

La littérature en numérique

par Florent Coste

Qu’apportent les *big data* à notre interprétation de Hugo, Balzac ou Flaubert ? Beaucoup, parce que les humanités numériques, loin d’accumuler mécaniquement des données sur les textes littéraires, changent notre rapport aux œuvres et notre manière de les lire.

Recensé : Franco Moretti (dir.), *La littérature au laboratoire*, trad. V. Lëys, avec la collaboration d’A. Gefen et P. Roger, Paris, Ithaque, 2016, 224 p., 26 €.

La promotion des *digital humanities* affecte nos manières d’appréhender, de lire, de traiter, de conceptualiser la littérature. Sous l’impulsion de Franco Moretti, elle s’accompagne d’une nouvelle modalité critique et théorique, forte d’un parti pris expérimental : le *distant reading*, la lecture à distance. Mais comment organiser ces expérimentations de critique littéraire digitale, dite aussi computationnelle ? Et que changent-elles à notre grammaire générale de la littérature et à notre manière d’en écrire l’histoire ?

La littérature dans la vie de laboratoire

Inaugurant une nouvelle collection intitulée *Theoria incognita* qui revendique une pratique désinhibée et offensive de la théorie, *La littérature au laboratoire* résulte d’un travail de traduction venant opportunément poursuivre l’importation en France de la critique computationnelle – une critique littéraire instrumentée par l’informatique, dans le cadre des humanités numériques ou *digital humanities*¹. Placé sous l’égide de Franco Moretti qui a très

¹ Pour une compréhension fine des enjeux de l’acclimatation des problématiques littéraires aux humanités numériques, voir Alexandre Gefen, « Les Big data dans la littérature et dans les arts », in M. Bouzeghoub et R. Mosseri (dir.), *Les Big data à découvert*, Paris, CNRS Éditions, 2017, p. 240-241 ; Alexandre Gefen (dir.), *Des chiffres et des lettres, Critique*, n°819-820, 2015.

tôt joué un rôle d'initiateur en la matière², ce volume n'en reste pas moins un collectif compilant 8 articles signés parfois à 5 ou 6 auteurs et initialement mis en ligne sur le site du Stanford Literary Lab, que dirige aujourd'hui Mark Algee-Hewitt.

L'ensemble du volume a d'abord le mérite de nous faire porter un regard métathéorique sur les humanités numériques, qui prend à contre-pied la conception caricaturale que l'on peut s'en faire, à savoir des humanités froides, sérielles, algorithmiques, lourdement équipées et financées, faisant le jeu du néolibéralisme, reposant sur une externalisation expéditive de la lecture, évacuant les missions herméneutiques traditionnelles, faisant preuve d'une forme peu désirable de scientisme. Bref « le flot des clichés sur le positivisme simplet des humanités numériques » (p. 106). Loin de montrer des chercheurs se contentant de faire mouliner des machines et d'interroger des corpus gigantesques pour extraire paresseusement des données, il est au contraire important de prendre la mesure des nouvelles manières de travailler impliquées par les humanités numériques.

Le travail proprement digital se montre ici dans toutes les incertitudes qu'ouvre un mouvement de recherche exploratoire : on teste, on tâtonne, on remet en perspective, on se réunit autour d'une table (« un outil tout aussi essentiel que d'autres équipements plus coûteux »), on se répartit à nouveau les tâches, on discute de ses faux pas, on abandonne des hypothèses de départ et, tout en se réorganisant, on reconfigure ses corpus, on prend du temps, ballotté qu'on est entre « épisodes solitaires, discussions en petit comité et flots de courriels », mais surtout on s'arme de patience³ et on finit par publier des résultats toujours provisoires. *La littérature au laboratoire* est un livre qui reconnaît les vertus du dissensus et de la controverse et qui refuse de passer sous silence les opérations et les médiations contribuant à la production progressive et souvent tumultueuse du savoir⁴. À cet égard, le volume porte fort bien son titre et nous permet de voir à l'œuvre une sorte d'herméneutique munie de la *hard science*, mais surtout collaborative.

² Franco Moretti, *Graphes, cartes et arbres. Modèles abstraits pour une autre histoire de la littérature*, trad. fr. É. Dobenesque, préf. L. Jeanpierre, Paris, Les Prairies ordinaires, 2008 [2005] ; *Distant Reading*, Verso, Londres, 2013.

³ Car « si c'est la rapidité que l'on recherche, alors rien ne vaut les méthodes d'interprétation traditionnelles » (p. 15).

⁴ Cela inclut les étapes de travail scientifique, mais aussi l'ensemble des services mobilisés dans la constitution des échantillons : « N'oublions pas que ce projet a dû jusqu'à présent faire appel à des bibliothécaires expérimentés à Londres, Cambridge, Los Angeles et bien sûr Stanford ; à une demi-douzaine de chercheurs du Literary Lab ; et à des employés du Hathi Trust, de Gale, etc. (...) Le Literary Lab a un peu d'argent pour ses recherches (mais, détrompez-vous, pas *cette* quantité d'argent) : difficile de rêver de ressources plus abondantes. Et pourtant, il nous faut attendre environ six mois pour recevoir du Hathi Trust et de Gale les textes qui auraient pu nous permettre de passer des 30 % initiaux à environ 70-80 % d'un échantillon aléatoire. » (p. 225)

Big data et smart data

Alors que mesure-t-on précisément au Stanford Literary Lab ? On radiographie des corpus de textes, pour y repérer des traits et motifs linguistiques imperceptibles et propres à être reconnus comme la signature typique d'un genre (faisant par exemple d'un roman un roman gothique) ; on quantifie la fréquence et la redondance des mots selon leur nature grammaticale pour évaluer la diversité lexicale et la richesse informationnelle ; on fait de même avec des formes verbales, des occurrences sémantiques, des chaînes syntagmatiques, des combinaisons de propositions ; on évalue la part qu'occupent le dialogue et la narration dans des paragraphes de romans du 19^e siècle ou la proportion de noms et de verbes dans les rapports de la Banque mondiale ; on classe selon leur intensité les verbes de parole dans les romans pour mesurer le volume sonore dans ces œuvres.

Les équipes du Stanford Literary Lab affrontent par ailleurs avec un réel souci d'explicitation la question de savoir ce qu'apporte l'accumulation de données permise par les humanités numériques. Les avantages ne sont pas négligeables : la recherche computationnelle corrobore des théories et des catégorisations littéraires (les mauvais esprits diront qu'enfonçant des portes ouvertes, elle confirme trivialement ce que l'on sait déjà) ; la mesure apporte aussi plus de précisions et évacue les jugements impressionnistes dont est capable la critique littéraire (« si c'est nouveau, c'est, d'abord, parce que c'est précis »). Avec de telles bases de données, on reconfigure des corpus moins canoniques, quasi coextensifs à l'archive et plus ouverts sur la longue durée, le *big data* devenant le *long data*. Mais, parce que les données produisent parfois des résultats contre-intuitifs, qui vont à rebours du sens commun littéraire, ces enquêtes digitales ont également les moyens de réfuter avec des éléments de preuve solides des théories littéraires : ainsi la conception hégélienne de la tragédie comme conflit dialectique trouve un démenti assez ferme dans l'analyse des graphes actantiels des pièces de Shakespeare et de Sophocle où les personnages centraux n'occupent pas nécessairement la plus grande place. Les articles composant la *Littérature au laboratoire* sont portés par un souci de faire passer à nos théories littéraires le test de l'empirie, et il importe d'en souligner le paradoxe : le passage par l'abstraction est le plus court chemin vers un nouveau régime probatoire dans les études littéraires, qui serait fondé sur un empirisme de bon aloi.

Plus fondamentalement encore, l'apport computationnel consiste dans un bouleversement de notre rapport à la littérature et de notre manière de la lire. Selon F. Moretti, la critique computationnelle ne consiste pas seulement à voir les choses en plus grand :

[...] quand nous travaillons sur 200 000 romans plutôt que 200, nous ne faisons pas la même chose en mille fois plus grand ; nous faisons autre chose. (p. 266)

Loin de ne jouer qu'un rôle auxiliaire et ancillaire, l'informatique provoque un saut qualitatif qui engage une transformation des manières de fabriquer les objets littéraires. La

reconfiguration digitale des corpus littéraires modifie les échelles de lecture et avec elles les unités d'observation : « la taille change la nature de l'objet » (p. 268). D'une part, les *digital humanities* poussent le littéraire à corrélérer dans son raisonnement de très petites unités à des conclusions à très grande échelle, à dissoudre les textes dans un nuage de données, et partant à escamoter la phénoménalité littéraire permise par l'expérience ordinaire de la lecture, individuelle, anthropocentrée, silencieuse et pour l'essentiel herméneutique. D'autre part, comme le soutient « L'opérationnalisation ou, du rôle de la mesure dans la théorie littéraire moderne »⁵, les humanités numériques obligent la critique littéraire à construire activement de nouveaux concepts manifestant des réalités peu palpables et nécessitant d'être visualisées par des diagrammes, des nuages de points, des graphes, des graphiques, etc. La mutation est similaire à celle que la science astronomique a connue avec l'invention du télescope : il serait désormais permis au spécialiste de littérature d'explorer et de cartographier « l'univers littéraire avec toutes ses galaxies, ses supernovas et ses trous noirs » (p. 33). Nous savons bien que les données fournies par le télescope Hubble ne sont précisément pas données et résultent de protocoles savants faits de reconstructions techniques et de médiations complexes. Mais cette image construite de la réalité n'en reste pas moins une image-de-la-réalité.

Ces jeux d'échelles développent par ailleurs l'acuité et la réflexivité dans notre usage des concepts, un peu comme durant ces moments où nous essayons différentes paires de lunettes et où nous prenons conscience de la manière dont ces prothèses conditionnent notre perception de la réalité. Pour reprendre une distinction de Koyré (p. 104), nous raisonnions jusque là avec des outils littéraires qui prolongeaient nos organes, mais qui se prêtaient mal à la quantification (par exemple la notion de personnage principal) ; équipés que nous sommes de bases de données, de métadonnées et d'algorithmes pour les traiter, nous pouvons désormais raisonner avec des instruments, qui, en enclenchant une série de mesures et d'opérations, sont capables de capter des réalités qui ne tombent pas sous les sens (par exemple l'espace-personnage, à savoir la quantité d'espace narratif alloué à tel ou tel personnage ; ou encore le volume sonore d'un roman).

C'est pourquoi il faut voir dans l'approche computationnelle défendue par le Stanford Literary Lab un appel à rouvrir la fabrique des concepts littéraires. Mais à ceux qui déploieraient que les humanités numériques signent l'arrêt de mort de l'interprétation littéraire⁶, on répondra avec l'appui du livre qu'en réalité on n'arrête pas d'interpréter sous l'emprise d'algorithmes perfectionnés et qu'on ne fait jamais que déplacer les interventions intellectuelles du chercheur à qui il reste toujours à donner de l'intelligibilité à cette masse de données et à dégager des causalités dans les corrélations qu'il a pu observer (p. 274). C'est ainsi que les *big data* se présentent légitimement comme des *smart data*.

⁵ Ce chapitre a été traduit et publié auparavant dans Alexandre Gefen (dir.), *Des chiffres et des lettres, Critique*, n°819-820, 2015 p. 712-734.

⁶ « Il semblerait que le monde en expansion de cet écosystème à forte orientation quantitative tende à contracter l'économie de l'interprétation. » (Emily Apter, *Against World Literature*, New York, Verso, 2013, p. 56).

L'histoire littéraire est-elle une histoire comme les autres ?

Comme d'habitude, les réflexions de F. Moretti et de son équipe sont revigorantes du point de vue de l'épistémologie littéraire. Elles offrent des moments salutaires de lucidité conceptuelle et méthodologique. Raisonner de manière digitale et computationnelle sur la matière littéraire suppose, comme les auteurs de ce collectif le font, de tirer profit des remarques de Weber, Kuhn, Popper, Pomian, Braudel, Leroi-Gourhan, Canguilhem, sur des points aussi divers et essentiels que ceux de la norme et de l'exception, de la consistance ontologique d'un fait scientifique, de la représentativité d'un échantillon, des rapports entre corrélation et causalité.

Une nouvelle conception de l'histoire littéraire émerge ainsi, portée par un esprit vérificationniste et expérimental, qui assume la nature explicative et hypothético-déductive du raisonnement littéraire. Les hypothèses sont faites pour être le plus souvent falsifiées et abandonnées ; la reconnaissance des échecs et la capacité autocritique ont cet avantage qu'ils nous prémunissent non seulement des hypothèses autoportées auxquelles on est tenté intuitivement de s'agripper, mais donnent aussi leur robustesse aux dernières hypothèses que nous sommes amenés à défendre comme des résultats solides et corroborés – en somme des faits scientifiques comme les autres.

Ce que les humanités numériques peuvent faire à la littérature est en ce sens comparable à ce que l'introduction de la longue durée a fait à l'histoire, en évacuant l'histoire événementielle et en minorant le rôle des grands hommes pour privilégier des questions comme l'évolution salariale, la fluctuation monétaire ou la pression démographique. De ce point de vue, Braudel et l'École des Annales constituent des références régulièrement convoquées dans les « plaquettes » (« *pamphlets* »).

F. Moretti remporte là une double victoire contre l'herméneutique traditionnelle : certes, et c'est déjà beaucoup, il congédie toute lecture littéraire qui ne serait attentive qu'à l'exception, l'écart ou la singularité au profit du repérage, dans le chaos et le bruit formés par les données, des régularités, des séries, des cohortes ou des *patterns*. Cependant, par une sorte de ruse tout à fait ironique, il parvient à accomplir les missions de l'herméneutique bien au delà des attentes de ceux qui se prévalent d'en être les gardiens. De la même manière que les sciences sociales peuvent comprendre les conduites humaines à des niveaux qui dépassent la simple compréhension qu'en ont les acteurs eux-mêmes, la critique littéraire digitale se détourne de la conception mythologique, éculée mais à tout le moins persistante, selon laquelle le sens d'une œuvre serait une affaire essentiellement individuelle et limitée à une poignée d'acteurs (l'auteur et/ou le lecteur). Elle se donne de la sorte les moyens de comprendre des œuvres mieux que les acteurs individuels n'en seraient capables à leur humble niveau, en en faisant un phénomène social complexe dont on rendrait observables des corrélations demeurant hors de leur portée.

L'équipe de F. Moretti entend ainsi œuvrer à une « histoire totale de la littérature » (p. 222). Mais que faut-il entendre par là ? D'abord, il ne s'agit pas ici d'une histoire mondiale de la littérature : puisant dans Braudel des recommandations méthodologiques, les travaux de F. Moretti n'ont étrangement jamais paru aussi loin de considérations sur la littérature mondiale ou sur des systèmes-mondes. Ensuite, c'est une histoire au demeurant assez internaliste, une histoire littéraire plus qu'à proprement parler une histoire de la littérature connectée à d'autres pans de l'histoire⁷.

La réalisation de ces expérimentations digitales dans le laboratoire littéraire, en plus d'être coûteuse en temps, incline insensiblement à deux défauts complémentaires. Celles-ci incitent à l'abstraction propre à une entreprise de modélisation et poussent sans doute à se concentrer sur des corpus romanesques du 19^e siècle et à perdre de vue, provisoirement et non définitivement, cette ample perspective mondiale. Elles aiguissent également une sorte d'attention d'ordre morphologique, qui tient le chercheur éloigné de la complexité sociohistorique des écosystèmes dans lesquelles les œuvres littéraires circulent et se transforment⁸.

C'est que, pour F. Moretti, les formes littéraires ont quelque chose de l'étalon que constitue un idéal-type wébérien :

autant de constructions mentales qui ne se trouveront jamais comme telles dans les œuvres singulières, mais que nous pouvons mobiliser pour mesurer les différences entre celles-ci, dans ce geste propre à la méthode comparative de Weber qui consiste à soumettre le cours imprévisible de l'histoire à la saisie unifiante des concepts. (p. 276)

De là, il est important de savoir « revenir des abstractions à l'histoire littéraire » (*ibid.*), sans perdre de vue les « *big questions* » (p. 17). C'est pourquoi il faut lire ce livre pour ce qu'il est : une mise au point *in progress* de protocoles d'analyse plus ou moins exportables et adaptables, et une stimulante invitation à construire des données et à élaborer des modes de traitement applicables à d'autres genres, d'autres cultures littéraires (moins européo-centrées), d'autres périodes (moins modernes).

Publié dans laviedesidees.fr, le 8 mai 2017.

⁷ Certes les auteurs affirment vouloir « ouvrir les romans et découvrir si leur destin social – popularité, prestige, les deux ou aucun – entretient un rapport quelconque avec leurs caractéristiques morphologiques » (p. 236) Avouons qu'il ne s'agit là pour les œuvres littéraires que d'un décloisonnement historique somme toute modeste.

⁸ Alexander Beecroft (*An ecology of World Literature*, Londres, Verso, 2015) privilégie par exemple la complexité des écosystèmes, dont il faut rendre compte de la complexité des interactions, à la simplicité modélisée des systèmes à la Moretti. Emily Apter le souligne également : « Alors que Moretti agrmente lui-même souvent son formalisme quantitatif de tournures spirituelles et d'hypothèses audacieuses sur le changement culturel et politique, les plaquettes du Literary Lab adoptent le style plat de la technique (...) Si les laboratoires annoncent le statut futur de la théorie des systèmes dans les études littéraires, ils se préoccupent moins des écosystèmes (et de leur applications organicistes, génétiques et évolutionnaires) que des systèmes computationnels tout court, mais mis au service de la littérature et de l'histoire sociale. » (*Against World Literature*, *op. cit.*, p. 55-56.)