset)". doi: 10.23728/b2share.eb5c468d3dc3401c8b2fb4605d868a00. (12/2018).

Fólica, Laura, Diana Roig-Sanz & Stefania Caristia (eds.) (2020): *Literary Translation in Periodicals: Methodological Challenges for a Transnational Approach*. Amsterdam/Philadelphia: John Benjamins.

Gramuglio, María Teresa (2013): *Nacionalismo y cosmopolitismo en la literatura argentina*. Rosario: E(m)r.

Herzgsell, Teresa (2020): Revistas culturales históricas en lengua española desde el modernismo hasta las vanguardias: procesos de modernización y formación de redes transnacionales. Spanish-language Cultural Magazines from Modernismo to Avant-Garde: Processes of Modernization and Transnational Network Formation. Accompanying Publications. 04_Teresa-Herzgsell_Categories.pdf. DARIAH-DE. https://doi.org/10.20375/0000-000d-1d06-d

Latora, Vito, Vincenzo Nicosia & Giovanni Russo (2017): *Complex Networks*. Cambridge. UK: Cambridge University Press.

Lehmann, Jörg & Hanno Ehrlicher (2022): "Transnational Network Formation in the Medium of Cultural Magazines". *Journal of Historical Network Research*, n.º6 (1). https://doi.org/10.25517/jhnr.v6i1.90

Liu, Alan (2018): *Friending the Past. The Sense of History in the Digital Age*. London/Chicago: The University of Chicago Press.

Manovich, Lev (2015): "The Science of Culture? Social Computing." *Digital Humanities and Cultural Analytics*. doi 10.22148/16.004.

Manovich, Lev (2020): *Cultural Analytics*. Cambridge, MA: MIT Press.

Meyer, Eric & Schroeder, Ralph (2015): *Knowledge Machines. Digital Transformations of the Sciences and Humanities*. London/ Cambridge (US): MIT Press.

Middell, Matthias & Katya Naumann (2010): "Global history and the spatial turn: from the impact of area studies to the study of critical junctures of globalization", in *Journal of Global History*, 5, pp. 149–170.

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.

Pegenaute, Luis (2018): "Translation and cultural development: Historical approaches", in Sue Ann Harding & Ovidi Carbonell i Cortés (eds.), *The Routledge Handbook of Translation and Culture*. London/New York: Routledge, pp. 177–206.

Roig-Sanz, Diana (2023): "Mapear el género. La circulación de mujeres escritoras, traductoras e intelectuales en la prensa iberoamericana (1927-1959)", Revista Chilena de Literatura, special issue "Vivir entre lenguas": prácticas de traducción en la literatura latinoamericana y caribeña", edited by Laura Fólica and Thomas Rothe, forthcoming.

Roig-Sanz, Diana & Ana Kvirikashvili (forth.): "Global Translation Zones/ Spaces of Translation: New Paradigms for Decentering Literary History", in Zhang Longxi & Omid Azadibougar (eds.). The Routledge Companion to Global Comparative Literature. London/New York: Routledge.

Roig-Sanz, Diana & Laura Fólica (2021): "Big Translation History. Data Science Applied to Translated Literature in the Spanish-speaking World, 1898–1945", in *Translation Spaces*, volume 10, issue 2, pp. 1–29. https://doi.org/10.1075/ts.21012.roi

Roig-Sanz, Diana, Laura Fólica & Ventsislav Ikoff (2021): "Translation in Literary Magazines", in Esperanza Bielsa (ed.), *The Routledge Handbook of Translation and Media*. London/New York: Routledge, pp. 199–216.

Roig-Sanz, Diana & Reine Meylaerts (eds) (2018): *Literary Translation and Cultural Mediators in Peripheral Cultures. Customs officers or smugglers?* London: Palgrave MacMillan.

Rosenberg, Emily S. (ed.) (2012): *A History of the World. A World Connecting, 1870–1945*. Cambridge: Harvard University Press.

Rotger, Neus, Diana Roig-Sanz & Marta Puxan Oliva (eds.) (2019): "Historicizing the Global: An Interdisciplinary Perspective", special issue in *Journal of Global History*, 14. Cambridge: Cambridge University Press.

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

Thiesse, Anne-Marie (2001): *La Création des identités nationales: Europe XVIIIe-XIXe siècle.* Paris: Points.

Wilfert, Blaise (2020): "De la traduction comme *publication* et comme *glocalisation*", in Laura Fólica, Diana Roig-Sanz & Stefania Caristia (eds.), *Literary Translation in Periodicals. Methodological Challenges for a Transnational Approach*. Amsterdam/Philadelphia: John Benjamins, pp. 21–46.

Wilke, Claus O. (2019): *Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures*. Sebastopol, CA: O'Reilly Media, Inc.

Willson, Patricia (2004): *La constelación del* Sur. *Traductores y traducciones en la literatura argentina del siglo XX*. Buenos Aires: Siglo XXI (reed. 2017).

Pedro Ruiz Pérez

Little Big Data: The Poem Against The Database

Digitalization entails processes of uniformization that are not always desirable. In the Aristotelian model of science (to know by universals), the neutralization of specific differences is necessary and is the basis for intellection, and this guides the formulation of laws of general application. In the field of aesthetics, and even in that of history, being able to generalize is inescapable and valuable. This is so in the field of theory, in *longue durée* periodization or in the defining of genres (literary, artistic . . .). However, ultimately the value of singularity endures: the singularity of the artistic object and the receiver, be they student, researcher, or one who does so for pleasure. The notion and development of the "digital humanities" offers up an initial solution to this duality – or at least a space for dialogue of the double perspective. The potential and speed of growth of the technological component of the polarity multiplies the modalities of application and at the same time colonizes the supposedly shared terrain. Without going too deeply into the difference between information and knowledge, the play of tensions runs the risk of reduction in the face of a situation of hegemony when numerical logic,[1] and its extreme dimension, the realm of Big Data, imposes itself. What follows is a reflection undertaken from a position of one working in research connected to the Digital Humanities, and with the distrust of one who is wary of its risks and limitations.[2]

1 Obviously, I am not referring to the meaning that prevails in the use of *numerique* in French, for which the term *digital* is used in Spanish and English. Neither am I evoking the idea that the superseding of the analogical is produced through the 0/1 binary code, numbers that are also coming to replace letters, beginning with those of the alphabets.

2 The reflections that follow have arisen while working on the ongoing project, *Biografías de autor e institución literaria en la edad moderna* [*Author and Literary-Institution Biographies in the Modern Age*] (SILEM) RTI2018-095664-B-C21 of the National R&D&I Programme (<http://www.uco.es/servicios/ucopress/silem>). It is a continuation of *Sujeto e institución literaria en la edad moderna* [*Subject and Literary Institution in the Modern Age*], a National R&D&I Programme coordinated project, FFI2014-54367-C2-1-R. Prior to this, there were two projects (PHEBO) on late baroque poetry (<http://www.uco.es/phebo/es>). In this period I also coordinated the setting up of the Aracne Network (<https://www.red-aracne.es/presentacion>), along with the complementary *Humanidades digitales y letras hispánicas* [*Digital Humanities and Spanish Literature*], National R&D&I Programme FFI2011-15606-E.

https://doi.org/10.1515/9783110753523-012

Forms of Reading (And of Research)

If the steps I follow and the end I seek had to be reduced to a formula, it might be found around a dialectic between the quantitative and the qualitative, one in which balance was not imposed by an overwhelming power and which did not end up being diluted by the traditional relationship between means and ends. In other words, one in which mere technology does not end up dominating science, again with reference to Aristotle. The aim is to preserve and cast the essential factor of the aesthetic experience in a renewed space, while finding, through experience, a balance between certain bases in tradition and the expectations of new technologies and epistemologies, specified in research possibilities. I begin with a re-affirmation of the subject.

Yo nací (perdonadme) / en la edad de la pérgola y el tenis" ("I was born (forgive me) / in the age of the pergola and tennis"), as go the well-known lines by Gil de Biedma. I had no *château* with grass courts; I only belong to a time, or, at least, come from one. More accurate, both in general and for this case, would be Rafael Alberti's words, which influenced de Biedma: "Yo nací – ¡respetadme! – con el cine" ("I was born – respect me! – with cinema"). And I say this while witnessing the growing hegemony of series on television platforms – that is, I am aware that the basis of my scientific knowledge has shifted to another dimension, something that has happened with audiovisual culture ever since the Lumière brothers.

In a space that is so propitious to uniformity, it is worth stressing the statements that give character to a position, which is what I now maintain. In essence, this is the consideration of the nature and specificity of the literary text and of philology, as well as an active assumption of the need for qualitative advances in research projects and professional networks, with the role they play in this framework of Digital Humanities. My starting point is the experience of this journey with intellectual conviction, not blind faith. From this conviction, in which there should always be a seed of doubt, I contemplate the changes that are taking place.

We should not lose sight of the lines of continuity in the paradigms. I wrote my thesis on a typewriter and using index cards. It was common practice, as it was to incorporate statistics of verses, including rhythmic variants. We were not, therefore, adverse to working with data, and that included an aspiration to globality (aimed for by the habitual title of "Life and Work of . . ."), for an all-encompassing view (which, by the way, the new thesis model has fragmented), and for the consequent development of management protocols.

Although now is not the time to historicize or fall into the habitual narcissisms of the academic field, it may be a good time for a backward glance to have a clear awareness of the place I am speaking from. Since 2005, the research pro-

jects I have been involved in with some responsibility have tackled objects in which the quantitative factor was foremost, and with a not insignificant scope in the terms of our field.[3] In fact, a methodological premise of mapping the field was a constant in all of them, resulting in the producing of a more or less extensive corpus or repertory, although in the usual terms in the research of our field, it could be considered relatively *big*. This was the case in the study of fifteenth- to seventeenth-century poetic texts with lists of authors, in the analysis of the poetry that we ended up calling the Late Baroque, and, finally, in the research on discourses in which the concept and image of the author and the literary institution was forged – now in a longer period that stretched to the mid-nineteenth century. We thus addressed the work that could be considered a poetic subgenre, in a defined chronology and in a modality of discourse, with aspirations of attaining exhaustiveness. The instrumental aims were the drawing up of repertories or catalogues, databases and digital libraries. The differences denote a process that was both technological and conceptual: in the first project we worked with paper index cards and Word files, resulting in the publication of a book (Ruiz Pérez 2010), in which the traditional indices tried to maintain a certain flexibility of use. The second project included the design and creation of a database with an advanced search system and a library of static texts, all available on a website. Lastly, in recent years, digital libraries with treatments of texts and the capacity for conceptual searches have shown their potential for examining the compositional mechanisms of a discourse.

This experience can explain my perspective. Its timing, coincidental with what can be considered the widespread development of the Digital Humanities, gives it a value of some significance. Computerization has exponentially increased both pace and possibility, even turning them into a qualitative change, similar to that brought about by the printing press when it overcame the incunabulum phase (reproduction of manuscripts) and configured another model of volume; or, as Guglielmo Cavallo (1975) and Armando Petrucci (1979) have shown, in the previous shift from the scroll to the codex: changes in formats, changes in models of reading. We are immersed in an equivalent paradigm shift.[4] We need to seek suitable adaptation, and this involves a review of the problems and limits of re-

3 Please see the references and information given on the websites mentioned in the previous footnote.

4 On the speed of change, see the advances in information and reflection given in the monograph coordinated by Morrás and Rojas Castro (2015) and in the volume edited by González and Bermúdez Sabel (2019). Since the start of the century there has been a proliferation of associations, conferences and journals on Digital Humanities, which record the extent and progress of this discipline. References and links can be found on the Red Aracne website, cited above.

search, a reconsideration of the object (susceptible to being tackled in broader terms, also quantitatively), and an appropriate design of the tools, hierarchically subordinated to the previous principles.

In the shifts and turns undergone in the Digital Humanities, the initial stage multiplied the possibilities of the individual reading, making a greater volume of information available in the form of texts. Globalizing models in the consolidation of the practice and the refinement of its tools will soon be established, if they have not already. The development of the technique involves multiplication and acceleration; its results, in exponential growth, first affect the object and end up imposing uniformization, which affects the perception of the subject. The process, logically, is in perfect harmony with the discourse of globalization. The problem makes itself felt in its application to poetry. I use the strict meaning, which has been central to my work, but the term can be applied to all textual reality that cannot be reduced to strict categorization. Even without positioning myself at the extreme end of the paradigm of the singularity of the literary text, its resistance to the processes of neutralization is clear here – or, at least, the necessity for nuances and particularizations.

The methods of quantitative analysis, and their application with tools that exponentially increase the numerical dimension, open up exciting paths for a renewed study of literature. However, many of these paths are unexplored; down them lies the risk that the hope or impression of oases is but a mirage. In order for the numbers of digitalization not to be imposed upon the discourse of words in the humanities, we need to reflect, based on the premise that all technology implies ideology, and prepare for incursions into wild territories without due calculation of risk. In this perspective, the aspects considered in this reflection are those relative to the specificity of literature and its discourses, the role of digital tools and the dimension of big data in research in this field, the redefinition of the object of study, and the creation of corpora to scale. I offer this as a trial run of the possibilities of dialogue between the critical and philological tradition and the epistemology derived from a technology that is still expanding.

The dilemma is this: will we master the technology and put it to the service of certain criteria, or will we sacrifice the criteria to the omnipotence of the god of technology? In less Manichean terms, up to what point should we refresh our criteria with new perspectives?

Some Aspects for Reflection

Let's take the following postulate: the relationship between data and base, the published edition of a text and the digital library, the index card and big data, reproduces a comparable text-macrotext relationship, conceptually and functionally, which operates with particular intensity in a collection of poems. As a part of the whole the poem does not lose its autonomy, but it is enriched in the dynamic of the book. And it is updated in every reading. Can we establish and maintain such a relationship in the design of our digital tools?

To walk this road, we need to establish a perspective, in its double meaning of a depth of field and a line of sight. We can do so with a question that requires an answer: what differences are there between humanist concordances, of the Bible or the oeuvre of Virgil, and the possibilities of a database, aside from making it possible to endlessly expand the corpus? The advantages produced by technology, where we must focus, enable us to surpass the possibilities of the alphabetical search to find the use of a word or its recurrence in the corpus of a work or author. Its repercussions in the design and execution of a research project in its widest sense are found in several shifts: 1) from the static to the dynamic; 2) from the linear to the relational; and 3) from the strictly lexical to the conceptual.

1) From the static to the dynamic. This is the relationship that takes the economy of a book of poems as parallel: that is, how its components work with the established *dispositio* –talking in the terms of classical poetics – or, in the semantic field of computers, with its architecture, with the functionality that entails. One lesson with regard to overcoming the static nature of text is that of the linguistics of the text, with its concepts of coherence and cohesion, its notion of discourse and the turn to the syntagmatic. Its model is an invitation to abandon the consideration of isolated and static elements, typical of an outlook with its basis in the paradigm. The model requires a syntagmatic consideration, in which the elements are conceived relationally, by their functions and their capacity to establish nexuses and dialogues with the rest of the elements *in praesentia* – to continue with established terminology in Philology and structuralist linguistics. Upon introducing data, we imitate the constitution of a paradigm, that is, a more or less systematic set in which elements that maintain some kind of relationship between them are arranged linearly. This is not the only line possible but one established with a specific objective. In principle, these data can enter the series with the independence appropriate to their real situation. Looking at the framework we are interested in, a piece of writing is materialized in a book (on paper or digitalized) with autonomous existence, whether on the shelf of a bookshop or of a library, no matter their organizing principle (by author, genre, chronology . . . even by size or colour). Only in the activity of the reader who

feels moved by the message of the text they read is there the possibility of jumping from one book to another with productivity. The conception and mechanics of a database or a digital library make it possible to surpass that reading dynamic, enhancing and intensifying it, multiplying its potentiality. The quantitative difference (in the number of texts and possibilities of interconnection) can be turned into a qualitative turn. When conceiving it grouped together, the datum/text ceases to be an autonomous element and a value that literary studies emphasized some time ago comes into play: all literary work is a weaving of intertextual relationships, where we can find the beginnings of a principle of dynamic relationships that transcend the strict frontiers of the text in its strict immanence. Its semantic and aesthetic richness comes alive to differing degrees, depending on the culture and capacity of each reader. In a computerized repository, at least as far as research interests are concerned, such a circumstance should be maintained, while making the most of the potentiality provided by technology to increase this dynamic of relationships and materialize them on a screen.

A paradox emerges, however. In literary philology, for centuries we have been working not with a language (limitless productive force, *dynamis*, *energeia*), but with texts (paradigm of perfect, complete *ergon*) – Text with a capital T, moreover, emblematizing a clear paradigm: the reduction in criticism of all the variants generated by a process of dynamic and complex transmission.[5] The supposed reconstruction of the archetype or ideal text is, for a non-specialist reader, situated on a line of exclusion opposite to the virtuality of hypertext. Yet the researcher or reader moved by philological curiosity can make the most of the critical apparatus to rebuild the trajectory of the text's different avatars and contemplate it in its dynamism, whether to consider the process as a whole or whether to pause in consideration of one of the episodes of their communicative course, of their different historical realizations. As with the concordances, in the philological tradition we discover – with all the nuances brought about by technological changes[6] – a princi-

5 I am thinking here about the essential aim, established according to principles related to faith in the ideal text and, therefore, the consideration of errors for all its variants, so that critical work consists in reverting disfigurations to return to the point of correction, the original (whether existent or not). This is what should be offered to the reader. The utility that the traces of previous work leave for the specialist, and that can be recorded in the same volume where the archetypal text is published, with the critical apparatus, is something else.

6 It is worth recalling that the development of concordances, such as indices and other paratextual elements of reading orientation, was only possible with the shift from scroll to codex – that is, from linear arrangement to a system of folding that enabled page numbering and therefore the ease of finding an element in them, which before was only relatively possible in highly articulated works, such as in books, chapters and verses, as can be seen in the Bible. The Christian holy book, with its fideist conception, also has an influence on the methodological configuration

ple of reading (and even of textual arrangement) that very much takes into account the dynamics of the text, its influences and internal recurrences and its material transformations. Perhaps we can continue to learn from what has been opted for in a centuries-old philological task in which the setting of the ideal text is presented along with the testimonies of a tale of transmission that keeps a dimension of dynamic hypertext alive.

The digitalization of the literary humanities can be adapted to a mode of work and the principles that it is based on, as long as this is done by tipping the scales toward the opposite end of idealizing staticism. Into the black box of computer tools,[7] we introduce finished, closed texts; their very number can confine us to a merely quantitative consideration (such as counting an author's Sapphic hendecasyllables, as per the aforementioned model of theses of the past century). The challenge is to convert this new rhizomatic syntagm into an energy in motion, in which to maintain or generate qualitative reflection, preserving and planning a similar functioning to that of poems –as mentioned above – in the framework of a well-constructed book. In its volume, above all when this reaches extraordinary – or at least uncommon – dimensions, the data threaten to become an answer, the answer, with totalitarian temptations. How can we avoid this? Let us now turn our gaze to the experimental sciences. Set to collide, texts, like atoms, generate energy, but they also open vacuums, and in them we must find the questions – the questions that truly advance research. In physics, technological developments have enabled the corroboration of previous scientific theories that were born out of analysis, reflection and judgement. The Geneva Large Hadron Collider finally proved Higgs right in his theory of the boson, which pre-existed its experimental proof. The case seems a model for the situation I am setting forth, and we cannot stay on the sidelines of this situation and its implications in terms of the model of relationship between science and technique. When it is not a theory that precedes the research or data accumulation, the latter in their quantitative dimension should become the source of new research questions. In short, the amount of data is not sufficient by

---

of a certain model of ecdotic work, that of Lachmann and his followers, such as how the assessment of the *auctoritas* of Virgil or Horace shaped the practices of Alexandrine philology, until both models and perspectives converged in the Renaissance *studia humanitatis*.

7 I use the notion of a "black box" taken from the scientific paradigm to highlight the relative independence of philological research approaches regarding technology, in the sense of keeping the relationship between *episteme* and *techne*, between research and its tools. For the humanities researcher, it is not essential to be an expert in computing, however much an effective familiarization can produce better results. The dialogue at the heart of a multi-disciplinary team brings about the most propitious situation, since it prevents the autonomy or primacy of philological principles from becoming impermeability. As I have been stressing, technology invites the rethinking not only of our procedures but also our concepts and objectives.

itself; what we are need are modes of treatment to convert them into a dynamic element of investigation.

2) From the linear to the relational. The digitalization of a library catalogue, for example, or the creation of an archive, are not in themselves digital humanities, because they stay on the horizon of the paradigm and of linearity, in a very similar way to the text of traditional writing. The management computing tool facilitates searching in cases such as those mentioned, but does not generate knowledge by itself. The information is quicker, but it is wholly outside the questions of research – at least, if the data is kept on the level of a schematic file, such as in an OPAC. This is due to the lack or the weakness of relational models and tools. The possibility of making connections is a requirement, being essential to give a specific humanist meaning to digitalization, beyond the generic usefulness that streamlining produces in the handling of existing information. The objective of the application processes of computing to research in the humanities is the (well-oriented) creation of new discourses and knowledge. The key is the formulation of the questions. Thus it is possible to go from mere information (that of more or less numerous data) to knowledge (based on judgement/reason).

Computer memory, with its differences – above all in the order of quantity – reproduces that of the human mind, which constructs it to its image and semblance. In both cases, we can distinguish two planes. First, there is the plane of a merely receptive arrangement, in which memory is converted into a deposit for the storage of information, articulated in data with differing degrees of refinement or extent. Like material warehouses or the buried remains of an old civilization, they lack value as long as they are not put to use. This is the function of the other component of memory, as capacity to activate, to put to work and connect the available elements. This dimension is vital not only for the advance of scientific knowledge but also for the very survival of the individual or the justification of a tool.

Let us return to the mechanism of reading, from the most conventional to, in a little more than one and a half millennia, the reading fomented by hypertext. In the former, the updating of the text when received, by means of the memory (increasing with the degree of culture and capability of the receiver), calls upon other texts read, stored in the memory, but updated with a reading that, through intertextuality, activates the weave of relations. The reader has the materiality of a volume of paper and lines of ink in their hand, while in their memory they hold the vestiges of previous readings from similar objects. Intelligent reading adds to comprehension, and delight in the text revitalizes those vestiges that are only seemingly absent. The essence of this mechanism should be preserved in computerized procedures. At the same time, it should be enriched with its possibilities, since its technology makes almost instantaneous retrieval possible and from a considerably larger volume of

data, but, above all – and here we must progress – it makes retrieval possible in simultaneity. In the framework of research interests, this seems to me to be a more productive condition of hypertext than that of substituting linear reading for the attractions of browsing, and much more evocative and fruitful in the processes of active reception of artistic creation in the virtual space.

In the scientific sphere – with the humanities included within it – the outlook of objectivity means prioritizing denotation over connotation. Thus the content of big data resources does not place its value so much on the plane of the possibilities of browsing as on that of the handling of data with a certain simultaneity, that which gives the relational weave, of the available data. Unlike a reader who reads for pleasure, open to the suggestions of the (hyper-)text, whoever accesses the tool with aspirations for scientific knowledge tends to do so guided by prior questions, research questions that have emerged from earlier readings and the resulting curiosities. However, the condition of the Digital Humanities makes it possible to offer a repertoire of dimensions that was unmanageable until fairly recently, and this reality could become, in turn, fertile soil for new questions – research questions that would be tricky to consider and resolve without the existence of this tool.

In order for the answers to questions to emerge, there is the inescapable requirement of designing a conceptual and digital treatment to enrich the data and give them meaning. An image of this challenge is provided by the distance between how social networks typically function and the creation of a community framework with a bundle of relations. It is not enough to provide possibilities to generate a positive discourse. One route in our field is to incorporate not merely denotative information in the materials that are included in the database and go on to form part of big data. I don't believe it necessary to return to the example of a computerized library register in which only the mere data of bibliographical identification are recorded, and this case is transferrable to the formation of databases of critical bibliography without distinction between their data and their metadata. I propose broadening the dimension of texts in an opposite sense to that of "augmented reality": to deepen rather than adorn.

3) From the strictly lexical to the conceptual. The transcription or edition of a text for its addition to a repository only mechanizes and facilitates access, but it does not alter our knowledge, beyond the loss that is entailed when passing from the material to the virtual.[8] A search tool using words or sign sequences makes some

8 Material bibliography has shown us how to discover the semiotic potential of all components of a book, particularly one coming from manual printing, where it is meaningful from its format, lettering and paper type to the paratext or constitution of its folds. Much of this information is lost in digitalization and most of the transcriptions available online, and it is not always recov-

processes possible, but does not open the way to readings and new questions. For a new stage of knowledge in the application of digital tools to literary studies, we need to advance in the hypertextual, with added levels of information, and to design suitable search tools for a relational, not linear, scenario. In more philological – rather than computer – terms, this is about creating new reading tools that positively meet the possibilities of the technology. One possible reference here is the work of Franco Moretti, with his notion of *distant reading* (2013) and its application in the activity of the Stanford Literary Lab: tools for new readings.

Some of the proposals regarding artificial intelligence are oriented in this direction. However, they do not avoid but rather increase what for us is a risk of partitioning, bias and orientation in the image offered. The algorithms of the large search engines serve as reminders in this respect. When we use them, beyond the craftmanship of the criterion and the appropriate technique for their application, what degree of control do we have over how the software functions? Moreover, can we clearly discern whether we are in control or are being controlled? Volume of information and speed of access to it are in principle positive values; but they are not always synonymous with knowledge – and I'll let the evidence speak for itself. Indeed, they frequently replace it. From the drive that moves knowledge, the accelerators provided by technology only make sense insofar as they enable us to construct lines of access to reality (in our case of texts) with a greater chance of comprehension, free of distortions. Quantitative and technological intervention in the entity of the object of knowledge entails, at the speed we are moving, a metamorphosis of the subject's structures of thought – or, more radically, a mutation of the thinking subject.

In a process of decontextualization and neutralization brought about by its integration into the mass of data, a text – particularly a literary text – suffers a reduction in communicative efficacy. Like splashes of ink on paper, the binary code of computing cannot replace the active mechanism of reading, but neither should it limit it. In the workshops of the heirs of Gutenberg, in a process of coding reached by consensus, they fixed rules, elements and patterns of composition that regulated and conditioned what at the beginning of the modern age was considered mass diffusion and reception. A similar process took place in the computing laboratories and in the management offices of corporations that direct their work. The result is that secular textual models can become denatured when rendered into new formats and, above all, when subjected to consumer processes

ered when editions aimed at this circuit are made. Along with accessibility, other possibilities of virtualization should be considered in the treatment of texts that would make it possible to compensate for this loss.

that are far more massive than those seen back in 1500. And the phenomenon multiplies exponentially when it is the texts themselves that are diluted in the dimensions of big data. The linearity of plain text of the printed page is converted into a relic and is denatured with digitalization and insertion into a macrocorpus. Either it runs a serious risk of doing so if the particularities of the literary text are not taken into account, or its reception submits to a capitalist logic of the number. Consideration as a datum places it on the plane of pure information, which it does not belong to, in the strict sense, or which at least does not prove specific to it. Its profitability lies in the possibilities of massification. Not only must they be reduced to data; there must be many, vast numbers, like the profits on an income statement. Thus the *data* have to be *big* to have productivity. As a counterweight to this logic, what is moved by other interests is what must act. The digital humanities must not lose the first of their components, the substantive, but reserve the second for its auxiliary, instrumental function.

The condition of the datum, the minimum element of information, enables its insertion in a quantitative paradigm, even on a mass scale, where it functions like the lexical units in dictionaries – those graveyards of words, according to Julio Cortázar. Only when words become concepts is judgement possible, and we again turn to Aristotle. Thus, it is the conceptual treatment, the dimension of the humanities, that can prevent the great catalogues from becoming the macronecropolis of postmodernity and the new versions of a liquid capitalism with its uses and values. Philological editing has carved out a well-defined space to give this value to the printed word. The footnote made it possible to incorporate a deeper level of reading, or gave it a dimension that was not entirely explicit in the text. The explanation of a term, a reference or a rhetorical schema maintained in the (para-)textual *dispositio* the etymological meaning of *explicare*, 'unfold', since it opened a widened – one might say hypertextual – horizon in the reading. When the footnote established the appropriate connection with the text, it ceased to be a simple datum and created a space of knowledge. If the printed page allows this duplicity of planes, expandable with subheads and double or triple apparatus of notes, the flexibility of the volume gives the chance to add the informative or analytical appendices as required. The screen, in turn, exponentially multiplies these possibilities and takes the consideration and treatment of the text into an essentially different dimension.

The notion of data and the illusion created by the implicit and latent consciousness of their availability in a database generates unconscious mechanisms similar to those of the market and consumption. In the push to purchase, the supply from large data warehouses – now the great online retail platforms – creates the generally false impression of availability, which undoubtedly facilitates where the spending is directed, shoving it closer to wastefulness translated into profits. In the

marketing sphere, the effect is no less irresistible even though it is known. In the field of social engineering and political manipulation, it constitutes one of the biggest threats to the present and the immediate future, that of the next elections. In the sphere of the humanities, these can be diluted by its new travelling companion, digitalization.

Should our discipline remain in the state it was found in when Lorenzo Valla revealed the forgery of the Donation of Constantine with which the Church sustained its earthly power? It is obvious that it should not, as obvious as the need to think about what can be changed and what is worth addressing, without losing sight of certain values and the objectives they entail.

Towards a Liberation from the Fetishism of Data

Not entirely unintentionally, I continue with an image that is as schematic in its formulation as it is rough in its visual resolution (Figure 1):

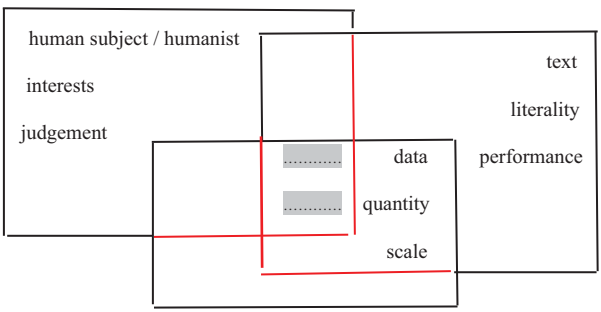


Figure 1: Reading and research processes. By the author.

Playing with graphic reduction serves to show the confluence of three elements of a disparate nature and the establishment of a space of intersection in the processes of reading and research. In a triangular relationship, we find the focal point represented by the interests of a subject, who cannot relinquish judgement; some objects of specific nature, based on an irreducible specificity that is revitalized in every reading; and, lastly, some procedures (conceptual and technological) that operate with intrinsic values unconnected to those of the other two elements – a philologist searching for a text in a digital library or a reference in a database. But let us return to the schematization.

Venn diagrams, used in my time at school to develop set theory, and the parallel of structuralist formalization, continue to serve, constrained by the limitations of Word and my computing competence. Thus its representation emblematically illustrates the dialectical triangle that I base my analysis upon and which shows the following: the opportunities and threats of digitalization in terms of big data, the specificity of texts such as literary texts, the unavoidable guiding presence of a critical subject and, overall, the need to map and colonize the intersecting space between these vectors. The limitations of the representation illustrate the spaces of perplexity and disorientation that we face in the endeavour. And in this we must distinguish between the constraints of technology and our limitations with that technology.

I shall put to one side – but not negate – reflection on the dependence of hardware, software and memory, or the potential depredation of their human correlates – operational until a couple of decades ago. Sticking to what I have outlined, and to the occasion, I shall confine myself to questions on the strategies for the positive resolution of the dialectic I have set forth. The orientation of these questions points, in my opinion, to the introduction of dynamisms in catalogues, which cannot be inert. Thus the potentiality of knowledge derived from numbers can enhance the experience – singular in nature – that arises from a more traditional reading. And this dynamism in the big data archive should not be entrusted to the typical algorithm. On the contrary, it must be established with a perspective that does not exclude the sense of the literary experience. The nature of this experience and the research regarding it should adapt to and fit the territory to be colonized, represented by the large catalogues; and, above all, the intervention must be regulated, through the investigation processes, in the definition, establishment and management procedures of these tools. Their purposes are too important to relegate their access to algorithms that are not controlled, or left to chance. It is not only a question of handling the existing tools and catalogues, those that are already given (or imposed upon us). It is imperative that, through our field of study, we take the lead in the design and making of the digital tools and the big data catalogues that not yet to be accomplished, so that in this task we keep the best remnant of a critical tradition that cannot be abandoned.

One line of work, developed by our project, is the use of the possibilities of hypertextuality, establishing a double plane – the textual and subtextual – in the materials, and adjusting the tools for the relational management of the data obtained in searches. Semantic labelling makes it possible to revitalize the text, by adding a plane to the mere sequence of signs and, above all, by introducing a semantic architecture beneath the texts, as finished accomplishments. This, through its logic, enables access to the planes of reading in which quantity becomes quality. In our case, the conceptual base is the construction of authorial notion and

image. Around this, the establishment of a systematic ontology, with the arborescent structure on the three planes defined by the TEI system (classes, attributes and labels), becomes the route for introducing a historical and philological criterion in the consideration of a good number of texts and a dynamic relationship between them, beyond their location in a shared repository. This model makes it possible to explore, for example, the recurrence of a value such as social status in the configuration of the image of the writer or the references to the networks of sociability in which they are registered.

Figure 2: Search results in the SILEM Biographies digital library. http://www.uco.es/servicios/ucopress/silem/buscador/busqueda-pro-final.php?query=%20.//note[@type=%27network%27%20and%20@subtype=%27literary%27]%20and%20%20.//socecStatus[@role=%27writer%27]&cadena=Sociabilidad%20-%20Redes%20-%20Literaria%20and%20Rol%20-%20Escritor&biblio=BIO (June 3, 2023).

This image of search results (Figure 2) serves as an example. The search was done in the library of author biographies (currently with just over two hundred documents that have been referenced, transcribed and labelled), and the aforementioned coding parameters were introduced to locate the passages in which there is mention of the writerly status of the biography subject and also of their inclusion in literary social networks. We can see how in 1622 a man of letters such as Tamayo de Vargas makes these traits clear when outlining the biography of Garcilaso in the introduction to the annotated edition of his works, as well as the language or resources he uses to express that information. The procedure makes it possible to gather not only the passages where express use of such concepts are used in their most recognizable lexicalization, but also those in which the allusions appear obliquely, even before the critical establishment of a notion. At the

same time, one can locate the other testimonies of this subgenre, in which one can extract the same connection and establish parallelisms and differences, as well as observe the density of its use in the established repertory. It is striking, for example, that less than 25% of the texts collected – even though they are biographies of writers in very different historical and publishing contexts – refer to these circumstances, which are so decisive in our consideration of a literary author.

Regardless of the details of a specific proposal, I think we can make use of the qualitative element, which can become prevention against the fetishism of big data. If you'll forgive me the wordplay, this is going from the virtual as dematerialization to the notion of virtuality as potentiality, which refers to a space of freedom and judgement that is still the patrimony of the subject. In the notional field of data mining, it is imperative to bear in mind that the extractive engineering that makes it possible to mine a seam is as important or more so than the actual wealth of that seam.

To be precise, in the field of literary discourse, one cannot abandon the qualitative dimension, with a certain relativization of the productivism of the *big*, imposed upon the vitality of the materials that have to be reduced to *data*. One factor in this consideration is the fact that we can work with a finite corpus, more or less extensive, but stable and established. Think of the complete works of an author or the texts that make up a genre, once there has been consensus on their definition and scope. It is essential, therefore, for there to be a specific adjustment of the statistical and projective models typical of the usual mining of big data, paying attention to the qualitative that resides in the singularity. As with human beings, texts, no matter how much they are digitized and added to databases, should not have their distinctive traits annulled – or if at all, only in a methodological and functional way, and this journey requires an end as much as a starting point. And at both points we must find the text, that small redoubt of reality.

Coda

Neither apocalyptic nor integrated – we must be travellers who should always be somewhat suspicious in order to stay alert; we survey a battlefield with unequal forces, the little philological David and the giant Goliath of computing. In terms of epic battles, at Little Big Horn Colonel Custer's cavalry perished with their boots on. Epic propaganda refused to let the example given by the military disaster show its tragic importance. It was so not because of the quantity of men lost, but because of a prestigious officer's sin of hubris, which led him to enter unfamiliar

territory without sufficient precaution. Today, the role of the warriors of Crazy Horse could be undertaken by the scientists of the *quadrivium*, the tribes of the number, for their ability to lay waste. But we can and should turn this around, and, if big data advances with the capability of an army and with its uniformity, we can reclaim the value of savagery and seize a space of our own. Like the poem, small and singular,[9] we can start to operate in a kind of *little big data*, almost craftlike, humanizing the concept without rejecting the tools that form it, taking possession of the tools of big data and adjusting them to the dimension of our needs, which is also a way of maintaining consideration for the singular nature of our purpose.

Bibliography

Cavallo, Guglielmo (1975): *Libri, Editori e pubblico nel Mondo antico: Guida storica e critica*. Laterza: Roma.

González, Déborah and Helena Bermúdez Sabel (eds. (2019): *Humanidades Digitales: Miradas hacia la Edad Media*. Berlin/Boston: De Gruyter.

Jiménez Heffernan, Julián (2002): "Pequeño, claro y libre: una poética para el poema lírico en la España del siglo XVI", in *Studi Ispanici*, 5, pp. 61–79.

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

Morras, María, and Antonio Rojas Castro (eds.) (2015): *Humanidades digitales y literaturas hispánicas*, in *Ínsula*, 822.

Petrucci, Armando (ed.) (1979): *Libri, scrittura e pubblico nel Rinascimento. Guida storica e critica*. Roma: Laterza.

Ruiz Pérez, Pedro (coord.) (2010): *El parnaso versificado. La construcción de la república de los poetas en los siglos de Oro*. Madrid: Abada.

9 This is the condition given it by the first Renaissance precepts on the genre (Jiménez Heffernan 2002), and it still survives in an underground way within the panorama of literature dominated by greater-length genres.

Daniel Torres Salinas & Ana Gallego Cuiñas

Big Borges: What Can Big Data Show About a Classic Writer on Social Networks?

1 Introduction

The Argentine writer Jorge Luis Borges is one of the most quoted authors in scientific articles from any discipline when it comes to talking about big data.[1] Stories such as "The Library of Babel", "The Aleph", "The Analytical Language of John Wilkins", "Tlön, Uqbar, Orbis Tertius", and "The Garden of Forking Paths", function as examples of the ideas of infinity, accumulation, condensation, causality, virtuality, and simultaneity, which govern twenty-first century digital *dataism*. On top of this we can add the many artistic interventions carried out using the big data of Borgesian works, such as: *Libros de arena* (*Books of Sand*) (2003)[2] or *Cuatro días con Borges en mente* (*Four Days with Borges in Mind*) (2012),[3] by the artist Mariano Sardón; *El Aleph engordado* (*The Fattened Aleph*) (2009), by the writer Pablo Katchadjian; *Remake* (2011) by Agustín Fernández Mallo; and *Borgestein* (2012) by Sergio Bizzio.

This is all a response to the double effect that the poetics of Jorge Luis Borges produces. On the one hand, it is ontological – in a technical and philosophical sense – given that Borges's texts are used to understand the nature, conceptualization and problems of the new digital media, whose paradigm would be the story, "Pierre Menard, Author of the Quixote" (Martínez 2007). On the other hand, the effect is epistemological, as Borges becomes method, which is another way of saying he becomes a model of interpretation – we could almost say he becomes algorithm – of the digital sphere and social media, to the point of being considered a precursor of the internet by the way in which the new technologies seem to have been sketched

1 Carolina Ferrer shows, in a study on the presence of Borges in bibliographic databases, that Borges is quoted – not only quotations but also concepts – more in scientific texts (mathematics, biology, genetics, anthropology, environment, geography, archaeology, sociology, linguistics, etc.) than in works of literature and criticism (2012: 505–506).

2 See https://marianosardon.com.ar/books/books_esp.htm.

3 See https://www.marianosardon.com.ar/day_borges/borges_mind_esp.htmón.

Notes: This study is produced by the Iber-Lab Scientific Unit of Excellence: Criticism, Languages and Cultures in Ibero-America of the University of Granada (ref. UCE 18-03) and by the A-SEJ-638-UGR20 project (FEDER - Junta de Andalucia).

https://doi.org/10.1515/9783110753523-013

out in his work (Brook 1995; Lapidot 1999; Sasson-Henry 2007; Brown 2009; Callus & Herbrechter 2009; Acuña Zumbado 2012; Newhouse 2019). In this area, the paradigmatic story is "Kafka and His Precursors", due to the fact that, after the rapid emergence of digital techniques and of Artificial Intelligence, we read Borgesian fiction in another way.

Yet Borges is not only a "classic" (Calvino 1993)[4] and a cult figure in humanist, scientific and technological discourse, but also – paradoxically – he is equally revered in mass media, particularly on social media like Twitter. There is no doubt that this media discourse – based on his myth and his oral output – forms part of his oeuvre and of his poetics of fiction. It is therefore necessary to analyse it from both a quantitative and qualitative point of view in order to measure his current impact on media holistically. Moreover, the case study that we hereby present opens up a new line of research in the field of literary studies, which means that we need to expand the notion of the *writer figure* (Gallego Cuiñas 2020) based on the (re)production, circulation and consumption of the authorial image and of the literary text – oral and written – on social networks. It will even help us to reconsider the value of literature in (digital) mass culture and the need to resignify the aesthetics of reception in the era of big data.

1.1 Borges and (digital) Mass Culture

In the decades of the seventies and eighties, Borges had already become a public figure, a writer in demand by communication media from all over the world, and he made regular appearances on the radio, television and in the press (cf. Borges & Ferrari 1992, 1999; Borges & Carrizo 1997).[5] His oral performances overflowed with intelligence, erudition and humour, talking of the most sublime and the most prosaic, making constant references both to books and to personal anecdotes (Bruni 1999; Pauls 2004).[6] And that is precisely what is striking about an author whose written work is a sign of *unreadability* (Gallego Cuiñas 2019): his ability to "trans-

4 Premat states that Borges is read (and is consumed, we would add) as a classic: "In any case, Borges can be deemed to be the leading classic writer of Latin American letters" (2022: 9).

5 Pineda Cachero recounts that in those years he would do up to three or four interviews a day (2002: 52). Annick Louis specifies that this happened above all when he left his job at the Biblioteca Nacional (National Library) and withdrew into domesticity, when he began to receive journalists, students and critics at his home (2020: 276).

6 Even when journalists asked him his opinion on political topics – due to his evident anti-Peronism – or insisted on asking him about his private life – of his relationship with his mother or his controversial marriage to María Kodama – despite his "policy of modesty", as Pauls calls it (2004: 45), being well known.

late" the codes of high culture to mass culture without making distinction and fascinating a non-specialist public.[7]

This peculiar relationship that Borges had with mass media has not received much critical attention (i.e., Bruni 1999; Saítta 2018; Pineda Cachero 2022), at least not specifically. However, there are two fundamental studies that address this issue: the celebrated chapter devoted to it by García Canclini in *Culturas híbridas* (*Hybrid Cultures*) (1990), and the recent article by Annick Louis, "*A momentary lapse of history*" (2020). In the former, Canclini studies the protagonistic position that the figure of Borges acquired in the media in the second half of the twentieth century, and the way in which he contributed to the professionalization of the writer when he reached out to mass culture – like Octavio Paz – with his literary discourse. He then became the authorial benchmark of what a writer in the Spanish language could do with the media (Canclini 1990: 95), of how to construct a reading framework of the actual work and of the self for academic doxy and, simultaneously, for mass media.

Annick Louis, who has worked on the author figure of Borges for decades, for his part has focused on the construction of the public character of Jorge Luis Borges between 1976 and 1986, the figure that articulates a "poetics of the media" in dialogue and relationship with his literary oeuvre (2020: 271). It comes as no surprise, therefore, that his oral work is almost as extensive as his literary production in his later years (Premat 2022: 86). Although it was his fiction – national recognition, its translation first in France and then in the US, the Formentor Prize – that won him fame (being named as Director of the National Library and as professor of the University of Buenos Aires, and awarded various honorary doctorates), it was his media appearances that made him popular (Louis 2020: 281). In what way? For Louis, there are two main factors: his "enfant terrible, controversial and provocative" role that he performed in interviews, and the parallel predominance of his image as a "wise and universal old poet" that prevailed in the media (Louis 2020: 277), particularly after he lost his sight in the mid-1950s.[8] This biographical fact, as Julio Premat notes, changed his way of conceiving literature and brought about his conversion into an oral writer (2022: 85), of himself and of others. Alan Pauls also stresses this transformation and the performative attitude of Borges in the media, where he appeared more and more frequently: "The celebration of his sullenness, the low voice as his hallmark, the fostering of malice and mockery that seep into a laconic phrase" (2004: 47). This idea seems essential to us because it helps to ex-

7 Borges ironized about his fame and the fact that people bought his books and did not read them. This idea is still around today: Borges's fiction is read little due to its intellectual difficulty.
8 An association is frequently made between the coming of Borges's fame and his blindness. Even he himself considers it to be "a defining trait of his position as writer" (Premat 2022: 91).

plain his strong presence on social media today – what we could call the "iconographic capital" of Borges; or, what amounts to the same thing, the way in which his image – his body – is a response to the pristine, romantic imaginary of the writer: a blind old man, solemn and in*genious*, who speaks and does not write – at least not in the traditional way, because the older Borges *dictated*.

However, up until now, the way in which Borges, his image and his oral texts are (re-)produced and circulated in digital mass culture has not been studied. By way of clarification, we understand 'digital mass culture' to mean all that culture that is created, transmitted, and experienced in a digital medium. Mass culture, as understood by Adorno in his *Dialectic of Enlightenment* (1944), has not only expanded exponentially over the last quarter century through the digital medium, but is also dominated by three typical components of this medium: the audiovisual, the aphoristic, and the affective.[9] These three *politics of mass culture* explain both the extraordinary success and the 'conditions of reproduction' of the word and image of Borges in the twenty-first century.

The first component, the iconography, corresponds to the image of Borges that has been assimilated to the prototype – one could also say cliché – of the universal genius: an old man with a walking stick, blind – like Homer himself[10] – and cosmopolitan, who is capable of conversing with an astrophysicist, an expert in the Kabbalah, or with Mick Jagger. The photographs of Borges that appear the most on all media are those of him in old age: "In the last few years of his life, the visual images of Borges framed, completed and defined his texts" (Premat 2022: 97). Yet even in the market, the face of Borges appears on coins, bags, T-shirts, mugs, and comics, to the same degree as Shakespeare and Virginia Woolf. He is also the subject of memes and YouTube montages, quotations from his oral discourse appear on social media, and even in trap music compositions, with his poem "Ajedrez" ("Chess") reproduced in the style of the artist Bizarrap.[11] What does this mean?

9 As Cabot states: "Above all we must point out that the culture of today is not a digital culture – this is only the medium – rather it is an audiovisual culture. Digitality is the last frontier, for now, of a process that is as old as human reason: the reduction of multiplicity to unity, or, if you will, the comprehensive reduction of the complexity of reality – complexity that increases at the same pace that our understanding of it increases." (2013: 24).

10 Premat specifies in this regard: "The identification with Homer, the first legendary writer of the West is explicit, as much as the value attributed to blindness: losing one's sight is a trigger of the writing; a writing that will construct a specific memory, an epic past, to culminate at last in the emergence of a virtual life, a life made of 'a rumour of glory and of hexameters'" (2022: 93–94).

11 See the video uploaded by Diego Palatino in 2022 on his Instagram account, @lectordeltren: https://www.instagram.com/p/CZaJvW3jsfT/?utm_source=ig_embed&ig_rid=2a4191da-5f55-4e31-bf19-241e6922933b.

One immediate deduction is that the classical and pop aura of the figure of Borges prevails over that of his fictional oeuvre in mass media and has become an emblem of "the Great Writer" (Premat 2022: 86).

The second politics is the aphoristic, which is woven together through the use of brevity, exactness and orality. Hence, Borges's oral discourse, born out of the many interviews he gave,[12] functions as a double of his written discourse. For Borges had a great capacity for speaking as he wrote, with sentences that appear simple but contain truly surprising periphrases, paradoxes, and analogies, in which terms that we could call Borgesian appear: "conjetural" ("conjectural"), "vasto" ("vast"), "fatigar" ("tire"), or "acaso" ("perhaps"). The Borgesian practice of the aphorism, both in fiction and in public, gave rise to the making of a documentary, *Borges para millones* (*Borges for Millions*) (1978) – the title, which seems to be an oxymoron, is symbolic of the fusion between high and mass culture in the figure of Borges – based on the utterances that, in the form of seductive sentences, the "maestro" gave in the media. The compendium of Borges's quotations, *Diccionario de Jorge Luis Borges*, that Blas Matamoros published in 1979, likewise arises out of this faculty. This was organized according to traditional and timeless subjects such as love, art, cinema, philosophy, history, the Argentinians, books, literature, politics, religion, tango, and society, and contains the most popular of Borges's quotations, the same that now circulate throughout the internet. The attractiveness of Borges's aphoristic opinions lies in the illusion of truth, wisdom and authenticity that the oral format in which they were uttered, through precise and careful language that makes one forget the artificiality and fictional construction of all media discourse. Hence it is the oral writer who is favoured in social networks, the Borges who is more readable and more reproducible; rather than the writer of stories and essays, who is unreadable and cryptic.

The third and last politics alludes to the era of affective capitalism (Illouz 2007; Santamaría 2018) or a capitalism of the emotions (Ahmed 2011) that greatly strengthen digital mass culture through the hybridization of the sentimental with the commercial. The consequences in the cultural field are clear: on the one hand, art is associated with emotion and fun first, before emancipation and resistance. On the other hand, the more symbolic capital an artist has, the more (economic) value is attributed to their biographical space, thus a shift ends up occurring from the myth to the man. As Canclini had already indicated in the nineties: "What is most common is that the public shift their concentration from the work to the biography of the artist and replace the struggle with forms by historical anecdote." (130). However, with respect to Borges, although the inte-

12 One must remember that for Borges, the interview is a literary genre.

rest in his biographical space increased after he got married to Kodama, his authority in the emotional sphere (his well-known quotations on love, friendship and life) is born out of the iconographic[13] – that is to say, out of the image of the artist "as the representative hero of big emotions" (Canclini 1990: 139) – that comes from the romantic imaginary of the nineteenth century. Thus, as we will see, the Borges that is praised in media is the "communicator" (Lipovestdky 2020: 94) who operates like a guru of feelings or of *eudaimonia* – as a medium of subjective expression and self-legitimization for the users of social media.

In conclusion, in these three politics of (digital) mass culture that explain the visibility of Borges, we find: (i) an increase in the exchange value of the signified over and above the signifier (that is to say, of the content before the treatment and formal innovation); (ii) greater social value of the oral and aphoristic than the narrative and aesthetic Borges; (iii) an appropriation and use value of the myth of the cult writer as mode of subjectivation and affective self-legitimization; and (iv), the persistence of romantic values in the reception and use of the image of the writer that circulates in mass culture.

1.2 The Borges Writer Figure

In the last quarter century, neoliberal capitalism has entered into a new ontological phase, where the market, emotions, the processes of subjectivation and the new technologies have saturated all spheres of life, including literary culture. This has led to a significant change in the legitimization and valuation of the cultural and social status of the writer. Their resemantization and overexposure as a media personality, their setting up as a consumer article and the multiplicity of gestures or scenes that mediate their public (digital) image has made studying the *writer figure* fundamental in the agenda of twenty-first-century literary criticism (Gallego Cuiñas 2022b). One could even maintain that it is one of the most revealing signifiers of the changes that have occurred in the relationship between literature and mass culture. Few social actors depend as much as writers on a context, on a readership and a market, "for what they are and for the image that they have of themselves on the image that other people have of them and of what they are" (Bourdieu 2003: 21). In this media image, in this performance of the public personality that is cause and effect of the meaning of the work, the social value of literature is also at stake. Thus, "the aesthetics are relative to the positions that writers occupy in the field"

13 As Ana Peluffo (2015) shows, scant attention has been paid to Borges's relation with the culture of the emotions within his written work.

(Sapiro 2016: 37) because their public interventions form part of the work and build another interpretative direction for their poetics.

Despite the centrality the writer has acquired in our culture, until the twenty-first century there was no growing interest for *author figures*[14] in literary studies (cf. Díaz 2007; Meizoz 2007; Premat 2009; Louis 2013; Gallego Cuiñas 2015; Fontdevila & Torras 2016), since it was still seen – under an essentialism of a romantic ilk – to be a topic and problem outside of the text, where literary value is not on the line. This explains why the Borges writer figure in the public sphere has not been studied as much as one would expect (with the exception of García Canclini, Louis, Premat, and Saítta, along with Lefere). There are still areas that need covering, as is the case with the uses[15] of the oral texts and of the Borgesian image in social media. This lack of attention to the figurations of the writer perpetuates the artificial separation between high and mass culture, as if the author were not also (re-)produced, circulated and consumed by the global and media market, and as if this did not constitute a(-nother) frame of visibility and of readability. Borges, in contrast, did value this question of "becoming an author" (Premat 2022: 8) and was very attentive to other writer figures in his biodiscourse, where he constantly quoted "authors, author figures, with their gestures, their manias, their idiosyncrasies, as one who quotes texts" (Molloy 1999: 231). What interested him was the myth, the way in which a writer forges an image, as he did, to define "successive fields of production and of reception" (Lefere 2015: 159) that not only self-legitimize but also enshrine it.[16] Moreover, as Annick Louis states, "Borges's early reflections on the question of fame (in essays from *Inquisiciones*, *El tamaño de mi esperanza* and *El idioma de los argentinos*, see Louis, 2014: 353–354) created in him an intense awareness of the implications of the processes of canonization, which he explored and translated into textual forms and into positions throughout the rest of his career" (Louis 2015: 18).[17]

14 In the critical field, the denomination "author figure" predominates, but we prefer to speak of the *writer figure* to transcend the notion of authorship tied to the (intellectual) property of the text and place emphasis on the specific exercise of the literary profession inside the cultural market, whose forms of rating value are different to those of other artistic professions (Gallego Cuiñas 2020).

15 We understand the category of 'use' not only in the Marxist sense but also as Virno (2017) proposes it: an "appropriation" that emanates from the relationship between life and language, between subject and object.

16 Remember that, for example, in "Presencia de Miguel Unamuno" we note how he reads his texts under the protection of the image of the writer, which is the one he projects in the work.

17 In this sense, Louis speaks of two periods in the construction of the Borges-author: one from 1919 to 1955, and the other from the return of Peronism in 1973 until his death in 1986 (2015: 18). She also takes brilliant charge of exploring the way in which the textual fiction of Borges forms an image of the writer (e.g., The Aleph, Tlön, Pierre Menard, etc.). For Lefere, the most autobio-

In (digital) mass culture, the modes of authorial (self-)representation are no longer confined to the book-text, but to the texts and images that make up the media figure of the writer (in interviews, lectures, notes, photographs, social networks, and so on). The social value thus shifts from the work, from the book-text that becomes a *zombie* category, to the author, the image and the oral-text. This entails thinking about an *epistemology of the writer figure* that needs to be addressed through three gestures (Gallego Cuiñas 2020):

- *the posture:* the way in which the writer occupies a position in the market and is visible in various instances of mediation: publications, translations, teaching, festivals, Master's degrees, conferences, social networks, et cetera;
- *the pose:* the performative strategy of image circulation and the scenography they deploy in the public sphere;
- *the myth:* consequence of the positive reception in academia, of the legitimization of high culture, and of meeting the levels of expectation of readers.

The case of Borges is revealing in this sense since he not only represents these three epistemic instances to perfection, but has also been raised up as the postmodern paradigm of the classic *writer figure* who triumphs on social media. It is evident that the Borgesian "pose" of the seventies and eighties has contributed significantly to furthering the "myth" of the erudite and cosmopolitan writer – described in the previous section – that is still being reproduced in digital mass culture almost forty years after his death. This becomes a specific digital "posture" on social media like Twitter, where Borges occupies the public place of the romantic writer (the figure of the writer par excellence), but also that of the visibility of the literary in mass media, which in spite of its loss of social influence is still tied to positions of prestige that the users of social media perpetuate with their tweets.

1.3 Borges and Social Media

The advance of the creative industries, the democratization and spectacularization of culture, have all favoured the proliferation and professionalization of new instances of mediation of the literary such as social media. The figuration of the writer, as we said earlier, is no longer disseminated only in texts of fiction, oral discourses, newspaper articles, and interviews; but in posed photos on Instagram,

graphical book by Borges is *El Hacedor* (2015: 154), because it is where he established his last image.

in memes, videos, blogs, bots and social networks. From among all those symbolic and material "I's" that proliferate in mass digital culture, the one that has gone the most unnoticed by literary criticism concerns precisely the topic that we are addressing: social media, that exceptional device for literary promotion and for the construction of the writer figure today (cf. Gallego Cuiñas 2022a).

There is therefore a need to go further in the analysis of the figurations of the writer based on social media such as Twitter – where the most attention is paid to literary culture and the emotional[18] – through *dataistic*, sociological and critical reading, to see how this network operates with the literary field, through mechanisms of reproduction and appropriation of certain writers and texts. Social networks function with an economy of affective representation that is modulated through the image and the (auto)biographical discourse of the writer, two textualities that certify and signify not only the *authenticity* or singularity of the authorial myth, but also that of the (re)producers and consumers themselves. Indeed, the first thing that we observe is that tweeters quote Borges because it causes an intelligent "reading effect" in whoever uses it, even if they have not read him, along with emotional capital in search of social recognition: "I read/know Borges". As Sosa Escudero explains, Borges has always caused "a contradictory sensation: on the one hand, that one ought to read him: to do so requires an intellectual education that only a chosen few possess, validated by a kind of secret sect that grants permission to "be Borgesian" after some initiatory manoeuvres" (2020: 16). For this reason, the Borges that is circulated on networks is the oral, aphoristic and affective Borges: the consumable and readable Borges, who goes well with a distant reading for life; rather than the literary work that is resistant and hard to read, which is more suited to close reading for specialists in literature.

Nevertheless, the digital space seems to be a beneficial medium for critical reflection because it makes it possible: (i) to measure the visibility of a classic writer in current mass culture, which allows us to revitalize the dialectic between literature and society; (ii) to think about the category of the writer figure via new formats like social media, which perpetuate the romantic icon and extol the short form and emotional content; (iii) to compare the 'being an author' (the writing) with the authorial image (writer figure) to demonstrate the way in which the author as intellectual property of the work – which does not sell – is becoming subordinated to the writer as actor of the writer-subject that sells (or is sold) as a work, the first example of which, in the history of Spanish-language contempo-

18 As Helgueta Manso argues, Twitter is one of the "predominantly (hyper-)textual, and therefore literary, platforms, as opposed to the audiovisual applications" (2022: 45), where the textual has a secondary role. Twitter, however, has as a driver the affective dimension, in its positive and negative sides, not like Facebook, not allowing insults or direct confrontation.

rary literature, is undoubtedly Jorge Luis Borges. Thus the writer figure of Borges is a material sign of a particular idea of the (classic) writer, of literature and of subjectivity that operates in the mass digital culture of our times, which has become a vital route for the construction of the social value of the literary, crystallized in the ability of reproducibility – digital – of a writer-subject, and no longer of a work. The writer figure of Borges is only comparable to Shakespeare's, to the point at which both the global academic reception of his texts and his presence on social media have turned him into one of the most recognized contemporary writers and known as an intellectual – as 'classic' – throughout the world.

1.4 Objectives

Borges's presence is increasingly prominent on Twitter, a medium on which messages are constantly being spread about him, and different appropriations of his image and quotations constantly appear. This undoubtedly affects the construction of his authorial figure and the reception of his work. Taking this as our basis, the framework objective that assembles and guides this study is focused on the analysis of the diffusion, reception and assimilation of the figure of Borges on Twitter, by means of informetric techniques (Moed, 2017) and Big Data (Zgurovsky & Zaychegnko 2020; Domingo Barroso et al. 2021) that allow us to construct a theoretical framework of readability. On the practical level, we have organized and divided our general objective into three specific sub-objectives:

- *Volume:* first, to find out the volume and frequency of tweets on Borges and to determine some basic characteristics, such as the language, and the measurement of the diffusion and interactions that generate such publications.
- *Content:* second, to find out what is shared, concerning ourselves with two of the aspects included in the three politics of mass culture: the aphoristic, on the one hand, and the iconographic, on the other.
- *Community:* third, we focus on identifying who are the actors – in this case Twitter accounts – that are the most relevant when spreading Borges's work, identifying their basic characteristics (type, sector, followers, etc.).

Our results shed light in two complementary directions, one theoretical, the other methodological. From the theoretical perspective, our investigation expands the scope of the critical discussion concerning the relationship between literature and mass culture through the way in which classic writer figures are reproduced in social media. This appears to be an extraordinary chance to rethink, in turn, the theory of reception, the writer figure, and the social value of the literary in the digital medium. From a methodological perspective, we confirm the viability

and advantage of using the most advanced informetric techniques and big data for the gathering of mass data on a writer – in this case Borges – on a social network. In this way, we show the different analytical possibilities, both epistemic and technical, that Twitter offers for their application and development in later research that we could include within the general category of "(Literary) Cultural Analytics".

Lastly, we have organized this paper into three sections: one on the method, one on the results, divided into three subsections (quotations, audiovisual content, and community), and the final section with our concluding thoughts.

2 Big Data Methods

In order to carry out the proposed analysis, we have used the big-data app, *Tractor*, for the data capture (Hurtado et al. 2021), to download the tweets that mentioned Borges during the time frame 01/01/2018 and 12/31/2021. We thus concentrate on a four-year period. In order to identify the tweets, we have sought the different ways in which the Argentine author was quoted without generating noise, which are as follows: "Jorge Luis Borges", "Jorge L Borges", "JL Borges", "Borges JL", "Borges, Jorge Luis" and "Borges, Jorge L". We recovered tweets published in all languages, although in certain studies, such as the aphoristic, only tweets written in Spanish were utilized. We refer to these sets of tweets as the "global collection" and "collection in Spanish", respectively. Lastly, regarding the search strategy, it should be specified that we have taken into account the typologies of conventional tweets but also those typologies that entail the diffusion and interaction of users, such as replies, retweets and retweets with comments (quotes).[19]

The total number of tweets recovered, including the four typologies, amounts to 205,216 (global collection). These tweets were exported in *.csv* format, subjected to computer-aided normalization processes, such as for the quotations, which involved the design of a semi-automatic routine for their identification and unification. To analyse the content and communities, the main tool used was *Graphext* software,[20] another paid online platform directed at big data and knowledge discovery. Thus we were able to transform the data, visualize and examine them to discover patterns, tendencies and for their subsequent critical elucidation. More specifically, *Graphext* enabled us to learn which emoticons and hashtags were used the most, analyse account biographies, establish the professional sector of

19 See: https://help.twitter.com/es/using-twitter/types-of-tweets.

20 See: https://www.graphext.com/.

the users and, lastly, for the analysis of the communities, to create an illustrative network of co-mentions. In other words, in this network, two accounts will appear linked in the visualization if they have simultaneous mentions, with which we have been able to define a series of communities and/or similar accounts (Robinson-Garcia et al. 2019).

3 Results

3.1 Volume: The Big Numbers of the Borges Phenomenon on Twitter

Table 1: Annual evolution of the number of tweets published on Borges in the period 2018–2021, according to year and language of publication.

Tweets: Language of publication	Year 2018	Year 2019	Year 2020	Year 2021	Total 2018–2021	% language
Spanish	32480	37099	39566	34902	144047	70%
English	7197	7404	7636	7345	29582	14%
Portuguese	1338	1517	2002	1718	6575	3%
Italian	1407	1480	1403	1512	5802	3%
Turkish	999	1198	1367	1282	4846	2%
French	993	1021	1319	956	4289	2%
Dutch	947	876	1156	1092	4071	2%
Catalan	708	780	895	724	3107	2%
Other languages (41)	779	788	840	890	3297	2%
Total nº tweets ▸▸	**46848**	**52163**	**56184**	**50421**	**205616**	**100%**

The global set of tweets that mention Borges amounts to 205,616, written in 49 languages (Table 1), where Spanish is the most represented with 144,047 (70%), followed by English with 29,582 (14%). The other languages obtain values equal to or less than 3%, and therefore their representation is not particularly significant. The gross total of annual tweets is fairly stable, with figures of around 50,000 tweets, and with 2020 being the year with the most Borges messages tweeted, at 56,184. The average number of daily tweets (Figure 1) was 141, with prominent peaks on key biographical dates for Borges, such as the day of his birthday, the 24th August (2,300 average tweets), or the date of his death, the 14th June (1000 average tweets). Curiously, on World Book Day, the 23rd April, Borges also received a lot of attention, which confirms him as the epitome of

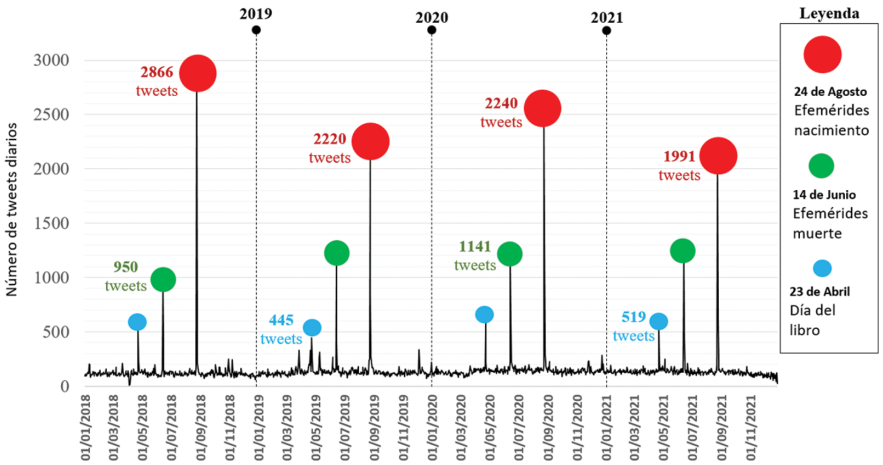


Figure 1: Daily evolution of the number of tweets on Borges from 01/01/2018 to 01/12/2021, highlighting those days (anniversaries of birth and death, and World Book Day) when the greatest number were written.

the classic writer figure: the great reader, surrounded by books, possessed of immense erudition.

Having determined the volume and frequency of tweets on Borges (Figure 1), we proceeded to examine the number of interactions and reactions generated by those tweets. For this we calculated indicators, both for the global collection of tweets (blue bars), and for those published in Spanish (orange bars). In 2.1 we can see that the messages have been spread through a total of 1,115,341 retweets; moreover, on 3,582,837 occasions they have caused a reaction in the reader, since they were added to favourites.

Almost all the interactions and reactions were in Spanish, which clearly exemplifies the predominance of this language community. Furthermore, examining the averages, we can see that this language also accounts for the most retweets and tweets added to favourites (2.2). We can therefore affirm that not only are there a large number of tweets published in Spanish on Borges, but also these are the most shared and appreciated by users. In short, what these figures demonstrate is that Borges is a stand-out model in the Ibero-American twitter community, which is why in the following analyses, we focus exclusively on their collection of tweets.

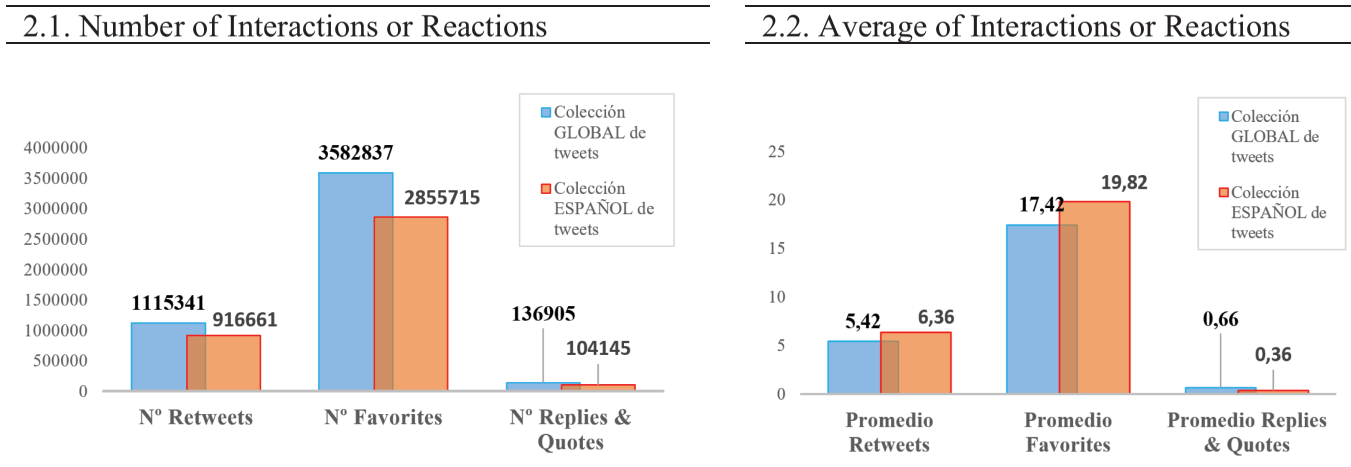


Illustration 2: Global indicators of tweets related to diffusion and interaction, according to both the global collection of recovered tweets and the collection of tweets in Spanish.

Table 2: The 20 quotations of Borges that have received the most attention/dissemination on Twitter, according to various indicators of diffusion: retweets, favourites, replies and commented retweets.

Borges quotations mentioned on Twitter	**Indicators of dissemination (interactions and reactions)**				
	Nº Tweets	**Nº Retweets**	**Nº Favourites**	**Nº Replies + r. with comment**	**Total Attention**
1. The pursuit of serenity seems to me to be a more reasonable ambition than the pursuit of happiness.	1091	26991	89066	2113	**118170**
2. I won't speak of revenge or forgiveness; forgetting is the only vengeance and the only forgiveness.	1995	27994	80212	1672	**109878**
3. Of the various tools invented by man, the book is the most amazing; the rest are extensions of his body . . . Only the book is an extension of the imagination and memory.	1602	26643	66279	1667	**94589**
4. Don't speak unless you can improve on silence.	683	19130	63434	1489	**84053**
5. One can give what one does not have. For example, a person can give happiness and not be happy; can scare and not be scared. And one can give wisdom and not have it. Everything is so mysterious in the world . . .	519	20554	58964	1009	**80527**
6. One can fake many things, even intelligence. But one can't fake happiness.	643	19376	56508	1015	**76899**
7. I owe you the best and perhaps the worst hours of my life, and that is a bond that cannot be broken.	720	17191	56605	851	**74647**
8. There are defeats that hold more dignity than a victory.	860	16686	46876	1083	**64645**
9. Of course I believe in dreams. To dream is essential, it could be the only real thing that exists.	446	12724	43702	863	**57289**

Table 2 (continued)

Borges quotations mentioned on Twitter	Indicators of dissemination (interactions and reactions)				
	Nº Tweets	Nº Retweets	Nº Favourites	Nº Replies + r. with comment	Total Attention
10. Neither have I lacked the friendship of a few people in my life, which is what matters. I don't believe I have a single enemy, or, if I do, they never made me aware of it. The truth is that no one can hurt us except the people we love.	216	13165	37676	996	**51837**
11. If a book is tedious for you, don't read it – it hasn't been written for you. Reading should be one of the forms of happiness.	141	12437	35256	834	**48527**
12. I always imagined that paradise would be some type of library.	1079	8659	31338	1563	**41560**
13. Gratitude is one of the highest forms of being.	194	9264	31377	831	**41472**
14. Don't hate your enemy, because if you do, you are in some way their slave. Your hatred will never be better than your peace.	222	9589	29557	480	**39626**
15. When one hates something, one thinks about the other constantly, and, in that sense, one becomes their slave. The same thing happens when we fall in love.	254	7803	28851	563	**37217**
16. We have the right and the duty of hope.	113	7596	26264	590	**34450**
17. "Journalist: Do you think young people are interested in politics? Jorge Luis Borges: I don't know. I was never interested in politics. I'm more interested in ethics. I think that if everyone acted ethically that could have a very large political effect."	113	8975	23936	504	**33415**

Table 2 (continued)

Borges quotations mentioned on Twitter	Indicators of dissemination (interactions and reactions)				
	Nº Tweets	Nº Retweets	Nº Favourites	Nº Replies + r. with comment	Total Attention
18. Friendship does not need frequency; love does.	318	6591	23928	517	**31036**
19. I would like a minimal state. I lived in Switzerland for five years and there nobody knows the president's name. I would propose that politicians were not public personalities.	149	7734	22292	438	**30464**
20. There are communists who state that being anti-communist is being fascist. That is as incomprehensible as saying that not being a Catholic is being a Mormon.	326	9178	20611	536	**30325**

3.2 Content: Scope and Characteristics

3.2.1 Of Aphorisms and Affects: The Big Quotations of Borges

When we focus on the analysis of the tweet collection in Spanish, both initially and using *Graphext*, we can quickly detect that one of the most disseminated types of content are the quotations. Hence we undertook the task of identification and homogenization to learn their real weight. In total, we identified 98 different quotations, which have been quoted on Twitter 40,255 times, which represents almost a third of the tweets published in Spanish. In other words, one of every three tweets on Borges has the aim of sharing a quotation of his, thus becoming the essential Borges content on Twitter. These tweets, moreover, have received a great deal of attention and reception in the medium, since they have been retweeted 482,861 times (52% of retweets in the Spanish collection) and marked as favourites 1,437,644 times (49% of favourites in the Spanish collection). These data verify the validity and social value of the Borgesian word in (digital) mass culture.

The most seen Borges quotation is a true aphorism, tied – as we noted at the start of this paper – to the management of subjectivity and of the emotions: "The pursuit of serenity seems to me to be a more reasonable ambition than the pursuit of happiness." The entire amount of different indicators of diffusion (retweets, favourites, etc.) comes to 118,170, a high figure that denoted that it has had an extraordinary readership. In this regard, if we consider that it has been retweeted 26,991 times, we can do a small arithmetic calculation of the potential audience of this Borgesian aphoristic message. If we begin from the basis that each person that has retweeted it has an average of 150 followers (a conservative estimation – see the followers column, Table 3), this single quotation could have had a remarkable reception and audience of 4,048,650 users.

This is not the only Borges quotation that has had an impact. In Table 2, we compiled the twenty most popular quotations of Borges, which give a very clear pattern of expression, since most are related to the short and sentimental form, which adheres to the concept of personal "development" or "growth", typical of emotional capitalism and of universal teachers such as Seneca, Confucius, or Buddha. Examples of these are the following Borgesian aphorisms: "Don't speak unless you can improve on silence"; "One can fake many things, even intelligence. What one can't fake is happiness"; "Gratitude is one of the highest forms of being"; "We have the right and the duty of hope"; and, "Friendship does not need frequency; love does." We can also identify some subthemes that are repeated, particularly in reference to books and libraries. The quotations on this theme are published most often on 23rd April, which is World Book Day (Figure 1). One example is the quotation, "I always imagined that paradise would be some kind of library", which is fourth in the total number of tweets in our ranking.

Thus, the identified quotations operate as universal dogmas that are perfectly suited to rapid consumption and to the character constraints of Twitter messages. Their content, moreover, is sufficiently impersonal that it would be possible to attribute them not only to Borges but to any philosopher or intellectual in the world, even to famous authors of self-help books like Eckart Tolle with *Stillness Speaks* (2003) or Rafael Santandreu with *Las gafas de la felicidad* (*The Lenses of Happiness*) (2015). It is no surprise, therefore, that the quotation with the most retweets occupies this realm: "I won't speak of revenge or forgiveness; forgetting is the only vengeance and the only forgiveness." Borges's power lies in his erudite figuration, in his myth of genius, that he has knowledge not only of literature and culture but of morality. Borges the opiner, who would today be quite the *influencer*, for in his media discourse he condenses the elemental structures of seduction: "beautiful rhetoric, slowness, ambiguity" set to serve an "emotional branding" (Lipovestdky 2020: 21), which is what triumphs on social media. Thus his success on Twitter comes from the spreading of aphoristic quotations uttered in his oral biodiscourse –

particularly in television interviews and in the press – rather than in his written literature, because the creation of emotions, swift and collective, prevails over intellectual reflection, deliberate and individual.

3.2.2 Iconography, Audiovisual Content and Notable Messages

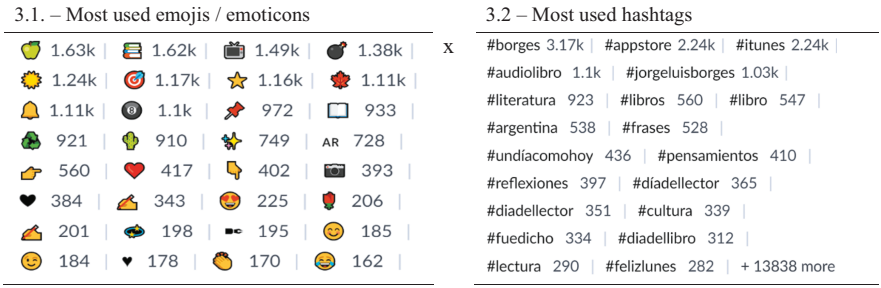


Illustration 3: Emojis and hashtags included in the tweets on Borges written in Spanish.

Examining the semiotic elements that accompany the tweets can help us to accurately trace the outline of Borges's reception in (digital) mass culture. Illustrations 3.1 and 3.2 show the emojis and hashtags that appear when Borges is mentioned in the Spanish collection. For example, an apple as a symbol of Apple alongside the logo of iTunes indicate the privileged medium of technical (re-)production used to mention Borges. Likewise, the most used emojis are associated (in order of frequency) with the following elements and contexts: books; television (which underlines the proliferation of Borges's image in the audiovisual media); a bomb or a target (which stress the nature of certainty and truth that his quotations have); and at the same time affective aspects such as a sun, a star, or hearts. The hashtags, for their part, show the relevance of his quotations (reflections, thought), as we saw in the previous section, while labels connected to the recommendation of his work (audiobook, books, reading) also abound.

As well as this, a significant proportion of the Twitter messages tend to be accompanied by videos and images. We have identified a video that illustrates the huge reach of audiovisual content: we are referring to the tweet on Borges posted by the account "literland" (*Community of reading and literature lovers*) on 19th February 2019 (literland [@literlandweb1], 2019). This has text entitled: "Diferencias entre amor y amistad según Jorge Luis #Borges" ("Differences between love and friendship according to Jorge Luis #Borges"), and shares a video, one minute in length, that received 8,175 retweets and 16,000 likes. On Twitter

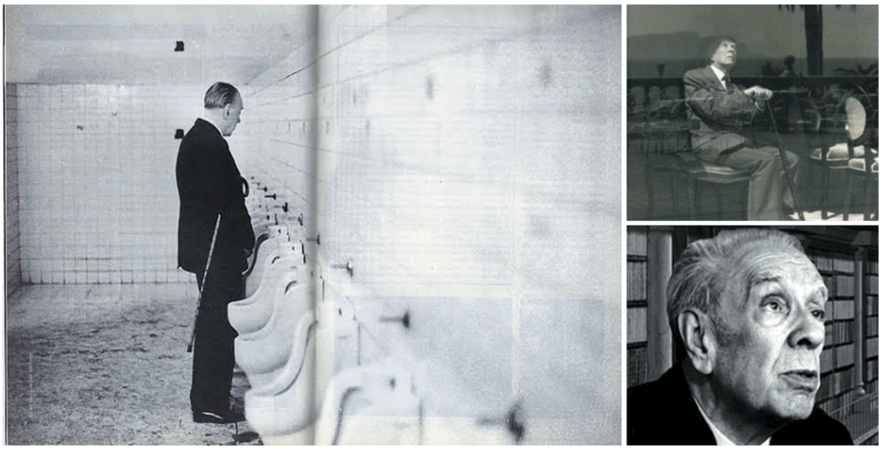


Illustration 4: Photographs uploaded with some of the tweets that have been shared the most.

alone, it has been played 322,000 times, and if you look for it on YouTube, it has more than 900,000 visits.[21] This video could be considered the most popular of Borges, and brings together precisely the three politics of (digital) mass culture that we have discussed: the iconographic figure of the classic, blind and lucid writer; the use of oral discourse; and the expression of affects, whose maximum signifiers are precisely love and friendship.

The analysis of images likewise gives us another side of the multifaceted digital diffusion of the Borgesian universe (Illustration 4). In this category, we have found the photograph of Borges that has the greatest circulation on networks, which is one in which the author appears to be using some urinals (Escritores haciendo cosas ("Writers doing things") [@CosasEscritores], 2021), and looks as though he was caught unawares. The photo is from 1973, when Borges went to Mexico for the first time, and was taken in the toilets of the fabled Colegio de San Ildefonso, where he recorded a television programme. As the photographer Rogelio Cuéllar recalls, Borges heard the camera shutter, but instead of getting angry he took it with good humour and did not censure the photograph. In this image, as you can see, Borges still represents the icon of the elderly writer, blind and with walking stick, but here showing his human side. Thus, although we have seen that the quotations occupy a central place in the reception of Borges, the audiovisual content and its iconographic value are also significant.

21 See: https://www.youtube.com/watch?v=7K-Hk1qt_mk.

3.3 Digital Community around Borges

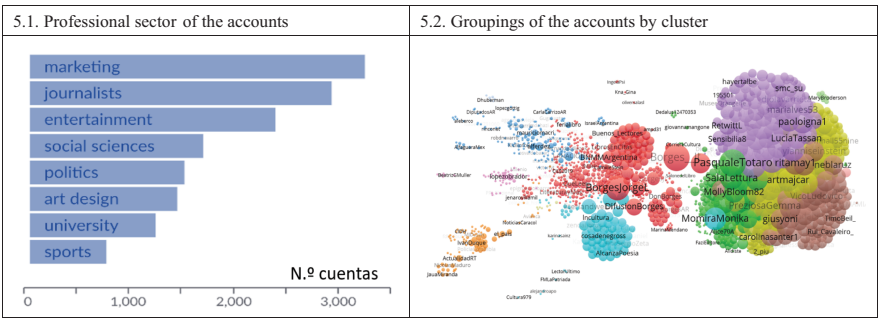


Illustration 5: General and summarized overview of the professional sectors the Twitter accounts belong to and of the groupings/clusters of them according to their interactions through retweets and annotations.

The Twitter community (Illustration 5) makes up a discursive and social ethos that is of interest to literary studies due to the high circulation of cultural content that exists on this network. If we focus on the accounts that bring together their community of reception, we see that there is a total of 85,960 active users, although 63,918 have only published one tweet. If we draw a parallel with the consumers of books, these users would be akin to those who buy just one book by the author and do not declare themselves to be fans or specialists. If we consider the classification that *Graphtext* gives according to the user biography, we can get a clear idea of the sector the biggest tweeters of the Borgesian message belong to. There are many accounts related to marketing and content marketers, above all as a marketing strategy to gain followers and, also, the world of communication, where we can identify many Argentinian newspapers that use Borges as a form of enticement.

There are also communities linked to entertainment, social sciences, politics, art, sport, business, photography, medicine, economy, travel, and video games. In an intermediate position we find the academic community, connected to the "university". On the network, we show how the accounts group together according to their interactions, thus verifying that there is a connected community that spreads content and comments on it, channelled through specific accounts on Borges and other literary accounts from Argentina (red cluster). In addition, in the large cluster on the right, we can make out personal accounts linked to various countries. Ultimately, we can state that the (digital) Borgesian community is broad, made up of sectors with diverse aims and interests, of which a relatively small proportion is able to interact and stay connected.

If we now pause to examine the famous personalities who retweet the most widespread messages on Borges, we can again identify a great deal of heterogeneity in the profiles. The tweets that stand out first are those of the president of Argentina, Alberto Fernández (Alberto Fernández [@alferdez], 2019), announcing the donation of 6000 books from the Borges collection (4037 Retweets; 24.8 Likes); and of the president of Mexico, Andrés Manuel López Obrador, who made a laudatory commemoration on the anniversary of Borges's death (2757 Retweets; 12,200 Likes). Then there is the use of the term, "Matria", which Borges used in a debate in Spain, revived (Ángel L. Hernández [@Angel_L_Hern], 2021) as a sign of authority and legitimacy; and the mistake made by the king of Spain, Felipe VI, when getting Borges's name wrong, saying José Luis Borges (Juan Carlos Monedero [@MonederoJC], 2019). Curiously, in these two cases, the tweets are linked to Podemos, a left-wing Spanish political party that campaigns for more and better public education and the promotion of culture and reading as instruments for critical emancipation.

Lastly, in Table 3 we show the most relevant accounts according to the circulation their tweets have attained. We should explain that that it is not always the accounts that post the most on Borges that are the most relevant.[22] The first is Cultura Bang, which has posted 482 tweets on Borges, with aggregate indicators that amount to a total of 335,283 retweets and interactions. In this list, which already has a marked cultural and intellectual character, the accounts dedicated to the dissemination of literature and reading, such as Literland, El Lector, Libros y Escritores, Cementerio de Libros, and Letras Breves, predominate. These are serious accounts, usually with miscellaneous content, but with a large following of tweeters, as illustrated by the 757,735 followers of Literland. We can also find a few personal accounts on this list, such as the Argentine film director Juan José Campanella, the Colombian journalists María José Castaño and Félix de Bedout, and the director of communication of the publishing house Planeta, Laura Franch from Spain. In these cases, the name of Borges is spread by well-known personalities, tied to the world of culture and, as can be seen, from different Ibero-American countries.

22 These are some of the accounts that have posted the most tweets about Borges: importantbot = 1254, LibretoStar = 2349, Autoayuda_Es = 1173, jinyounglandss = 1002, Libromovil = 793.

Table 3: Twitter accounts / users that have reached the greatest propagation with the publication of tweets on Borges. Only those that write in Spanish have been selected.

Account	Aggregate Indicators of Diffusion & Interaction					
	Nº Followers	Nº Tweets	Nº Retweets	Nº Favourites	Nº quotes Replies	TOTAL
Cultura Bang	191065	482	67774	267027	3397	335283
Literland	757735	121	69872	211277	4994	281270
El Lector	501716	352	63333	205548	3502	269233
Libertario	163395	135	69704	193971	3533	263810
Fuedicho	445333	229	53869	158063	3170	212161
Páginas Redondas	142181	73	28758	85773	3000	114604
Poetas Hispanos®	112032	107	26167	61651	1547	87925
Libros y escritores	88329	37	21383	49389	1044	70809
Alexis Pérez	87694	670	12136	34152	568	46958
Winston	154150	14	11915	28430	802	40359
Cementerio De Libros	225609	42	7597	25345	203	32984
Alberto Fernández	2155004	1	4047	24927	3273	28975
¿QuéLeer?	1929333	149	7259	23841	798	31249
El reinado de las flores	28745	297	6906	23491	364	30694
Maria Jose Castaño	47198	28	6283	22947	1337	29258
Cúpula de Libros	191486	12	7309	18440	185	25761
laura franch	80011	31	4849	18730	282	23610
Jorge Luis Borges	12788	377	6323	16624	334	23324
Ana Bolena	79569	56	4391	18408	519	22855
Juan José Campanella	886143	1	5096	15825	372	20922
Andrés Manuel	8376274	1	2757	12200	2037	14958
Félix de Bedout	2454515	4	2797	13177	543	15978
Escritores haciendo cosas	18337	3	1391	13631	379	15025
Letras Breves	177588	8	3969	9402	260	13379
Eres Inteligente	780616	14	3798	9544	119	13356
La Parada Poética	59212	73	3296	9845	122	13214
Buenos Aires en el recuerdo	90945	25	2019	9612	188	11656
Cristina CR	17184	14	1957	9704	98	11675
¿Por qué es tendencia?	1100509	8	426	10694	280	11128
Leer es Vivir . . .	79372	57	3067	7843	84	10967

4 By Way of Final (open) Reflection

In this chapter, we have studied a classic writer such as Borges through informetrics, a hyperquantitative perspective, and using a digital communication medium such as Twitter, which is usually overlooked in literary studies. Thus we have been able to measure the diffusion and reception of the literary message of Borges

and approach a reconfiguration of the concept of 'writer figure' thanks to the analysis of the volume, content and communities that generate tweets about Borges. Our analysis of 205,216 tweets has shown that the practical application of big data to literature opens up a whole field of study and offers an opportunity to reformulate and update classic concepts of literary theory. In this regard, our first conclusion is that the concept of literature tied only to the category of writing published in book form is insufficient to grasp the new modes in which "literary culture" (Gallego Cuiñas 2022) is expanding today, through other (digital) codes and values that can be analysed with the help of *dataistic* tools, as our study has been able to demonstrate.

In the specific case of Borges, we can confirm that on Twitter the same phenomenon is taking place as in academic production and in the mass media, whereby his name is synonymous with erudition – "insolent, exasperated, exacerbated" (Premat 2022: 69), encyclopaedic, and transdisciplinary – which guarantees the impact of his image and of his discourse in (digital) mass culture, governed by "the supremacy of the law of being pleasing and emotionally moving" (Lipovestdky 2020: 17). It is evident that the oral Borges of mass culture, with his aphoristic potency, is winning the battle of social value over the Borges that circulates in books and is praised in academia for the aesthetic values of unreadability, intertextuality and interdisciplinarity. We have also shown that the quotations of Borges that are reproduced on Twitter crystallize the stylistic traits of his poetics of fiction: erudition, concision, parody, efficacy, linguistic precision, humour, use of the oxymoron and of the analogy, and so on. Hence the oral, audiovisual and media discourse of Borges, associated with "intelligence capital" (Lipovestdky 2020: 258),[23] morality and the ontological revelation of the truth, strengthen his own writerly myth and vice versa. In other words, they expand and enrich his poetics.

What implications does this have for criticism? On the one hand, the need also to expand its objects of study, as occurs with 'writer figures', in order to legitimize them as literary episteme and to think about them against the backlight of the notions of 'author', 'literary work', and 'reader'. As Julio Premat states, the iconic figure of Borges "is inseparable from his writing [. . .], therefore, it fulfils a function in the reading of the texts" (Premat 2022: 97). On the other hand, the importance of renewing the aesthetics of reception through sociology, (digital) mass culture and big data, not as a mere empiricist record of audiences, public taste and opinion (García Canclini 1990: 125), but through the way in which a non-specialist (digital) community co-produces meaning by appropriating an author, or better still, a

23 In this regard, Premat writes: "Borges puts us before the obligation to interpret, to reason, which is why his readers feel that they are on the same level as the author. The reading of these texts gives us the conviction, pretty difficult to define, of being intelligent, of being almost as intelligent as him" (2022: 153–154).

name and a corpus that function as paratext of the work. These communities are not interpretative, there is no textual exegesis; rather, they are reproductive of certain content, which arises from a selective gesture that simultaneously constructs social and literary value.

What, therefore, do the communities of tweeters consume when they quote a tweet on Borges? An *augmented* Borges. There is no doubt of the strong impact the Argentine has, not only on social media but in academic discourse, which turns him into a cult or pet writer for high culture and for mass culture, whose separation – as we said at the beginning of this chapter – is revealed to be artificial. Because both appropriate his published and/or oral writing out of the same "literary forms, which are, in Borges, inseparable from the aesthetic effect" (Premat 2022: 13). The joining of academic value – symbolic and experimental capital – and media value – iconographic and affective capital – is interwoven with the correspondence that occurs between his life (his iconographic image) and his work (his oral texts), as a paradigm of the romantic utopia of writer that is commercialized in digital mass media (Illouz 1997), which the literary field should not turn its back on. Thus social media acts as an ideal laboratory for practising theory with the quantitative results that the big data analysis provides. Through Borges, we have tackled the relationship between literature and (digital) culture today, which is set up as a new and productive route for sociological interpretation for the "literary criticism of value" (Gallego Cuiñas 2022) of the twenty-first century. In conclusion, what we have attempted to demonstrate in this study is that the values of Big Borges are many and varied, and thus many varied critical, literary and dataistic parameters are needed to give a (good) account of them.

Bibliography

Acuña Zumbado, Eduardo (2012): "Trazos proto-hipertextuales en la narrativa moderna latinoamericana: 'La biblioteca de Babel", in *Hispania*, 95, pp. 640–649.

Ahmed, Sara (2004): *The Cultural Politics of Emotions*. New York: Routledge.

Barroso-Hurtado, Domingo, Wenceslao Arroyo-Machado, and Daniel Torres-Salinas (2021): Formación Profesional Dual: Evolución de red de actores en Twitter, in *Educación XX1*, 9, Article 2.

Berlanga Fernández, Inmaculada, Francisco García-García, and Juan S. Victoria (2013): "Ethos, pathos y logos en Facebook. El usuario de redes: nuevo "rétor" del siglo XXI", in *Comunicar*, 41, pp. 127–135.

Block, William (2008): *The Unimaginable Mathematics of Borges' Library of Babel*. Oxford: Oxford University Press.

Borges, Jorge Luis, and Antonio Carrizo (1997): *Borges el memorioso. Conversaciones*. México: Fondo de Cultura Económica.

Borges, Jorge Luis, and Osvaldo Ferrari (1999): *Reencuentros. Diálogos inéditos*. Buenos Aires: Sudamericana.
Borges, Jorge Luis, and Osvaldo Ferrari (1992): *Diálogos*. Buenos Aires: Seix Barral.
Bourdieu, Pierre (2003): *Campo de poder, campo intelectual*. Buenos Aires: Quadrata.
Brook, James (1995): *Resisting the Virtual Life: The Culture and Politics of Information*. San Francisco: City Lights.
Brown, Andrew (2009): "Retasking Borges: Technology and the Desire for a Borgesian Present", in *Variaciones Borges*, 28, pp.231–239.
Bruni, Nina (1999): "Borges y su actitud ante la comunicación masiva: una propuesta desde la sociología de la cultura", in *Revista de Literaturas Modernas*, 29, pp.73–88.
Cabot, Mateu (2013): "Cultura digital de masas y nueva educación", in *Impulso*, 23, pp.19–28.
Callus, Ivan, and Stefan Herbrechter (2009): *Cy-Borges: Memories of the Posthuman in the Work of Jorge Luis Borges*. Bucknell: Bucknell University.
Calvino, Italo (1993): *Por qué leer a los clásicos*. Barcelona: Tusquets.
Diaz, José Luis (2007): *L'écrivain imaginaire. Scénographies auctoriales à l'époque romantique*. Paris: Champion.
Ferrer, Carolina (2012): "La obra de Jorge Luis Borges, las ciencias y la crítica un análisis de las bases bibliográficas", in Patrizia Botta, Aviva Garribba, María Luisa Cerrón Puga, Debora Vaccari (eds.). *Rumbos del hispanismo en el umbral del Cincuentenario de la AIH*. Roma: Bagatto Libri, pp. 500–507.
Gallego Cuiñas, Ana (2015): "Cortázar legitima a Cortázar. Entrevistas ante el espejo", in Jean-Philippe Barnabé and Kevin Perromat (eds.). *Julio Cortázar: nuevas ediciones, nuevas lecturas*. Paris: Indigo, pp.185–207.
Gallego Cuiñas, Ana (2019): *Las novelas argentinas del siglo 21. Nuevos modos de producción, circulación y recepción*. New York: Peter Lang.
Gallego Cuiñas, Ana (2020): "El gesto Borges en Piglia", in *Letras*, 81, pp. 245–269.
Gallego Cuiñas, Ana (2022a): "Esto no es un autor: futurabilidad y valor literario en *Granta* 2021", in *Insula*, 904, pp. 18–21.
Gallego Cuiñas, Ana (2022b): *Cultura literaria y políticas de mercado. Editoriales, ferias y festivales*. Berlin: De Gruyter.
García Canclini, Néstor (1990): *Culturas híbridas. Estrategias para entrar y salir de la modernidad*. México: Grijalbo.
Helgueta Manso, Javier (2022): "Antiguos y modernos en la comunicación multimedial de la literatura. Citas, videoreseñas, tuiteros y booktubers", in *Insula*, 907–908, pp. 43–46.
Illouz, Eva (1997): *Consuming the romantic Utopia. Love and the Cultural Contradictions of Capitalism*. Berkeley: University of California Press.
Illouz, Eva (2007): *Intimidades congeladas. Las emociones en el capitalismo*. Buenos Aires: Katz editores.
Lapidot, Ema (1990): *Borges y la inteligencia artificial. Análisis al estilo Pierre Menard*. Buenos Aires: Editorial Pliegos.
Lapidot, Ema (1999): "Borges entre la imprenta el hipertexto", in Fernando de Torode Toro and Alfonso de Torode Toro (eds.), *Jorge Luis Borges: pensamiento y saber en el siglo XX*. Madrid/ Frankfurt am Main: Iberoamericana/Vervuert.
Lefere, Robin (2005): *Borges. Entre autorretrato y autobiografía*. Madrid: Gredos.
Lefere, Robin (2015): "La autofiguración en Borges: perspectivas de contextualización", in *Variaciones Borges*, 40, pp. 153–160.
Lipovestdky, Gilles (2020): *Gustar y emocionar. Ensayo sobre la sociedad de la seducción*. Barcelona: Anagrama.

Louis, Annick (2015): “El autor entre dictadura y democracia, fama nacional e internacional. El caso de Jorge Luis Borges (1973-1986)”, in *Revista Letral*, 14, pp. 17–32.
Louis, Annick (2020): “A momentary lapse of history. Borges y la crítica moderna argentina bajo la última dictadura y en la postdictadura (1976–1986”, in *Letras*, n 81, pp. 270–338.
Matamoros, Blas (2008): *Diccionario privado de Jorge Luis Borges*. Molina de Segura: Nausícaä.
Martínez, Guillermo (2007): *Borges y la matemática*. Buenos Aires: Seix Barral.
Martínez-Gamboa, Ricardo (2016): “Big Data en humanidades digitales: De la escritura digital a la “lectura distante”, in *Revista Chilena de Literatura*, 94. <https://revistaliteratura.uchile.cl/index.php/RCL/article/view/44969> (09-12-2021).
Meizoz, Jérôme (2007): *Postures littéraires. Mises en scène modernes de l’auteur*. Genève: Slatkine.
Moed, H. F. (2017). Applied Evaluative Informetrics (1st ed. 2017). Springer International Publishing: Imprint: Springer. https://doi.org/10.1007/978-3-319-60522-7.
Molloy, Sylvia (1999): *Las letras de Borges y otros ensayos*. Rosario: Beatriz Viterbo.
Newhouse, Gatlin Terrol (2019): *Network Babe. Jorge Luis Borges and the Cultura Logic of Digital Networks*. Portland: Reed College.
Olaizola, Andrés (2018): “Bots sociales literarios y autoría. Un aporte desde la retórica digital”, in *Virtuales*, 9, pp. 237–259.
Pauls, Alan (2004): *El factor Borges*. Barcelona: Anagrama.
Peluffo, Ana (2015): “Borges y la cultura de las emociones”, in *Variaciones Borges*, 39: pp. 3–22.
Pérez Fontdevila, Aina, and Meri Torras Francés (2016): *Los papeles del autor/a. Marcos teóricos sobre la autoría literaria*. Madrid: Arco Libros.
Pineda Cachero, Antonio (2002): “Borges y la comunicación de masas”, in *Discurso*, 16/17, pp. 33–67.
Premat, Julio (2009): *Héroes sin atributos. Figuras de autor en la literatura argentina*. Buenos Aires: Fondo de Cultura Económica.
Premat, Julio (2022): *Borges. La reinvención de la literatura*. Buenos Aires: Paidós.
Robinson-García, Nicolas, Wenceslao Arroyo-Machado, and Daniel Torres-Salinas (2019): “Mapping social media attention in Microbiology: identifying main topics and actors”, in *FEMS microbiology letters*, 28, fnz075.
Rojo, Alberto (2015): *Borges y la Física Cuántica*. Buenos Aires: Siglo XXI.
Saítta, Silvia (2018): “Borges mediático”, in *Variaciones Borges*, 46, pp. 3–22.
Santamaría, Alberto (2018): *En los límites de lo posible. Política, cultura y capitalismo afectivo*. Madrid: Akal.
Sapiro, Gisèle (2016): *La sociología de la literatura*. Buenos Aires: Fondo de Cultura Económica.
Sasson-Henry, Perla (2007): *Borges 2.0. From Texts to Virtual Worlds*. New York: Peter Lang.
Sosa Escudero, Walter (2014): *Que es (y que no es) la Estadística*. Buenos Aires: Siglo XXI Editores, Buenos Aires.
Sosa Escudero, Walter (2020): *Borges, big data y yo: Guía nerd (y un poco real) para perderse en el laberinto borgeano*. Buenos Aires: Siglo XXI Editores.
VVAA (2021): *How Collaborating with Artificial Intelligence Could Help Writers of the Future*. Literary Hub. https://lithub.com/how-collaborating-with-artificial-intelligence-could-help-writers-of-the-future/ (08-12-2021).
Virno, Paolo (2017): *La idea de mundo. Intelecto público y uso de la vida*. Buenos Aires: la marca editora.
Zgurovsky, Michael Z., and Yuriy P. Zaychenko (2020): *Big Data: Conceptual Analysis and Applications*. Berlin: Springer, https://doi.org/10.1007/978-3-030-14298-8.