Algerian Journal of Manuscripts Volume 17 / Special Issue: (2021), pp. 1-12

Hardening Soft Sciences, Softening Hard Sciences: Literary Studies in the Age of 'Big Data'

Sihem Saibi [1*]

[1] University of Bejaia- LESMS Lab, (Algeria),

sihem.saibi@univ-bejaia.dz

Received: 06/10/2021 Accepted: 24/12/2021 Published:31/12/2021

Abstract

Today, technology is breaking down the barriers between disciplines and subjects, and one of the immediate consequences of what Milad Doueihi calls 'the digital turn' is interdisciplinary. The intersection and interaction of many disciplines have not spared literary criticism. The present paper is a reflection on the impact of the digital turn on literary studies. It is divided into four parts and is structured as follows. First, I start by questioning the similarities and differences between soft and hard sciences. The following part is a brief survey of scholarly sources on the emerging field of DH. Some basic concepts like big data, DH, distant reading, and close reading are defined. The following part reviews some available tools that support close reading and distant reading. The paper ends with some remarks on the necessity to adapt literary studies to these changes and use qualitative and quantitative methods in teaching literature.

Keywords: Big data; Digital Humanities (DH); close reading; distant reading; literary studies.

1. INTRODUCTION

It is commonly thought that the marriage of science and art is impossible. "The arts," according to Nelson Goodman (1978), "must be taken no less seriously than the sciences as modes of discovery, creation, and enlargement of knowledge in the broad sense of advancement of the understanding" (Goodman, 1978, p. 102). Goodman's perceptive remark on

[1] ***Corresponding author/ Sihem Saibi***

the values of science and art is significant to our investigation. Are sciences and arts different? Are they similar? Science, unlike art, should be impartial; the method used by the scientist should be rigorous. Art, on the contrary, should be subjective, and the artist has freedom that the scientist does not have. This view about art and science is very common. In fact, it is the same idea that has moderated many discussions in human sciences and other sciences. Alex Rosenberg (2012), for example, commenting on humanities said:

> When it comes to real understanding, the humanities are nothing we have to take seriously, except as symptoms. But they are everything we need to take seriously. When it comes to entertainment, enjoyment, and psychological satisfaction. Just don't treat them as knowledge or wisdom. (Rosenberg, 2012, p. 307)

The two quotations summarize the continuing divergence between researchers and mainly between humanists and non-humanists. The debate around art and cognition has interested many intellectuals and scholars. Gaut's most famous quotation has been used later by humanists and post-humanists:

> There is no doubt that art can be a source of knowledge, yielding facts about period or place right up to complex psychological or r moral truths. To the extent that any learning is valuable then this too is valuable. But learning in itself does not look like a distinctive artistic achievement; maybe the same truths could be learnt in other way; maybe some of what is learnt is incidental to the work itself (Gaut, 2006, 115)

Accordingly, if we want to reach 'truth', we have to rely on science as it is the only means to acquire knowledge.

2. Literature Review

The *dangerous liaison* between technical fields and literary studies as mentioned above has attracted the attention of many researchers. In what follows, we will review some of the most prominent studies in Digital

Humanities and define basic concepts to understand that the literary text as 'a black box' is just a myth.

Two definitions of big data are proposed here before moving to DH. Mario Aquilina (2017) defines big data as "the enormous size of data and the process of analyzing data" (Aquilina, 2017, 494). Boyd and Crawford (2012) define Big Data as "a cultural, technological, and scholarly phenomenon" (Boyd & Crawford, 2012, 663). Big data has drawn the attention of many researchers in humanities as well as computing. Digital literary studies, an emerging discipline, rely mainly on big data. Jean-Gabriel Ganascia, in "The Logic of the Big Data Turn in Digital Literary Studies", claimed that "big data can renew with the use of computers, the Humanities, i.e, the disciplines rationally studying human works and cultural production. Digital literary studies are emblematic of these new approaches, certainly because they constitute the oldest subfield of D.H" (Ganassia, 2015, p.1).

Before moving to the impact of DH on literary studies, it is important to make a distinction between the sciences of nature and the sciences of culture. Heinrich Rickert, a German philosopher said that humanities are empirical sciences. His student and fellow researcher Ernest Cassierer in 1923 suggested that the two sciences do not have the same logic. In German, Kulturwissen refers to the sciences of culture while the word Naturewissenschaft refers to the sciences of nature. The purpose of Kulturewissen is to study and give meaning to human works and cultural records; however, Naturewissenschaft deals with physical perceptions (Ganascia, 2015). The two sciences use different methods and have different logics. While the sciences of culture attempt to comprehend particulars, the sciences of nature endeavor to discern general laws. Ganascia holds that there is no rivalry between the logic of sciences of culture and digital tools; in fact, digital methods and tools support traditional humanities.

2.1. Digital Humanities: A Project or an Institutionalized Discipline?

In his study of DH, Anne Burdick et al (2012) said that DH was born of "the encounter between traditional humanities and computational methods" (p.3). This new discipline allows a better visibility of buried data. In another definition, DH refers to "Research that uses information technology as a central part of its methodology, for creating and/or processing data. The use of the term "digital humanities" reflects a growing sense of the importance that digital tools and resources now have for humanities subjects"[2].

The processes of archiving works and the development of free licences like Creative Commons had an impact on the development of DH. Thousands of texts were not accessible before their digitization. Consequently, researchers and students can have access to and share documents around the world. In literary criticism, the digitization of literary works permitted a free circulation of works by unknown and unheard of writers.

In order to understand if DH is a field, a project, or an institutionalized discipline, one has to understand its complex history. Patrick Svensson notes that the evolution from computing to digital humanities "was not much more than a change in names, and in actuality, the epistemic tradition of humanities computing has remained strong in the digital humanities" (Svensson, 2016, p.36). In the 1980s, DH, as Svensson explains, was an emerging field:

> In the initial 1987 welcome message for the *Humanist* e- mail list, Willard McCarty wrote that "computing in the humanities is an emerging and highly cross- disciplinary field."1 The quality of being emergent is often associated with the uncertain institutional and disciplinary status of the field as well as with much discussion about what it is and what it can be. It is not surprising, perhaps, that we can detect a certain amount of weariness among old- timers who have been debating these issues for a long time, especially now when there is a sense of a stronger institutional position and the possibility

[2] *Oxford University Digital Humanities Center*, quoted in Meunier (2018-2019), p. 19.

> of leaving behind some of the uncertainty and hardships. (Svensson, 2016, p.37)

Xavier Laurent Salvador in his article "Littérature et sciences humaines a l'heure numérique" (2019), summarized well the first consequences of DH on other disciplines. According to the researcher, DH has broadened the fields of disciplines and moved the arbitrary frontiers between human sciences and other sciences (Salvador, 2019, p.6). One should not see DH as a replacement but as an extension of classical / traditional humanities. The humanists' apprehension can be justified because of the rapid evolution of DH from a mere project to an institutionalized discipline. In this regard, Milad Douihi in "Quelles humanités numériques?" put in plain words that:

> Les humanités numériques ne cessent de susciter critiques et interrogations quant a leur statut institutionnel, leur histoire et surtout leur position dans le paysage intellectuel et académique. Cette situation s'exprime en partie par la manière dramatisée dont le monde savant vit la conversion numérique de nos sociétés par la floraison des manifestes annonçant une rupture avec le passé et la naissance de nouvelles méthodes permettant des explorations radicalement inédites des objets culturels. (Douihi, 2015, pp.819-820)

DH is seen by some humanists as a post-discipline (Welger-Barboza 2012) and a transdiscipline by Le Deuff (2014) while others prefer to see it as a new field of research (Four 2013), or a turning point by Caryol and Morandi (2016)[3]. Today, there are hundreds of projects around the world that are financed by labs, centers, and universities to study known and unknown authors. Among the well-known projects we can cite The Shelley Godwin Project, Francis Bacon Projects, and The Women Writers Projects.

[3] For more details, see Gefen (2018), Janickel, et al (2015), and Scrivener (2017).

2.2 Close Reading and Distant Reading

As mentioned in the introduction, science and art are different and have different logics and epistemologies. Matthew L. Jockers (2013) notes that "careful observation" and "sustained and concentrated reading of text", in literary studies are important (Jockers, 2013, p.6). He continues saying that experimentation in science allows "a method through which competing observations and conclusions may be tested and ruled out. With a few exceptions, there is no obvious corollary to scientific experimentation in literary studies" (Jockers, 2013, p.6). In literary studies, in contrast, researchers attain conclusions that are not always testable and repeatable. Because literary researchers and readers often rely on (subjective) interpretation, the conclusions they draw are not verifiable, and there is a possibility to give an unsound, inaccurate reading.

Close reading, a highly notorious and contested method of text analysis today among teachers and critics, became popular in the twentieth century with I. A. Richard's essay "Practical Criticism" (1929) and other works by the New Critics like John Crowe Ransom, Cleanth Brooks, and T.S. Eliot. Close reading is defined as "a neutral first step in understanding literature" (Showalter,2003, p.56), and a "slow reading, a deliberate attempt to detach ourselves from the magical power of storytelling and pay attention to language, imagery, allusion, intertextuality, syntax, and form" (Showalter, 2003, p.98). This method is used by teachers of literature to encourage students to focus on the words on the page and ignore the writer and the context.

As already noted, distant reading[1] was introduced in the twenty first century by the Italian scholar Franco Moretti who once said that distant reading is "a little pact with the devil: we know how to read texts, now let's learn how not to read them"[4]. Moretti's words may seem malevolent because they announce the death of close reading. Moretti's project, which was suspicious to literary scholars, has proved today that the future of

[4] https://newleftreview.org/issues/ii1/articles/franco-moretti-conjectures-on-world-literature

literary criticism in the light of technological advances is far from being 'dead'.

3. Close Reading and Distant Reading Tools and Techniques

After defining key concepts, we may now turn to introduce text mining and web-based tools and techniques. Visualization is a very common DH tool which used by humanists and scientists. Data visualization tools are ideal when literary critics and readers want to a have a picture of data in different forms. Franco Moretti in 2005 introduced visualization in his famous yet controversial article "Graphs, Maps, Trees". In "Visual Text Analysis in Digital Humanities", the scholars listed numerous text analysis tasks that are classified as follows: similar patterns, topics, named entities, corpus analysis, and texts of interest.

Literary criticism and literary studies can largely take profit from technological advance that is offering to teachers and researchers every day new pedagogical toolkits. In fact, there are many digital tools that may support traditional close reading. For example, eMargin https://emargin.bcu.ac.uk/ is an interesting collaborative annotation tool that can be used to highlight, color-code, and annotate. We have another tool which is Poem Viewer https://nms.kcl.ac.uk:8443/poemvis/index.html that can be used by students of literature when dealing with poems. Poem Viewer is a web-based tool for visualizing poems. This tool uses glyphs that show phonetic and semantic features.

Data visualization tools are very helpful because they process and transmit information easily and quickley. POEMAGE http://www.sci.utah.edu/~nmccurdy/Poemage/ is a visualization system for exploring the sonic topology of a poem. This tool was developed at the University of Utah as part of a collaboration between data visualization experts and poets/poetry scholars. Another toolkit is PRISM http://pedagogy-toolkit.org/tools/PRISM.html. PRISM is a tool for crowd sourced interpretation created by Scholarslab. It allows contributors add

their commentaries and interpretations of texts online by highlighting key terms, concepts, and sections in the text.

Distant reading tools refer to software tools that are also commonly used by researchers in computational text analysis. There are several tools that detect citations (Phoebus) and semantic reformulations (DeSeRT) and analyze the genetics of literary texts (Medite). The most common distant reading techniques include the following: maps, timelines, heat maps, and interactive dot plots. Maps are good when we want to visualize geospatial information in a literary text or even a group of texts. Timelines are effective in visualizing time in one text or a group of texts. In contrast, Heat maps (block matrices) are useful in corpus analysis when we want to analyze similar patterns or look for text reuse or plagiarism. They can be used with other techniques like networks, nodes, and trees to visualize relations between texts. Interactive dot plot, on the contrary, are useful in detecting text reuse. To this list of distant reading techniques, we can add tag clouds, tag pies, dust-and-magnet visualization, and topological landscape.

4. CONCLUSION

Many skeptical scholars and critics warned against technological tyranny and the use of artificial intelligence. André Moritz once wrote in "Objects and Methds of Literary History" (1966) that:

> This craving for knowledge is truly speaking, the only scientific part of their efforts. There is no scientific method in literary history in the sense that there is no method, however well adapted to a given science, that literary history can transplant and apply that this is possible is responsible for much poor and childish work: statistics and charts, evolution of species, and quantitative analysis are processes, methods, and hypotheses excellent in their place, but their place is not in literary history. (Moritz, 1969, p.3)

Many have predicted the failure of computer-aided literary studies; for example, Paul Delany in "L'ordinateur et la critique littéraire" (1994) has made a strange parallel between Shelley's monster and computers. Mark

Olson in "Signs, Symbols, and Discourses: A New Direction for Computer-aided Literary Studies" (1993-94) clearly claimed that computers were unable to interpret certain devices like humor and irony. Today, recent research has proved the contrary; text-mining techniques and web-based tools can help teachers and students with the analysis of literary texts. However, distant reading should be used *to support and enhance* close reading and learn more about literary texts.

5. Bibliography List:

1. Aquilina, M. (2017). The Work of the Literary Critic in the Age of Big Data. *Interdisciplinary Literary Studies*, 19 (4), 493-516.

2. Allison, S., et al. (2013). "Style at the Scale of the Sentence." Literary Lab, Pamphlet 5. Available in http://litlab.stanford.edu/LiteraryLabPamphlet5. pdf. 22 September 2015.

3. Allison,S. et al. (2011). Quantitative Formalism: An Experiment. Literary Lab, Pamphlet 1. Available in http://litlab.stanford.edu/LiteraryLabPamphlet1. pdf. 22 September 2015.

4. Barry, A., & Born, G., eds. (2013). *Interdisciplinarity: Reconfigurationsof the Social and Natural Sciences*. Routledge.

5. Bode, K. (2012). *Reading by Numbers: Recalibrating the Literary Field*. Anthem Press.

6. Boyd, D., & Crawford, K. (2012). Critical questions for big data provocations for a cultural, technological, and scholarly phenomenon. *Information Communication &Society*, 15(5), 662-679.

7. Boyles, N. (2013). Closing in on Close Reading. *Educational Leadership*, 70 (4), 36–41.

8. Cassirer, E. (1961). The *Logic* of the *Humanities*. Yale University Press.

9. Brooks C R., Warren, R. P., eds. (1938). *Understanding Poetry.* Henry Holt and Co.

10. Brooks, C. (1943). *Understanding Fiction.* Henry Holt and Co.

11. Burdick, A, et al. (2016). *Digital_Humanities.* MIT Press.

12. Caryol, V., & Morndi, F. (dir.). (2016). *Le tournant numérique des sciences humaines et sociales*, Pessac, Presses de la Maison des Sciences de l'Homme d'Aquitaine. DOI : 10.4000/books.msha.1305

13. Four, P.A. (2013). *Comment les digital humanities (ou les humanités numériques) transforment les SHS*, Document de la Direction de la Prospective et du Dialogue Public/Grand Lyon. http://fr.slideshare.net/pierrealainfour/comment-les-digital-humanities-ou-les-humanits-numriques-transforment-les-shs

14. Coles, K., & Lein, J. G. (July 2013). Solitary Mind, Collaborative Mind: Close Reading and Interdisciplinary Research. In Proceedings of the Digital Humanities, http://dh2013.unl.edu/abstracts/ab-217.html

15. Coles K., et al (2014). Empowering Play, Experimenting with Poems: Disciplinary Values and Visualization Development. In Proceedings of the Digital Humanities 2014.

16. Delany P. (1994). L'ordinateur et la critique littéraire. *Littérature*, (96), 6-18. doi : https://doi.org/10.3406/litt.1994.2349

17. Doueihi, M. (2015). Quelles humanités numériques ? », *Critique*, 8 (819-820), 704-711

18. Doueihi M. (2008). *La grande conversion numérique*. Seuil.

19. Ganassia, J. G., et al. (December 2015). The Logic of the Big Data Turn in Digital Literary Studies. *Frontiers in Digital Humanities*, 2 (5), 1-5. 10.3389/fdigh.2015.00007

20. Gaut, B. (2006). Art and Cognition: In M. Kieran (Ed.), *Contemporary Debates in Aesthetics and the Philosophy of Art* (pp.115–126). Blackwell.

21. Gefen, A. (2018). The Empirical Turn of Literary Studies. Nicoletta Pireddu. Reframing Critical, Literary, and Cultural Theories (pp.119-135). Palgrave Macmillan Cham.

22. Goodman, N. (1978). *Ways of Worldmaking*. Hackett.

23. Kassous I. Z , Sarnou H . (2021). Investigating The Effectiveness of Digital Storytelling in Developing Algerian Learners' Emotional Intelligence: A Case Study Of Third Year Lmd Students Of English At Abdelhamid Ibn Badis University. *Journal of Studies in Language, Culture and Society (JSLCS)*. 4(1), 109-120.

24. Jänicke1, S. Et al. (2015). On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges Eurographics Conference on Visualization (EuroVis) STAR – State of The Art Report R. Borgo, F. Ganovelli, and I. Viola (Editors),1-21.

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

26. Jockers M.L. (July 16 - 22, 2012). Computing and Visualizing the 19th-Century Literary Genome. *Proceedings of the Digital Humanities*. University of Hamburg. Hamburg, Germany.

27. Le Deuff, O (dir.) (2014). Les temps des humanités digitales. La mutation des sciences humaines et sociales. Limoges, FYP éditions.

28. Meunier, J.G. (2019). Le paradoxe des humanités numériques. *Quaderni*, 98, 19-31. 10.4000/quaderni.1407

29. Moretti, F. (2005). *Graphs, Maps, Trees: Abstract Model for Literary History*. Verso.

30. Moretti, F. (2013). Operationalizing': Or the Function of Measurement in Modern Literary Theory. Literary Lab, Pamphlet 6. Available in http://litlab.stanford.edu/LiteraryLabPamphlet6.pdf. 22 September 2015.

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

32. Moretti, F. 2013. *The Bourgeois: Between History and Literature*. Verso.

33. Morize, A. (1969). Problems and Methods of Literary History. Biblo and Tannen Publishers.

34. Olson, M. (1994-95). Signs, Symbols, and Discourses: A New Direction of Computer-aided Literary Studies. *Computers and the Humanities*, 27 (5-6), 309-14.

35. Richards, I. A. (1930). *Practical Criticism: A Study of Literary Judgment*. Kegan Paul, Trench, Turbner and Co. Ltd.

36. Rickert, H. (1921). *Kulturwissenschaft und Naturwissenschaft*. Mohr.

37. Rosenberg, A. (2012). The Atheist Guide to Reality. Norton.

38. Showalter, E. (2003). *Teaching Literature*. Blackwed.

39. John, M., Et al. (2016). Visual Analytics for Narrative Text: Visualizing Characters and their Relationships as Extracted from Novels. Proceedings of the 7th International Conference on Information Visualization Theory and Applications (IVAPP '16).

40. Scrivner, O., & Davis, J. (2017) Interactive Text Mining Suite: Data Visualization for Literary Studies January 2017 Conference: Corpora in the Digital Humanities at Bloomington, Indiana.

41. Svensson, P. (2016). Digital Humanities as a Field. In *Big Digital Humanities: Imagining a Meeting Place for the Humanities and the Digital* 36–81. University of Michigan Press. https://doi.org/10.2307/j.ctv65sx0t.6

42. Welger-Barboza C. (2012). Les digital humanities aujourd'hui : centres, réseaux, pratiques et enjeux, Mounier Pierre (dir.), *Read/Write Book 2 : Une introduction aux humanités numériques*, Marseille, OpenEdition Press.10.4000/1895.2

43. Wieviorka, M. (2013). L'impératif numérique ou la nouvelle ère des sciences humaines et sociales, Paris, CNRS editions. 10.3917/cnrs.wiev.2013.01