

Distant reading: some musings between world literature and digital humanities

di Giorgio Guzzetta

The publication in 2013 of Franco Moretti's *Distant Reading* marked the definitive *canonization* (a word that the author might not like) of «a growing field of unorthodox literary studies» and was particularly praised by digital humanists as some sort of vindication of their work by an established literary scholar¹. Ross argues that it provides a constructive alternative to the «conflict rhetoric characterizing DH meta-discourse» and self-justification («the divisive conceptual apparatuses that structure DH meta-discourse at present») at the «rhetorical and institutional crossroads» of the discipline we are currently experiencing. The book, moving away from the usual opposition between traditional and digital humanists, models Moretti's career as «a kind of alternate academic history in which the digital humanities developed organically from within traditional methods of scholarship» providing a «compelling [...] anti-conflict narrative that emphasizes the shared goals of the 'traditional' academy and DH»². For sure, distant reading contributed to the hype of Digital Humanities in recent years, especially in literary studies departments³.

This would be an important achievement, of course, but the space opened up at this crossroad is not as peaceful as it seems, in my view. To what extent Moretti's work can vouch for a sound and suitable, uninterrupted and organic evolution is a statement that still needs to be validated. A comparison between the endorsement of distant reading by Digital Humanities scholars and its more controversial reception in

literary studies casts some reasonable doubts about this mediating role. In the initial chapter of *Reading by Numbers*⁴, Bode distinguished three main areas of criticism: reductionism («such approaches reduce the inherent complexity and multiplicity of literature and language to uniform data»), absolutism («the quantitative methods make false claim to authoritative and objective knowledge») and acquiescence to neoliberalism («such studies resonates, in problematic and complicit ways, with contemporary institutional discourses, especially neoliberal or economic rationalist managerial practice»)⁵. In her view, «some aspects of Moretti's research justify some of the criticisms», but the three «are not intrinsic to quantitative methods». In fact, in spite of those criticisms, she maintains that Moretti's «work is well worth the engagement»⁶.

While sketching a new history of the Australian novel, Bode was able to recalibrate not only the literary field (a claim she made in the subtitle of the book), but also the field of quantitative research methodologies. Her analysis of different methodologies that Moretti mixed together in its own personal blend, such as book history, literary darwinism, and, more recently, digital humanities, is very detailed:

His [Moretti's] application of quantitative methods extends well beyond any other work in book history, especially in his use of what might be called textual as well as material or historical data. While the first chapter of *Graphs* is (as Moretti acknowledges) es-

entially an exercise in quantitative book history – drawing on historical data to explore trends in book publication, authorship and genre across a range of national fields – the other two chapters are based on datasets created from elements within particular literary texts, such as character, setting, plot and device. [...] In more recent work, Moretti employs quantitative methods to analyse language patterns in much larger groups of texts. These other studies have their own antecedents: Literary Darwinists used textual data, and Moretti's analyses of language patterns draw on methods developed in linguistics and digital humanities (or humanities computing). But no one else incorporates these range of approaches, or combine them in ways as original and provocative as Moretti⁷.

The scope of her survey is limited, for obvious reasons, to Anglo-American culture. The main area of expertise of both Moretti and Bode is English literature, the latter from a peripheral position (Australian novel), the former with forays in European and World literature (the significance of which I will discuss later on). So, in a sense her bias is understandable. Still, the lack of reference to Italian culture is, in my view, problematic, and in my paper I will try to fill the gap, especially in the first section where I analyze the origin of the idea of distant reading.

My main interest is not to discuss the history of Moretti's ideas and methods as such, but how they are connected with digital humanities, how they fit in it. Moretti's writing is «wide-ranging and, above all, *sparky*», and I have no doubts that he is «a writer who has interesting and important things to say about the practice of literary history in general and the development of the novel in particular»⁸. Therefore, my discussion is focused only on a very specific aspect of his scholar work, that goes under the umbrella-term of distant reading, which, as I have said before, represents the main common ground with digital humanities. My paper could be considered, I guess, a close reading of *Distant Reading* focused on three main topics: the origin of the methodology; the problem of world literature; the issue of data.

The origin of Distant Reading

Given the hype that surrounded the label he created, it is not surprising that Moretti decided to write a book to reconstruct «the intellectual trajectory of his philosophy

of 'distant reading' [... following] two decades of critical explorations» (I quote from the publisher's website). It is not a monograph, but a collection of essays, originally published from 1991 to 2011, without even an introduction to connect them together. This choice is a bit odd, but not entirely new in Italian culture. In 1957, Elio Vittorini, at that time a leading figure in Italian literature and culture, especially left-wing, collected his writings from 1929 to 1957 for *Diario in pubblico*, a controversial book in which each piece was accompanied by comments written in a different font that contextualised and updated the content of the original texts. The way in which he reworked the material stirred some criticism, rekindled by the posthumous enlarged edition of 1970 (Vittorini died in 1966), where a new section was added covering the period between 1957 to 1966. He was attacked mainly on the grounds that he distorted the originals to conceal his alleged sympathies with Fascism, an issue on which tons of ink have been spent in the past, but that does not concern us here. What is interesting, on the other hand, is the possibility that Moretti used a similar structure when collecting his own writings.

Moretti's academic career started in the 1970s (his first edited book was published in 1975, his first monograph a year later), when he was trying to «create a happy union between materialist and formalist approaches to literature»⁹. At that time the ghost of Vittorini was still present in Italian culture, and would have for sure haunted a young, left-wing literary scholar like Moretti. This can explain the similarities between *Diario in pubblico* and *Distant Reading*. Like Vittorini, Moretti used italics to introduce each essay, commenting and giving information on the origin and significance of it, on the writing circumstances and the issues addressed, and sometimes even on the difference between the Italian and the English versions. A possible connection with *Diario in Pubblico* is interesting because of the way in which Vittorini dis-assembled and re-assembled the original texts, abstracting and virtualizing them with a 'distant reading' method similar to the one proposed by Moretti later on. In doing this, Vittorini was actually saying something about the way in which cultural and linguistic memory works (something that one could apply also to his work as a translator, but that is another story). In short, what he is implying is that texts and books are not (forever) cast in stone, but are negotiated, textured linguistic systems that combine together subatomic particles that keep operating, passing

through them, and connecting cultural and linguistic life before and after the event of a book publication. In other words, we could say that he is sidelining the text as a cultural object, privileging or at least bringing into view elements from a different scale. Similarly, as we know, Moretti will look at components and features above and below the level of text.

Apart from this innovative approach to cultural and textual memory, that was not very well received at that time, the one thing that the 1970s generation for sure inherited from Vittorini was his attention for the problem of the two cultures and his attempts to bridge the gap between science and humanities. Inspired by C.P. Snow's 1959 pamphlet, in the early 1960s Calvino and Vittorini debated the question in the last issue of *Il Menabò*, the literary journal they co-directed (that stopped when Vittorini died in 1966, except for this 1967 issue that was a tribute to him). Furthermore, in a collection of notes written by Vittorini, posthumously published as *Le due tensioni*, «one of the dominant themes [...] was the need for contemporary literature to catch up with modern scientific knowledge»¹⁰. The notes covered a wide range of ideas (sociology, linguistic, anthropology, even cybernetics), but the common element was «la passione dominante della letteratura», omnipresent, «anche dove se ne pronostica o se ne sconta già prossima la fine»¹¹.

The centrality of literature as the main symbolic system that allows a society's expression of itself, ultimately a humanistic concept, was a basic assumption for both Vittorini and Moretti. The latter, in the first essay of the collection, *Modern European Literature: A Geographical Sketch* (1991), abandoned Curtius' idea of a homogeneous European literature, but still takes for granted a strong, organic connection between literature and cultural identity, at the national level:

The notion of European literature, singular, was replaced by that of an archipelago of distinct yet close national cultures, where styles and stories moved quickly and frequently, undergoing all sorts of metamorphoses¹².

It is this organic nature of literature that allowed him to follow Vittorini's invitation to bridge the gap between science and humanities. Specifically, he is able to sketch the historical evolution of European literature and re-interpret the notion of canon using concept

from evolutionary biology, such as *variation* and *selection*. The literary canon is the result of a process of selection amongst the variants (i.e., the many different texts) produced by individual agents (i.e., the writers). Caesar summarizes Moretti's position quite clearly, highlighting the question of writer's agency as a particular problematic issue:

In the dialectic between biological variation and selection, the variation, or innovation, for Lamarck is always functional to selection [...] Lamarck's organisms are rational agents and, according to Geoffrey Winthrop-Young, 'it is this combination of conscious adaptation and rational evolution [...] that makes theorists account for cultural evolution in Lamarckian rather than Darwinian terms'. Because in Darwin, the link between variation and selection is not necessary: selection itself obeys the laws of necessity, but not all variations are selected. While biological evolution in Darwinian terms proceeds by divergence, cultural evolution proceeds, it seems, by convergence. *This creates a problem for cultural historians because it appears to deny agency. But Moretti succeeds in breaking this problem down quite cleverly.* On the one hand, the discontinuity between variation and selection would mean that more literature is always produced than will ever be 'selected', and also that no effort on the part of an individual author can ensure that his or her work will be preserved for posterity; by the same token, however, it allows Moretti to preserve an individual agency that gives birth to texts, and then to declare it irrelevant when it comes to analyzing the text's social function¹³.

The agency of the writer is problematic also in Vittorini, as his idea of *natural engagement* shows. It might be too daring to say that «Vittorini barely ever, I hazard, expresses the relationship between himself and society as a relationship between a writer-individual and a reader-individual» and that «the author-reader relationship is [...] a subset of a much greater, even universal, set of relationships»¹⁴, but clearly there is in Vittorini a strong identification of the writer with a society, or a community, similar to the national literary identification envisaged by Moretti. The ambiguities of Vittorini's notion of literature and culture are best expressed by Burns when she said that he

refers frequently to a central concept: 'lo spirito di ricerca'. This 'spirit' is the essence and energy-

source of culture, and the reason why it is so important to human development. The very phrase encapsulates two directions which persist in Vittorini's critical thinking: the 'spirito' expresses a metaphysical quality, whilst the 'ricerca' indicates a more practical and earthly activity¹⁵.

In Moretti too there is a dualism that he never managed to reconcile, between plot and language (or story and style: «narrative theory has always known that a story is composed of the distinct layers of story and style», a statement far too trenchant, in my view) or between variation and selection, that will lead to two different lines of research¹⁶.

This dualism is epitomized in Vittorini's «due tensioni», described as «tensione di affettività/espressività come chioma d'albero di un tronco di tensione razionale»¹⁷. To visualize this abstract model of literature, he sketched an actual tree on the left side of the page, a practice that Moretti will inherit and further develop.

Moretti's visualizations indeed open up «exciting vistas»¹⁸, anticipating more recent interdisciplinary collaborations between science and humanities that revised the idea of the canon:

Ciascuno degli studiosi che ha risposto al mio invito ha contribuito a definire i contorni e sviluppare le varie sfumature del concetto e della pratica di "canone". [...] La perdita di centralità del testo da tempo non è più un tabù nelle discipline umanistiche [...] La sempre più inattuale (e nociva) separazione tra scienze della cultura e scienze della natura riporta alla luce un senso non-canonico degli artefatti cognitivi: la *varianza* svela il ruolo del margine, del subalterno, dell'escluso¹⁹.

Should we call this *distant reading*? I would say no. The text has lost centrality in both cases, but for different reasons. In the case of Fiori, we can say that text and literature have both lost centrality, becoming part of the (new) media system:

Mettendo al centro della storia i mezzi di comunicazione, l'asse culturale del novecento subisce uno spostamento – per esempio riportando le forze che guidano la creazione dei concetti di autore e opera (la letteratura stessa) dentro il *sistema dei media* [...] e rappresenta il primo attacco dall'interno all'assetto di quelle discipline umanistiche tradizionali – storia, filosofia e letteratura – che avevano

dominato la cultura occidentale degli ultimi cinque secoli²⁰.

Moretti, on the other hand, is not interested in the transformations and the mutations of the text triggered by digital encoding and digital convergence. In the *Introduzione* to *La letteratura vista da lontano*, the Italian version of *Graphs, Maps, Trees*, he clearly states that «qui si parla di letteratura: l'oggetto rimane più o meno quello di sempre, a differenza della recente virata del *new historicism*, e poi dei *cultural studies*, verso altri ambiti di discorso»²¹. He takes a clear stand here against new media and, even more so, hybrid forms such as digital games²².

The shift is not from text to media, but from text to literary system. The difference is more clear in the second essay of his collection, *Conjectures on World Literature*, where he spoke for the first time of «distant reading». Even Moretti had some doubts about the label, as he stated in introducing the piece:

*Conjectures provoked heated reactions [...] aimed at the idea of 'distant reading'. That fatal formula had been a late addition to the paper, where it was initially specified, in an allusion to the basic procedure of quantitative history, by the words, 'serial reading'. Then, somehow, 'serial' disappeared, and 'distant' remained. Partly, it was meant as a joke; a moment of relief in a rather relentless argument. But no one seems to have taken it as a joke, and they were probably right*²³.

Serial reading foregrounds purely methodological aspects of a quantitative approach, whereas *distant* positions the reader *above* and *against* the text, and establishes a cognitive, if not moral, distance. Distant reading is indeed a change of scale, but also a «condition of knowledge [that] allows you to focus on units that are much smaller or much larger than the text: devices, themes, tropes- or genre and systems»²⁴. The scale is that of World Literature, which requires the use of «different categories» and «a new critical method» in order to make sense of the «great unread» (a concept originating from Cohen 1999). How Moretti interpret it is what I will discuss in the next section.

World Literature

The idea of *World Literature*, an offspring of the *Respublica Literaria* of the Renaissance, gained currency after Goethe used it in several occasions, particularly in 1836 Eckermann's *Conversations*. His interpretation was similar to Moretti's idea of European literature described above, that is to say an international network of writers from France, England and Germany that mutually exchanged creative texts and opinions in journals and in letters. Not surprisingly, in Goethe the shift between European and Universal (*Weltliteratur*) was frequent, the former being a step in the progression that will lead to the latter:

For Goethe, world literature is, to start with, European literature. It is in process of realising itself in Europe. A European literature [...] is the first stage of a world literature which from these beginnings will spread in ever-widening circles to a system which in the end will embrace the world. World literature is a living, growing organism, which can develop from the germ of European literature [...]²⁵.

In the last decades of the twentieth century, *Weltliteratur* came back with a vengeance, and Moretti's version, clearly derived from Goethe's euro-centric approach, is part of this resurrection. (Sinopoli 2010) Discussing *world literature* was a way to bring literature back to the fore, after a period in which discussing and defining it was increasingly difficult. From the nineteenth century onwards, says Reiss, literature «has suffered from a kind of repetitive inertia, as it has also been changing its meaning and cultural role»²⁶. Paraphrasing what Kirsch wrote about digital humanities, we can say that a focus on world literature was an attempt to bring back to life the «long-suffering noun» using a «high-powered» word²⁷.

Cultural globalization played a part, as well as the emergence of postcolonial and cultural studies:

Nel corso del novecento [...] la questione della mondialità della letteratura si articolerà ulteriormente in un vero e proprio confronto tra la progressiva, e temuta, globalizzazione delle culture e il sempre più limitato raggio d'azione degli studi letterari, incapaci di star dietro a un sistema letterario in continua espansione e al contempo in forte debito nei confronti dei processi storici e delle forze economico-

politiche internazionali, rispetto ai quali esso appare sempre meno autonomo e autosufficiente²⁸.

Among the historical processes involved, the emergence of the digital, with the ensuing cultural convergence into new media in the 1980s and the creation of a *worldwide web* in the 1990s, is obviously key. We cannot establish a direct connection between the two events, but the fact that the renewed interest in world literature happened in the same years in which Tim Berners-Lee *invented* the web, creating a global *cultural network* on top of the *technological* one, is indeed significant. When Edward Said discussed Auerbach's ideas on world literature, he spoke about a problematic connection between new media, cultural studies and globalization. Said described the current cultural situation as an «age of media-produced attitudes», in which «the ideological insistence of a culture drawing attention to itself as superior», allowing forms of resistance and opposition in anticolonial and postcolonial movements, «has given way to a culture whose canons and standards are invisible to the degree that they are 'natural', 'objective' and 'real'»²⁹. In other words, what was still visible has become invisible, but it remains in both cases problematic.

The creation of a global digital archive will clearly affect the materiality, if not the notion itself, of world literature. It cannot be separated from the emergence of the web as a global communication system and from issues prompted by the encoding and digitization of texts. Moreover, the problem of the creation and manipulation of data was the main focus of humanists and literary scholars that used computers long before big data. An early document discussing digital libraries (without calling them that, though) is interesting in this respect.

In 1965, a team led by J. C. R. Licklider, a key-figure in shaping the digital age as we know it, published a report on «the applicability of some of the newer techniques for handling information to what goes at present by the name of library work»³⁰. The title, *Libraries of the Future*, was to be taken literally, insofar as their aim was not the contemporary situation but what they imagined the libraries will or should be in the year 2000. There is no space here to analyze the full content of the report, in all its ramifications; I will only emphasise the scope of the study, defined at the outset. They started by saying that the book was not the main object of

interest, and they were focusing on something very similar to the things Moretti focused on:

The 'libraries' of the phrase, 'libraries of the future,' may not be very much like present-day libraries, and the term 'library', rooted in 'book', is not truly appropriate to the kind of system on which the study focused. We delimited the scope of the study, almost at the outset, to functions, classes of information, and domains of knowledge in which the items of basic interest are not the print or paper, and not the words and sentences themselves- but the facts, concepts, principles, and ideas that lie behind the visible and tangible aspects of documents³¹.

Moretti's ideas are perfectly consistent with this definition. We can consider his work as making full use of the libraries of the new millennium. What immediately follows in the report, as a consequences of what was said in the quotation above, is more problematic:

The criterion question for the delimitation was: 'Can it be rephrased without significant loss?' Thus we delimited the scope to include only 'transformable information'. Works of art are clearly beyond that scope, for they suffer even from reproduction. Works of literature are beyond it also, though not as far. Within the scope lie secondary parts of art and literature, most of history, medicine, and law, and almost all of science, technology, and the records of business and government³².

Oddly enough, it seems clear that in their vision primary literary works were not part of the libraries of the future, or digital libraries as we call them now. Only secondary work (literary history and criticism, for instance) played a part. One might wonder if in fact distant reading was not doing something similar, that is to say ignoring primary works and using only 'secondary' data, being therefore entirely consistent with their scope. In this respect, however, I think we can say that their bold prophecy was wrong. Between then and now something happened that they could not have predicted.

Literary texts found a place in digital libraries thanks to the labor of humanities computing scholars. The roots of today's global digital archive (or of the near future's ones, if you prefer) are to be found in the work of those pioneers. Consequently, our idea of world literature is influenced by their work. Even constructed by

it. To understand where we are we need to look, however briefly, at this history. In the next section I want to discuss Moretti's use of data and how it evolved in time, what is the role of humanities computing and digital humanities in this evolution, and to what extent the idea of distant reading is actually the result of it, or something coming out from a more traditional approach. My point is that it is not the existence of the global digital archive that makes data important in literary studies, as Moretti seems to imply sometimes, but the other way around. It is the discussion on literary data that made digital archives possible. Data came first, then digital archives, or at least they came into life together.

The question of data

When he called his method *distant reading*, Moretti did not do something entirely new. In fact, he became part of an earlier tradition. The emphasis on *close* reading already implied the idea of a *distant* reading against which new critics reacted, even though they did not use the term explicitly. Ironically enough, two books published in the same years – one inaugurating close reading against the habit to «isolate certain aspects of poetry for special investigation» an impulse that «in its crudest manifestation [...] leads to statistical surveys of one kind or another»³³, the other one promoting the use of statistical techniques³⁴ – were both advocating the need to approach *wholeness*. Brooks and Warren insisted repeatedly on the «whole context» and the «whole picture», on the text «taken as a whole» as an «organic system of relationships», while Yule wanted to study the vocabulary of a writer globally, using statistical methods:

What I felt I wanted in the first place, prior to any detail, was some summary, some picture of the vocabulary as a whole. Surely the colour and flavour of a text [...] are determined not by the exceptional words [...] but in the main by the common words used by the author [...] The sort of picture I wanted could only be given by a list of the words used classified by the number of times they were used [...]³⁵.

The use of statistics to study literary texts is much older, and we can say that many of the ideas that are currently «the staples of the digital humanist's diet»³⁶

were «presented and lost several times between the mid-1700s and the late 1900s», often because of the lack of «a technological invention [...] necessary for it to be fully developed»³⁷. A comprehensive history of stylometry (a term original invented by Lutoslawski, a Polish philosopher, in the 1890s) is yet to be written³⁸, but we know enough to say that Moretti's statistical methods are rooted in a long tradition of stylometry, a tradition that was clearly boosted when analogue and digital computers arrived.

In most cases these early *distant readers* were statisticians with a passion for literature that did some stylometric exercises in their free time. L. A. Sherman, a professor that regularly used statistics in teaching and researching literature, was an exception, and was famously (or infamously) satirized by the young Willa Cather³⁹. Sherman was probably the first to create a literary lab (in his book, *Analytics of Literature*, he called it «literary laboratory») where he conducted experiments, using quantitative methods, splitting the text into elements, counting words etc. He lamented that texts were treated as:

ultimate phenomena, without much recognition that there may be [...] less complicated forms of the same kind, as well as constant ultimate elements whose presence in new proportions and new combinations make up all differences observed. [...] In other words, Chaucer and Shakespeare are considered simply as Chaucer and Shakespeare, with no reference to the fact that there must be in both common constituents and factors which, in different frequency and degrees of potency, make up the very diverse effects of their respective poetry. The same must be true also of our prosaists⁴⁰.

A discussion of similarities and differences between Sherman's and Moretti's lab is not possible here, and perhaps not even interesting. However, even if it was as «ignorant» and «ruthless» as Cather said, his idea travelled and was refined by a long history of discussions about literary data, in which humanities computing and digital humanities played a significant role.

Counting words was becoming more common, and, thanks to Hollerith and his punched card machines, mechanical devices were more and more used. In 1935, long before Buser, IBM published a book describing *Practical Applications of the Punched Card Method in College and Universities*, with a chapter on literary stud-

ies⁴¹. Later on, in 1964, all these converged into an international conference on *Literary Data Processing* in New York. There is no need here to discuss this in detail. It is sufficient to say that these scholars were slowly building up the discipline of humanities computing.

Although Moretti never mentioned it explicitly this first wave of digital literary studies (not surprisingly: this was indeed a very small group of academics), nonetheless, in *The Slaughterhouse*, his first attempt to refine his own methods for *not reading* a book, he stated that the new scenario «requires other skills: sampling; statistics; work with series, titles, concordances, incipits»⁴². At least three of those (sampling, statistics, concordances) were already used by humanities computing practitioners since the 1950s and 1960s, and even earlier, as we have seen. Their work changed and improved their understanding of literary data.

What kind of data Moretti was using in his experiments? In *Conjectures* distant reading meant that we need *second hand* data, relying on «a patchwork of other people research, without a single direct textual reading»⁴³. There is no *digital* reading of any kind, because the data he worked with at first were not derived from computational text analysis. He draws conclusion out of twenty «independent critical studies» on the dissemination of the novel in areas as diverse as West Africa, Arab countries, Latina America, Southern and Eastern Europe, Turkey, Philippines, China, Japan etc. In *Graphs* he builds upon «work originally done by McBurney, Beasley, Raven, Garside and Block for Britain; Angus, Mylne and Frautschi for France; Zwicker for Japan; Petersen for Denmark; Ragone for Italy; Marti-Lopez and Santana for Spain; Joshi for India; and Griswold for Nigeria». Neither the reason why these particular ones were selected, nor, more importantly, how he abstracted data from their monographs is discussed. Obviously, in *traditional* literary criticism, secondary literature is not usually discussed nor justified, but in digital humanities, and humanities computing as well, the creation and manipulation of data became key. The absence of such a discussion confirms, in my opinion, that he was not aware of the work made in humanities computing.

Pragmatically enough, in *Slaughterhouse* he went back to reading the 108 crime stories he discussed. It was not close reading, however, but some kind of browsing for clues. As Moretti says, «it felt very different from the reading I used to know,» and was prob-

ably not as enjoyable. The idea of doing it again did not seem very attractive, and we can assume that the repeated announcement of the arrival of digital archives was a desire as much as a statement of fact. Having to deal with a new and enlarged archive, it is understandable that he initially lost his bearings and had to proceed tentatively: «quantitative data [...] produced such a large new body of evidence, for which I was so completely unprepared, that the need for a theoretical framework was for a few years forgotten in the heady mood of permanent exploration»⁴⁴.

The results of these explorations were presented in *Graphs, Maps, Trees*, whose abstract models and visualization provoke the most aggressive criticisms, which I already mentioned before. Here I just want to add that the «false claim to authoritative and objective knowledge» is related, in my view, not only to the use of quantitative data, but also to the combination of the world system theory of core and periphery with the evolutionary framework. In Moretti's interpretation, I dare say, the quantitative results that posit Europe at the core of the system are backed up by his perception how European literature's «capacity to generate new forms, which seemed so historically unique»⁴⁵. In other words, European literature generates more variation than any other culture, and it seems natural that it became the core. Taken separately, the three vistas are fascinating, but their combination makes European superiority almost inescapable. Consider, for instance, the «law of literary evolution» of the modern novel. This law states that:

in cultures that belong to the periphery of the literary system (which means: almost all cultures, inside and outside Europe), the modern novel first arises not as an autonomous development but as a compromise between a western formal influence (usually French or English) and local materials⁴⁶.

Controversies generated by Moretti's «grand narrative of intellectual diffusion with Europe as the core»⁴⁷ are well-known and well-documented⁴⁸, and there is no reason to insist on them here. The notion of periphery, and the accompanying euro-centric attitude, is already problematic enough; but what I find even more troubling is the fact that the compromise is not between a western formal influence and local literary forms (which might be considered a study on cross-fertilization), but between *western form* and *local materials*. This is prob-

ably the most problematic aspect of Moretti's distant reading, and justifies Spivak accusation of a «scopic ambition of mapping the world's literatures»⁴⁹. Granted, instead of a law this is an experimental hypothesis that need to be tested, but this, I believe, makes things even more problematic for digital humanities. What I am trying to say is that separate hypothesis making and data collection is not easy, and the risk is that the bias of the hypotheses will be reflected in the data.

In his response to criticism, Moretti started mentioning digital archives. In *Distant Reading* the digital is barely discussed, and almost always in the actualizing comments in italics (the ones written in 2015). In the actual essays, digital databases and archives (not digital humanities) are only mentioned in *The Novel* (2008/2013), and in *Style, Inc.* (2009/2013). In both cases Moretti is merely predicting a time when «we will have a digital archive with the full texts of (almost) all novels ever published» in which «a quantitative stylistics [...] may find some answers»⁵⁰. Even in 2013 the creation of this archive is still a problem, as the use of the adverb *slowly* in the two italicized occurrence of digital, the first one regarding the creation of a global digital archive «(slowly) coming into being», and the second one concerning the fact that «a desire for a general theory of the new literary archive is slowly emerging in the world of digital humanities»⁵¹.

In itself, the existence of a global digital archive was not without consequences, transforming the actual object of study: «It's another fascinating aspect of the 'size is not just size' idea: world literature, not just as a unique object of study in and of itself, but as a sort of 'natural laboratory' for all sorts of theoretical experiments.» For a literary scholar, the shift from library to archive is problematic, says Moretti:

[...] one cannot study a large archive in the same way one studies a text: texts are designed to 'speak' to us [...] but archives are not messages that were meant to address us, and so they say absolutely nothing until one asks the right question. And the trouble is, we literary scholars are not good at that: we are trained to listen, not to ask questions, and asking questions is the opposite of listening: it turns criticism on its head, and transforms it into an experiment of sorts. (Moretti 2008/2013)

Indeed, there is an evolution in the way Moretti collects and uses data, in the kind of experiments he

is able to do. And I cannot deny that he anticipated some of the interesting things Digital Humanist cherish. See, for example, how his rationale for quantitative work data seems to anticipate the current emphasis on open data and collaboration:

I mention these names right away because quantitative work is truly *cooperation*: not only in the pragmatic sense that it takes forever to gather the data, but because such data are ideally independent from any individual researcher, and can thus be shared by others, and combined in more than one way⁵².

Also the use of maps, diagrams, trees, waves and other kinds of visualizations is very similar to what is happening for instance with “surface” text analysis tools such as Voyant, with the significant difference that in this case data are constructed using digital textual analysis⁵³. The playfulness that characterizes some of Moretti’s experiments makes even more sense when applied to digital tools. Rockwell, for example, discussing what text analysis *really* is, speaks of “playful experimentation”:

Disciplined play privileges experimentation and modelling over hypothesis testing or concordance publishing. Playful experimentation is a pragmatic approach of trying something, seeing if you obtain interesting results, and if you do, then trying to theorize why those results are interesting rather than starting from articulated principles⁵⁴.

Moretti’s approach is also pragmatic: “abstraction is not an end in itself, but a way to widen the domain of the literary historian, and enrich his internal problematic”⁵⁵. But at the same time, it appears that he is far from being immune from hypothesis testing, something that Rockwell did not encourage.

I am stopping here, following the span of the book, immediately before the work of the Stanford Literary Lab, in which, to be fair, more attention was given to some of the issues I mentioned. I will conclude by saying that Moretti’s work can indeed connect traditional literary studies with digital humanities in the future, but without a critical analysis of some of his views, first of all distant reading, there is a risk that new gaps will open between digital literary studies and other fields of research, such as postcolonial and cultural studies.

Bibliography

Aravamudan S., *Enlightenment Orientalism: Resisting the Rise of the Novel*, Chicago, The University of Chicago Press 2012.

Baehne G.W., *Practical Applications of the Punched Card Method in Colleges and Universities*, New York, Columbia University Press 1935.

Bessinger J.B., Parrish S.M., Arader H.F. (eds.), *Literary Data Processing Conference Proceedings*, White Plains, N.Y., IBM 1965.

Bode K., *Reading by Numbers: Recalibrating the Literary Field*, London-New York, Anthem Press 2012.

Bonsaver G., *Elio Vittorini: The Writer and the Written*. Leeds, Northern Universities Press 2000.

Brooks C., Warren R.P., *Understanding Poetry*, New York, H. Holt and Company 1938.

Burns J., *Fragments of Impegno: Interpretations of Commitment in Contemporary Italian Narrative, 1980-2000*. Leeds, Northern Universities Press 2001.

Caesar M., *Franco Moretti and the World Literature Debate*, in «Italian Studies» 62.1 (2007), pp. 125-35.

Cohen M., *The Sentimental Education of the Novel*, Princeton, Princeton University Press 1999.

Fiorimonte D. (ed.), *Canoni liquidi: variazione culturale e stabilità testuale dalla Bibbia a internet*, Napoli, Scriptaweb 2011.

Fiorimonte D., *Scrittura e filologia nell’era digitale*, Torino, Bolletti Boringhieri 2003.

Goodwin J., Holbo J., *Reading Graphs, Maps & Trees: Responses to Franco Moretti*, Anderson (SC), Parlor Press 2011.

Jockers M.L., *Macroanalysis: Digital Methods and Literary History*, Chicago, University of Illinois Press 2013.

Kirsch A., *Technology Is Taking Over English Departments*, in «New Republic» 2 May 2014, Web. 25 Jan. 2016.

Licklider J.C.R., *Libraries of the Future*, Cambridge, Mass., MIT Press 1969.

McCarty W., *Humanities Computing*, Basingstoke [England], New York, Palgrave Macmillan 2005.

Moretti F., *Distant Reading*, London, Verso 2013.

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

Moretti F., *La letteratura vista da lontano*, Torino, Einaudi 2005.

Reiss T.J., *The Meaning of Literature*, Ithaca, N.Y., Cornell University Press 1992.

Rockwell G., *What Is Text Analysis, Really?*, in «Literary and Linguistic Computing» 18.2 (2003), pp. 209–19. llc.oxfordjournals.org. Web.

Rockwell G., Sinclair S., *Towards an Archaeology of Text Analysis Tools*, Lausanne, N.p., 2014. Web. 19 Oct. 2015.

Ross S., *In Praise of Overstating the Case: A Review of Franco Moretti, Distant Reading (London: Verso, 2013)*, in «DHQ: Digital Humanities Quarterly» 8.1 (2014). Web.

Said E.W., *The World, the Text, and the Critic*, Cambridge, Harvard University Press 1983.

Schulz K., *What Is Distant Reading?*, in The Mechanic Muse - «New York Times». N.p., 24 June 2011. Web. 25 Jan. 2016.

Sherman L.A., *Analytics of Literature*, Boston, Ginn 1893.

Sinopoli F., «Dall’universalismo letterario alle forme attuali della mondialità letteraria», in *La letteratura del mondo nel XXI secolo*, Milan, Italy, B. Mondadori 2010.

Slote B., «Introduction», in *The Kingdom of Art: Willa Cather’s*

First Principles and Critical Statements, 1893-1896, Lincoln, University of Nebraska Press 1967.

Snow C.P., *The Two Cultures and the Scientific Revolution*, Cambridge, Cambridge University Press 1959.

Strich F., *Goethe and World Literature*, London, Routledge & K. Paul 1949.

Stubbs M., «Technology and Phraseology. With Notes on the History of Corpus Linguistics», in *Exploring the Lexis-Grammar Interface*, Ed. Ute Römer and Rainer Schulze, Amsterdam et al., John Benjamins 2009, pp. 15-31.

Vittorini E., *Diario in pubblico*, Milano, Bompiani 1957.

Vittorini E., *Le due tensioni*, Milano, Il sagggiatore 1967.

Yule G.U., *The Statistical Study of Literary Vocabulary*, Cambridge, Cambridge University Press 1944.

Notes

¹ S. Ross, *In Praise of Overstating the Case: A Review of Franco Moretti, Distant Reading* (London, Verso, 2013), in «DHQ: Digital Humanities Quarterly» 8.1 (2014): <http://www.digitalhumanities.org/dhq/vol/8/1/000171/000171.html>.

² Ibidem.

³ A simple google search on "distant reading and digital humanities" will be proof enough of this close connection. See, for instance, Schulz, *What is distant reading?* on the «New York Times»: <http://www.nytimes.com/2011/06/26/books/review/the-mechanic-muse-what-is-distant-reading.html>

⁴ K. Bode, *Reading by Numbers: Recalibrating the Literary Field*, London-New York,, Anthem Press 2012, pp. 7-25.

⁵ Ibidem, pp. 9-10.

⁶ Ibidem, p. 7.

⁷ Ibidem, pp. 9-10.

⁸ M. Caesar, *Franco Moretti and the World Literature Debate*, in «Italian Studies» 62.1 (2007), p.126.

⁹ Ibidem, p.125.

¹⁰ G. Bonsaver, *Elio Vittorini: The Writer and the Written*, Leeds, Northern Universities Press 2000, p. 239.

¹¹ «The dominant passion for literature, omnipresent, even when he foretells his imminent death.» (translations are mine, unless otherwise indicated): E. Vittorini, *Le due tensioni*, Milano, Il sagggiatore 1967, p. xi.

¹² F. Moretti, *Distant Reading*, London, Verso 2013, p. 1.

¹³ M. Caesar, *Franco Moretti*, cit., p. 127; italics are mine.

¹⁴ J. Burns, *Fragments of Impegno: Interpretations of Commitment in Contemporary Italian Narrative*, 1980-2000, Leeds, Northern Universities Press 2001, p. 16.

¹⁵ Ibidem, p. 18.

¹⁶ F. Moretti, *Distant Reading*, cit., p. 92.

¹⁷ «tension of affectivity/expressivity as foliage of a tree trunk of rational tension». E. Vittorini, *Le due tensioni*, cit., p. 5.

¹⁸ M. Caesar, *Franco Moretti*, cit., p. 126.

¹⁹ «Each of the scholars who responded to my invitation helped to define boundaries and nuances of the concept and practice of canon [...] The loss of centrality of the text is not anymore a taboo in the humanities [...] The increasingly outdated (and harmful) separation between the study of culture and the

study of nature brings to light a non-canonical significance of cognitive artefacts: variation unveils the role of the marginal, the subaltern, the excluded»: D. Fiormonte, (ed.), *Canoni liquidi: variazione culturale e stabilità testuale dalla Bibbia a internet*, Napoli, Scriptaweb 2011 p. 6.

²⁰ «By placing media at the center of history, the cultural axis of the twentieth century is shifted - for example locating the forces behind the creation of the concepts of author and text (literature itself) inside the media system, [... which is] the first time traditional humanistic disciplines that had dominated Western culture for the last ve centuries such as history, philosophy and literature- were attacked from within.»: D. Fiormonte, *Scrittura e filologia nell'era digitale*, Torino, Bollati Boringhieri 2003, p. 31.

²¹ F. Moretti, *La letteratura vista da lontano*, Torino, Einaudi 2005, pp. 4-5. I am using here the Italian version because it differs slightly from the American one and makes my point more clear. Both version were published in 2005, therefore the differences are cultural and linguistic, not the result of a revision of the content. A more detailed comparison between the two books, and in general of the different versions of all Moretti's writings, could be interesting, especially if connected to his views on world literature (which I will discuss later on). But there is no space here. Instead, I will simply quote the English version (with the differences in italics, and some quick comments in brackets): «To begin with, this is an essay on literary history [no mention of history in the Italian book]: literature, the old territory (more or less), unlike the drift towards other discourses so typical of recent years.» [the "other discourses" being cultural studies and new historicism] F. Moretti, *Graphs, Maps, Trees. Abstract Models for a Literary History*, London, Verso 2005, p. 1.

²² F. Moretti, *La letteratura*, cit., p. 5.

²³ F. Moretti, *Distant Reading*, cit., p. 44.

²⁴ Ibidem, pp. 48-9.

²⁵ F. Strich, *Goethe and World Literature*, London, Routledge & K. Paul 1949, p. 16.

²⁶ T.J. Reiss, *The Meaning of Literature*, Ithaca, N.Y, Cornell University Press 1992, p. 28.

²⁷ A. Kirsch, *Technology Is Taking Over English Departments*, in «New Republic» 2 May 2014, Web. 25 Jan. 2016: <https://newrepublic.com/article/117428/limits-digital-humanities-adam-kirsch>.

²⁸ «During the twentieth century [...] the question of the worldness of literature will unfold a friction between the progressive, and feared, globalization of cultures and the increasingly limited range of literary studies, unable to keep up with a literary system constantly expanding and at the same time more and more dependent upon historical processes and economic powers»: F. Sinopoli, «Dall'universalismo letterario alle forme attuali della mondialità letteraria», in *La letteratura del mondo nel XXI secolo*, Milan, Italy, B. Mondadori, 2010, p. 58.

²⁹ E.W. Said, *The World, the Text, and the Critic*, Cambridge, Harvard University Press 1983, p. 9.

³⁰ J.C.R. Licklider, *Libraries of the Future*, Cambridge, Mass., MIT Press 1969, p. v.

³¹ Ibidem, p. 2.

- ³² Ibidem, p. 2.
- ³³ C. Brooks & R.P. Warren, *Understanding Poetry*, New York, H. Holt and Company 1938, p. vii.
- ³⁴ G.U. Yule, *The Statistical Study of Literary Vocabulary*, Cambridge, Cambridge University Press 1944.
- ³⁵ Ibidem, p. 2.
- ³⁶ M.L. Jockers, *Macroanalysis: Digital Methods and Literary History*, Chicago, University of Illinois Press 2013, p. 4.
- ³⁷ M. Stubbs, «Technology and Phraseology. With Notes on the History of Corpus Linguistics», in *Exploring the Lexis- Grammar Interface*, Ed. Ute Römer and Rainer Schulze, Amsterdam et al., John Benjamins 2009, p. 15.
- ³⁸ Actually, the history of stylometry and of early humanities computing is part of a book project I am working on at the moment, and I am anticipating some of it here.
- ³⁹ B. Slotte, «Preface», in *The Kingdom of Art: Willa Cather's First Principles and Critical Statements, 1893-1896*, Lincoln, University of Nebraska Press 1967, pp. vii-x.
- ⁴⁰ L.A. Sherman, *Analytics of Literature*, Boston, Ginn 1893, p. ix-x.
- ⁴¹ G.W. Baehne, *Practical Applications of the Punched Card Method in Colleges and Universities*, New York, Columbia University Press 1935, pp. 405-8.
- ⁴² F. Moretti, *Distant Reading*, cit., p. 67.
- ⁴³ Ibidem, p. 48.
- ⁴⁴ Ibidem, p. 122.
- ⁴⁵ Ibidem, p. 1.
- ⁴⁶ F. Moretti, *Graphs*, cit., p. 50.
- ⁴⁷ S. Aravamudan, *Enlightenment Orientalism: Resisting the Rise of the Novel*, Chicago, The University of Chicago Press 2012, p. 256(note 13).
- ⁴⁸ J. Goodwin, J. Holbo, *Reading Graphs, Maps & Trees: Responses to Franco Moretti*, Anderson (SC), Parlor Press 2011.
- ⁴⁹ G. Spivak, *Death of a Discipline*, New York, Columbia University Press 2003, p. 30.
- ⁵⁰ F. Moretti, *Distant Reading*, cit., p. 181.
- ⁵¹ Ibidem, p. 122.
- ⁵² F. Moretti, *Graphs*, cit., p. 5.
- ⁵³ G. Rockwell, S. Sinclair, *Towards an Archaeology of Text Analysis Tools*, Lausanne, EPFL - UNIL 2014.
- ⁵² G. Rockwell, *What Is Text Analysis, Really?*, in «Literary and Linguistic Computing» 18.2 (2003), pp. 214.
- ⁵³ F. Moretti, *Graphs*, cit., p. 2.