Digital Literature

by Florent Coste

How does big data contribute to our understanding of Hugo, Balzac or Flaubert? A great deal, because far from being a mechanical accumulation of data on literary texts, the digital humanities transform our relationship to works and the way we read them.

Reviewed: Franco Moretti (Ed.), *La littérature au laboratoire*, transl. V. Lëys, in collaboration with A. Gefen and P. Roger, Paris, Ithaque, 2016, 224 pp., 26 €.

The promotion of the digital humanities affects the way in which we apprehend, read, treat and conceptualise literature. Under Franco Moretti's impetus, it is accompanied by a new critical and theoretical approach, based on the experimental bias of distant reading. But how are these experiments in digital, also described as computational, literary criticism organised? How do they change our general grammar of literature and the way we write its history?

Literature in the laboratory life

Inaugurating a new collection called Theoria incognita that claims to engage in an uninhibited and offensive theoretical practice, *La littérature au laboratoire* is the fruit of a translation initiative that opportunely follows on from the import into France of computational criticism – a form of computer-based literary criticism, in the context of the

digital humanities.[1] Backed by Franco Moretti, who played a key role in initiating the research,[2] this volume is nonetheless a collective work, a compilation of 8 articles sometimes signed by 5 or even 6 authors. It was initially available on the Stanford Literary Lab site, currently directed by Mark Algee-Hewitt.

To start with, the overall volume has the advantage of providing a meta-theoretical view of the digital humanities, in opposition to the usual caricatural notion we have of them as cold humanities, serial, algorithmic, highly equipped and funded, supporting neoliberalism, based on a hasty externalisation of reading, eliminating the traditional hermeneutic missions and revealing a little desirable form of scientism. In short 'the pervasive clichés on the simple-minded positivism of digital humanities'. (Pamphlet 6, December 2013, page 9). Far from showing researchers content to run machines and question gigantic corpuses to lazily extract data, on the contrary it is important to evaluate the new ways of working using the digital humanities.

The specifically digital work appears here with all the uncertainties that an exploratory research movement opens up: you test, you try out, you resituate things in perspective, you gather around a table ('as essential a tool as the really expensive ones'), the tasks are divided up, you discuss slip ups, you abandon some of the original hypotheses, and while reorganising, you reconfigure your corpuses, take your time, as you shift between 'solitary work, small group discussions and rivers of emails' 'here only patience will do'[3] and you end up publishing results that remain provisional. *La littérature au laboratoire* is a book that recognises the virtues of dissent and controversy and that refuses to ignore the operations and mediations that contribute to the progressive and often tumultuous production of knowledge.[4] In this respect, the volume is extremely well titled and allows us to see the workings of a sort of hermeneutic, equipped with hard science, but above all, collaborative.

---

[1] For a nuanced understanding of the issues involved in the acclimatisation of literary problems to the digital humanities, see Alexandre Gefen, 'Les Big data dans la littérature et dans les arts », in M. Bouzeghoub and R. Mosseri (Ed.), *Les Big data à découvert*, Paris, CNRS Éditions, 2017, pp. 240-241; Alexandre Gefen (Ed.), *Des chiffres et des lettres*, *Critique*, n°819-820, 2015.

[2] Franco Moretti, *Graphs, Maps, Trees: Abstract Mdoels for a Literary History*, Verso, London, 2005; *Distant Reading*, Verso, London, 2013.

[3] As 'For rapidity nothing beats traditional interpretation'. Pamphlet 12, April 2016 (p. 5).

[4] This includes the scientific phases of the work as well as all the services mobilised to create the samples: 'Remember this was a search involving many excellent librarians in London, Cambridge, Los Angeles and of course at Stanford; a half dozen researchers at the Literary Lab; plus people at Hathi, Gale, and so on. (…) The Literary Lab has some money for research (though, make no mistake, not that kind of money). In other words, one could hardly hope for better resources. And yet it took about six months to receive from Hathi and Gale the set of texts that should have allowed us to move from the initial 30%, to around 70-80% of the random sample' Pamphlet 11, January 2016, p. 2.

Big data and smart data

So what do they actually measure at Stanford Literary Lab? Corpuses of texts are X-rayed, to identify imperceptible and specific linguistic features and motifs that can be identified as the typical signature of a genre (those, for example, that make a novel a gothic novel). The frequency and redundancy of the words are quantified according to their grammatical nature, to evaluate the lexical diversity and informational wealth; the same process is applied to verbal forms, semantic occurrences, syntagmatic chains and combinations of propositions. The space the dialogue and the narrative occupy in paragraphs of 19th century novels is evaluated, or the proportion of nouns and verbs in World Bank reports; the speech verbs in novels are classified by intensity, to measure the loudness in these works.

In a real attempt to explicate the issue, the Stanford Literary Lab teams also tackle the question of knowing what the accumulation of data enabled by the digital humanities contributes. The advantages are far from negligible: computational research corroborates literary theories and categorisations (detractors would say that this is breaking down open doors, trivially confirming what is already known); measurement also provides further details and eliminates the impressionist judgements literary criticism is capable of, ('it's new because it's precise'). With databases like this, we reconfigure less canonical and in the long term, more open corpora, almost as large as the archive, with big data becoming long data in this case. But because data sometimes produces counter-intuitive results that go against literary common sense, these digital surveys also have the means to draw upon solid elements of proof to refute literary theories; thus an analysis of the narrative graphs of plays by Shakespeare and Sophocles, where the key characters do not necessarily occupy the most space, refute the Hegelian view of tragedy as a dialectic conflict. The articles that make up *Littérature au laboratoire* are based on a desire to subject our literary theories to the test of empiricism, and it is important to underscore the paradox: abstraction is the shortest path to a new investigative regime in literary studies, founded upon genuine empiricism.

Even more fundamental, the computational contribution transforms our relationship to literature and the way we read it. According to F. Moretti, computational criticism does not only consist in seeing things at a larger scale:

> […] when we work on 200,000 novels instead of 200, we are not doing the same thing, 1,000 times bigger; we are doing a different thing. (Pamphlet 15, September 2017, p. 2).

Far from only playing a secondary and ancilliary role, computer-based studies create a qualitative shift that involves a transformation in the way literary objects are created. The digital reconfiguration of literary corpora changes the scales of reading, and with them the units of observation: 'The new scale changes our relationship to our object, and in fact it changes the object itself." (p. 1). Pamphlet 15 Sept 2017. On one hand, the digital humanities force the literary expert to reason in terms of correlating very small units with conclusions at a very large scale, to dissolve the texts in a cloud of data, and at the outset, to evade the literary

phenomenality permitted by the ordinary experience of individual, anthropocentric, silent and largely hermeneutic reading. On the other hand, as ' "Operationalising" or the function of measurement in modern literary theory'[5] claims, the digital humanities force literary criticism to actively construct new concepts that reveal intangible realities, which need to be visualised as diagrams, clouds of points, graphs and graphics, etc. The change is similar to what astronomical science underwent with the invention of the telescope: now specialists of literature 'could envisage generation-by-generation maps of the literary universe, with galaxies, supernovae, black holes ...' Pamphlet 1, Jan 15, 2011 (p. 10). We know very well that the data provided by the Hubble telescope are in fact not constructed and created by knowledgeable protocols made up of complex technical reconstructions and mediations. But this constructed *image* of reality is nonetheless an image of *reality*.

This interplay of scales also develops an acuity and reflexivity in our use of concepts, a bit like when we try on different pairs of spectacles and become aware of how these devices condition our perception of reality. To use a distinction made by Koyré (Pamphlet 12, April 2016) (p. 3), until now our reasoning has been based on literary tools that are an extension of our organs, but were ill-adapted to quantification (for example, the idea of the main character). Equipped with databases, metadata and algorithms to treat them, we can henceforth base our reasoning on instruments, which by sparking off a series of measurements and operations, are capable of capturing realities that are not conditioned by the senses (for example, the character-space, or the quantity of narrative space allocated to one character or the other; or even the loudness of a novel).

For this reason, the computational approach Stanford Literary Lab defends should be seen as a call to reopen the literary concept factory. And, based on a reading of the book, our reply to those who complain that the digital humanities sound the death knell of literary interpretation[6] is that, in reality, the influence of perfected algorithms does not mean the end of interpretation. It only implies a shift in the researcher's intellectual intervention, and it is still he or she who makes this mass of data intelligible, and identifies causalities in the correlations he or she observes (Pamphlet 5, June 2013). Hence, big data can legitimately be seen as smart data.

Is literary history the same as other histories?

As usual, F. Moretti and his team's reflections are reinvigorating from the point of view of literary epistemology. They propose advantageous moments of conceptual and methodological lucidity. Evaluating literary material from a digital and computational

---

[5] This chapter was first translated and published in Alexandre Gefen (Ed.), *Des chiffres et des lettres*, *Critique*, n°819-820, 2015 pp. 712-734.

[6] 'It would seem that the expansive 'world' of this quant-driven ecosystem tends towards contraction of the teconomy of interpretation'. (Emily Apter, *Against World Literature*, New York, Verso, 2013, p. 56).

perspective, presumes, as the authors of this collective show, applying remarks made by Weber, Kuhn, Popper, Pomian, Braudel, Leroi-Gourhan, Canguilhem, to points as varied and essential as those concerning the norm and the exception, the ontological consistency of a scientific fact, the representativeness of a sample or relationships between correlation and causality.

A new vision of literary history thus emerges, supported by verificationist and experimental thinking, in an explanatory and hypothetico-deductive style of literary reasoning. Hypotheses are posited most often to be proven false and abandoned; a recognition of failure and a capacity for self-criticism have the merit of protecting us not only from self-supporting hypotheses, which we are intuitively tempted to cling to, but also serve to develop the robustness of the last hypotheses, which we are led to defend as solid and corroborated results – in short, scientific facts like any others.

The effect the digital humanities can have on literature is, in this sense, comparable to what the introduction of the long-term did to history, by eliminating factual history and diminishing the role of great men, to focus on questions like salary progression, monetary fluctuation or demographic pressure. From this viewpoint Braudel and the Annales school are references that occur frequently in the '*Pamphlets*'.

Here, F. Moretti wins a twofold victory against traditional hermeneutics: of course and laudably, he does away with any literary reading that only pays attention to exceptions, differences or singularities to focus on the identification of regularities, series, cohorts or patterns amidst the chaos and noise created by the data. Nonetheless, by a sort of highly ironical trick, he manages to fulfil the role of hermeneutics, far beyond the expectations of those who claim to be its guardians. Just as the social sciences can comprehend human behaviour at levels that go beyond the simple understanding the actors themselves have of it, digital literary criticism turns away from the tired but at least persistent mythological view, according to which the meaning of a work would be a largely individual affair, limited to a handful of actors (the author and the reader). In this way it appropriates the means of understanding works better than individual actors would be able to at their humble level, by turning them into a complex social phenomenon, and making it possible to observe the correlations that remain beyond these actors' reach.

F. Moretti's team thus intends to work towards a 'history of literature as a whole' Pamphlet 12, April 2016, p. 6. But what does this actually mean? To start with, it is not a global history of literature. With its methodological recommendations drawn from Braudel, F. Moretti's works have strangely never seemed as far from the reflections on world literature

or world systems. Then, it is also a fairly internalist history, a literary history rather than strictly speaking, a history of literature connected to other aspects of history.[7]

Conducting these digital experiments in the literary laboratory, is not only time consuming, it imperceptibly tends towards two additional flaws. These encourage the abstraction specific to an attempt at modelisation and undoubtedly encourage the researcher to focus on the fictional corpora of the 19th century and provisionally, not permanently, to lose sight of the wide global view. It also sharpens a morphological type of attention, which keeps the researcher at a distance from the socio-historical complexity of the ecosystems in which literary works circulate and are transformed.[8]

In fact, for F. Moretti, literary forms are something like the yardsticks that constitute a Weberian ideal-type:

> "a mental construct" that "cannot be found empirically anywhere in reality", but that, once constructed, can be used "for comparison with and measurement of reality". (…) a mental construct which we will never find as such in individual works, but which we can use to "measure" their relationships. Form will never explain a single text, and is the only thing that can explain a series of them. (Pamphlet 15 Sept 2017 p. 9)

Hence it is important to know how to 'return from abstraction to literary history' (*ibid.*), without losing sight of the '*big questions*' Pamphlet 4, May 2012, p.49. For this reason this book should be read for what it is: an on going clarification of more or less exportable and adaptable protocols of analysis and a stimulating invitation to construct data and develop modes of treatment that are applicable to other genres, other (less Eurocentric) literary cultures, other (less modern) periods.

Published in laviedesidees.fr, 8 May 2017.

---

[7] It is true that the authors claim to want to 'open novels and discover whether their social function – popularity, prestige, both or neither – maintain any kind of relationship with their morphological characteristics' (p. 236). Let us admit that for literary works this is an attempt at a fairly modest historical decompartmentalisation.

[8] Alexander Beecroft (*An ecology of World Literature*, London, Verso, 2015) for example, focuses on the complexity of the ecosystems, with their Moretti style modelised systems, and their complex interactions, which need to be explained. Emily Apter also underscores that "While Moretti himself often enlivens his quantitative formalism with witty phrase and bold hypotheses about cultural and political change, the Literary Lab pamphlets adhere to the flattened diction of technics (…) If the labs are harbingers of the future status of the systems theory in literary studies, they spell out less concern with ecosystems (and their organicist, genetic, evolutionary applications) and more concern with computational systems *tout court*, albeit in the service of literary and social history'. (*Against World Literature, op. cit.*, p. 55-56.).