



Revue Sciences/Lettres

2 | 2014

Épistémologies digitales des sciences humaines et sociales

Nouvelles frontières de l'historien

Stéphane Lamassé et Philippe Rygiel



Édition électronique

URL : <http://journals.openedition.org/rsl/411>

DOI : 10.4000/rsl.411

ISSN : 2271-6246

Éditeur

Éditions Rue d'Ulm

Référence électronique

Stéphane Lamassé et Philippe Rygiel, « Nouvelles frontières de l'historien », *Revue Sciences/Lettres* [En ligne], 2 | 2014, mis en ligne le 07 octobre 2013, consulté le 22 avril 2019. URL : <http://journals.openedition.org/rsl/411> ; DOI : 10.4000/rsl.411

Ce document a été généré automatiquement le 22 avril 2019.

© Revue Sciences/Lettres

Nouvelles frontières de l'historien

Stéphane Lamassé et Philippe Rygiel

- 1 Il n'est plus, dans nos sociétés, d'activité savante sans usage de l'ordinateur et des réseaux électroniques. Le fait n'est pas tout à fait nouveau, même pour les historiens. La découverte des possibles ouverts par l'usage de l'informatique par de petits cercles pionniers est ancienne¹ et la première trace d'une activité historique sur la toile remonterait à 1993. Cela attira peu alors l'attention de la majorité des praticiens des disciplines érudites et particulièrement des historiens. Ils tendirent longtemps à tenir l'informatique pour une technique ancillaire, dont l'irruption ne bousculait pas les normes et les pratiques disciplinaires, et dont les spécialistes, en petit nombre, s'exprimaient en des revues spécialisées à l'audience modeste². La conjoncture s'est au cours des deux dernières années brutalement renversée. Plusieurs grandes revues ou collections ont consacré un numéro spécial ou un volumineux dossier au tournant numérique de l'histoire³, ou bien en préparent un⁴, des séminaires d'histoire numérique naissent dans les plus prestigieuses institutions, les blogs se multiplient⁵, la vénérable société des historiens modernistes et contemporanéistes a consacré sa dernière réunion annuelle à la question et plusieurs projets de recherches récemment lancés intègrent une forte dimension numérique⁶.
- 2 Si nous ne savons pas plus aujourd'hui qu'hier ce que les historiens font effectivement de leurs ordinateurs⁷, nous sommes soudain confrontés à une avalanche de textes programmatiques ou réflexifs et claquent comme des étendards de multiples et rutilants labels. À peine averti qu'il est l'heureux possesseur d'un ordinateur connecté à un réseau, voilà l'historien sommé de décider s'il se range parmi les tenants des humanités numériques (*digital humanities*), d'une histoire numérique (*digital history*) ou incline à se définir comme un historien pris en un contexte numérique, incité aussi à se demander s'il est déjà un historien 2.0, voire un expérimentateur audacieux inventant les pratiques qui seront celles de l'érudit 3.0⁸.
- 3 Cette soudaine effervescence est liée à la conscience, nouvelle, de ce que l'informatique, sous sa forme réseau, ne permet pas seulement d'accomplir un peu mieux ou un peu plus vite les mêmes tâches qu'autrefois, mais de ce qu'elle induit une mutation des conditions de production et de diffusion des connaissances historiques, voire de la nature de celles-

ci. La transformation la plus souvent mise en évidence est la possibilité nouvelle d'accéder immédiatement à de vastes réservoirs de données et d'articles. Le fait que beaucoup d'historiens, mais aussi de professeurs, peuvent aujourd'hui accéder à des textes et à des images numériques de documents d'archives utilisables dans le cadre de leurs recherches ou de leur enseignement est au principe de la prise de conscience récente de ce que l'informatique réseau n'est pas un simple perfectionnement de la machine à écrire et du téléphone.

- 4 La possibilité d'une communication immédiate et peu coûteuse entre spécialistes d'un même champ est souvent également soulignée, de même que celle d'une dissémination efficace des produits de l'activité historique par le moyen de la toile, parce que l'une comme l'autre apparaissent susceptibles de transformer tant la structuration de la profession historique que son rapport aux milieux proches (archivistes, industries culturelles, etc.) avec lesquels elle est en contact⁹.
- 5 Ces mutations, il est vrai spectaculaires, et d'autant plus vivement perçues qu'elles ont déjà transformé le quotidien des praticiens, sont cependant déjà depuis longtemps identifiées, au point d'avoir donné naissance à quelques discussions canoniques, tenant aux transformations supposées des pratiques de lecture (lecture profonde ou lente qui aurait été la norme autrefois et lecture flottante ou rapide qui le serait aujourd'hui¹⁰), à la démocratisation – notion particulièrement problématique quoique pas toujours interrogée¹¹ – de l'écriture de l'histoire, ou bien à l'éventuel triomphe des partisans d'une pensée graphique et réticulaire sur les tenants attardés d'une réflexion séquentielle et linéaire¹².
- 6 Ce ne sont pas donc ces matières que nous entendons discuter, non plus que les questions, moins fréquemment abordées par les historiens, mais tout à fait essentielles de l'archivage du monde numérique¹³, ou de l'apparition, par le biais par exemple du *crowdsourcing*, de nouveaux producteurs de récits historiques¹⁴.
- 7 Nous nous proposons ici de revenir sur les formes contemporaines de l'accumulation, sous forme numérique, de données utilisables par l'historien et certains des possibles alors ouverts en matière de structuration de l'information et de traitement de celle-ci. Nous nous attacherons particulièrement à la façon dont cela peut contribuer, pour un temps au moins, à redéfinir les échanges professionnels des praticiens de l'histoire et en particulier leur rapport aux représentants d'autres champs disciplinaires.
- 8 L'aspect le plus connu de ce processus, engagé depuis déjà de nombreuses années, est la numérisation des collections des grandes bibliothèques publiques et la mise à disposition de vastes ensembles textuels, sous formes d'abord de fichiers image, puis de plus en plus de fichiers texte par des opérateurs tant privés que public. La croissance est en ce domaine exponentielle. La seule BnF, qui mettait à disposition en 1997 environ 3 000 volumes, offre par l'intermédiaire du portail *Gallica* 90 000 ouvrages numérisés et 80 000 images en 2007, et environ 400 000 documents fin 2010¹⁵. Depuis peu, les institutions publiques ouvrent également l'accès à une quantité croissante de données d'enquêtes, issues souvent de la statistique publique et pour partie nativement numériques¹⁶. L'équipement DIME-SHS¹⁷ participe de ce mouvement et s'oriente, semble-t-il, vers de la mise à disposition de méthodes et d'enquêtes en Open data. De même nature est la mise à disposition de la littérature scientifique, récente ou moins récente, accessible au travers de portails dont l'accès souvent est payant. Dans tous les cas, l'utilisateur de ces ressources constate rapidement qu'il ne peut en faire usage sans que celles-ci ne soient documentées, et leur documentation accessible et manipulable, ce qui suppose

l'intervention des acteurs traditionnels que sont les archivistes, les bibliothécaires et les spécialistes de la documentation.

- 9 Plus récemment ont été lancées de grandes entreprises de numérisation de fonds d'archives¹⁸. En France celles-ci touchent plus particulièrement un type de sources que l'on appelait autrefois sérielles : enregistrement, cadastres, état-civil sont particulièrement susceptibles d'être disponibles sous forme numérique, et dans la plupart des cas accessibles par le biais d'internet. La majorité des archives départementales offrent ainsi en 2012 l'accès, selon des modalités différentes, à tout ou partie de l'état civil ancien par le moyen de documents numériques¹⁹. Ces plans de numérisation sont parfois plus ou moins directement liés à des entreprises d'édition scientifique qui renouvellent profondément les méthodes de celle-ci, mais aussi parfois ses catégories – l'idée d'une possible stabilité du texte médiéval par exemple²⁰ –. L'érudition est donc aujourd'hui déjà profondément numérique par les outils qu'elle emploie, mais aussi parce que ceux-ci peuvent être mis en lien avec une transformation des pratiques et modes de pensée constitutifs de ses disciplines.
- 10 Moins notées, parce qu'encore souvent le fait d'entreprises pilotes, sont les recherches s'accompagnant de dispositifs, au moins partiellement automatisés, permettant la production de bases de données structurées rassemblant des descripteurs des objets numérisés, ainsi parfois qu'une transcription des éléments textuels inscrits en ceux-ci, voire une explicitation de certaines des informations portées par les éléments symboliques incorporés au document, dynamiques dont témoignent plusieurs exemples.
- 11 Il est aujourd'hui possible ainsi de nourrir un système de gestion d'information au moyen d'une océrisation partielle de documents portant des inscriptions manuscrites, pour peu que ceux-ci soient structurés ou semi-structurés, ce qui est le cas de nombreuses sources utilisées par les historiens du contemporain (listes, dossiers, formulaires). Des prototypes sont élaborés dès le début des années 2000²¹ et des systèmes de ce genre sont aujourd'hui utilisés de façon routinière par les entreprises et les administrations²².
- 12 Nous pouvons également caractériser des documents anciens comportant textes et figures au moyen de paramètres calculés, sans connaissance *a priori* de la structure du document ni de son contenu sémantique, c'est-à-dire produire, de manière semi-automatique, une information structurée décrivant ceux-ci et les images (lettrines, enluminures, dessins²³) ou blocs de textes qui s'y trouvent, ce qui débouche sur la possibilité d'une indexation partiellement automatisée de ces matériaux et d'une analyse de leurs propriétés physiques. Cela autorise la mise au point d'algorithmes d'attribution, voire, par la détermination de classes d'objets, permet la reconnaissance de symboles et donc l'extraction d'informations, d'où la possibilité de générer des systèmes d'information associés à des documents de type carte ou plan²⁴. Ces outils sont encore des dispositifs de laboratoire, dont l'usage suppose un paramétrage fin opéré par des spécialistes au bagage mathématique conséquent (transformations de Fourier, chaînes de Markov, fractales sont parmi les outils les plus utilisés pour l'analyse semi-automatisée de documents complexes²⁵), ils n'en sont pas moins appelés à se développer rapidement, du fait même de l'ampleur des programmes de numérisation qui rendent indispensables la création d'outils autorisant l'exploitation des documents disponibles dont la masse croissante rend impossible l'indexation manuelle²⁶.
- 13 Sont également en cours de développement des outils permettant la description et l'indexation des flux vocaux tels ceux capturés par les émissions de radio ou les émissions de télévision à partir de la reconstitution automatisée d'un *verbatim*, en attendant sans

doute, progrès des capacités de calcul promises par les ordinateurs quantiques aidant, la naissance d'outils similaires permettant de traiter des flux vidéos conséquents. L'équipex Matrice ainsi, porté par Denis Peschanski, a pour premier objectif la production d'outils efficaces permettant, à partir de sources sonores très bruitées et hétérogènes, d'améliorer la qualité des dispositifs de reconnaissance vocale, ce qui autorisera, à terme, la transcription semi-automatique de vastes réservoirs d'archives orales ou des bandes sons de flux vidéos.

- 14 La particularité de ces évolutions rapides est de n'être pas liée seulement à la purification d'algorithmes de plus en plus efficaces, non plus qu'au raffinement toujours plus grand des outils mathématiques utilisés, mais de dépendre de l'extraordinaire croissance des capacités de calcul décrite par la loi de Moore et de l'augmentation, non moins stupéfiante, des capacités de stockage disponibles, ainsi que de celle de la rapidité du transfert des données, qu'accompagne la très rapide diminution des coûts de ces différents éléments.
- 15 Ce contexte a d'autres conséquences. L'accumulation de masses d'informations utilisables par les historiens se produit alors que se démultiplient les possibilités de manipulations distribuées de vastes amas de données, et que diminuent les coûts d'appropriation d'un certain nombre d'outils logiciels, sous l'effet, en partie, du développement de solutions logicielles libres exploitant à plein la souplesse du travail en réseau.
- 16 Pour ne prendre que quelques exemples, chacun peut aujourd'hui facilement disposer sur son ordinateur personnel de systèmes de gestion d'information performants et aisément mutualisables (Evernote par exemple ou Zotero, plus connu des historiens) permettant l'intégration de documents hétérogènes et leur indexation à la volée, alors qu'il fallait, il y a vingt ans, construire pour cela une base de données relationnelle, élaborer, souvent *a priori*, une arborescence de descripteurs et doter, si l'on voulait partager les données ainsi produites, chaque poste d'utilisateur d'une coûteuse licence. De même fallait-il, pour réaliser une simple analyse factorielle, voire obtenir rapidement une série de tableaux croisés, écrire une macro-commande, ou être capable de manipuler des logiciels professionnels dont l'ergonomie ne semblait remarquable qu'aux adeptes, rares, du langage Fortran.
- 17 Ces mutations, qui apparaissent particulièrement rapides à l'échelle d'une carrière de chercheur, sont plus spectaculaires encore quand sont considérées des pratiques qui demeurent l'apanage de cercles plus étroits de spécialistes. L'analyse textométrique des corpus anciens a ainsi récemment subi de rapides et profondes mutations techniques et conceptuelles²⁷. L'analyse de discours se déploie aujourd'hui dans toutes les périodes²⁸ et des techniques comme celle de l'étiquetage morphosyntaxique, qui, il y a peu, nécessitaient un temps considérable deviennent, en partie au moins, automatisables pour un grand nombre de langues contemporaines²⁹ et anciennes³⁰. Le lien avec la discipline historique est ici particulièrement fort : les moyens de traitements de la langue ont été investis par des historiens qui infléchissent la production logicielle³¹.
- 18 Il est donc aujourd'hui possible de concevoir des chaînes d'acquisition et de traitement du signal historique, qui permettent, partant d'une source, d'un fonds, de le décrire finement au moyen de documents numériques, d'en transcrire les inscriptions et de mettre en place des outils, accessibles aux historiens de métier, permettant une navigation efficace du tout et aussi, par des manipulations automatisées de ces répertoires de signes, d'en proposer des visualisations et des descriptions synthétiques aptes à nourrir des discours historiques tout à fait conformes aux canons de la discipline. En témoignent, par exemple

les nombreuses publications récentes étudiant le Paris ancien à l'aide des outils façonnés par le projet Alpage³².

- 19 L'histoire se dévoile alors mode de connaissance instrumenté dont les praticiens, aux prises avec une matérialité, forgent des dispositifs permettant de générer des traces graphiques interprétables. La pratique historique en d'autres termes assemble des techniques. Si la conscience de cela l'est parfois, la chose n'est pas nouvelle : les historiens du livre montrent depuis déjà longtemps que celui-ci est un objet extraordinairement complexe, dont l'évolution est liée à celle d'un secteur industriel important en même temps qu'indissociablement à des techniques de l'intellect que tout à la fois il incorpore et rend possible.
- 20 La nouveauté est que l'historien, en sa pratique quotidienne, est en contact direct avec des machines dont il ne peut saisir la complexité et que celles-ci permettent une automatisation d'une partie des tâches qu'il accomplissait sans le secours d'aucune, par exemple la production de listes, de cartes, de plans, la création d'index ou de tables. Changement d'échelle donc, de *modus operandi* et d'ordre de grandeur, qui n'abolit pas ce qui précède, mais dont les implications ne conduisent pas moins à une transformation profonde des pratiques, du métier et sans doute de la position sociale de l'historien, qui se faisait autrefois, s'il faut en croire du moins les textes normatifs d'hier, un devoir de contrôler toutes les phases de son travail, de la collecte à la synthèse, en passant par la description et l'évaluation de ses sources.
- 21 Celui-ci dépend, aujourd'hui comme hier, d'une infrastructure, c'est-à-dire d'outils, de machines, d'institutions, de procédures codifiées, du fonctionnement de secteurs industriels et l'informatique réseau dévoreuse de ressources naturelles, gourmande en énergie, particulièrement polluante et important secteur d'emploi est une industrie lourde. Elle a cependant ses propriétés propres. Elle incorpore un mille-feuille de textes et de langages informatiques, qui procèdent d'outils mathématiques dont parfois seuls les plus grands spécialistes du domaine avaient, il y a quelques dizaines d'années³³, notion, ce qui rend l'environnement au sein duquel se déploient nos pratiques savantes infiniment plus complexe qu'autrefois, au point qu'il n'est pas certain qu'un individu puisse encore avoir une intellection globale de ce système, ni même une représentation fonctionnelle à peu près assurée de celui-ci.
- 22 Le constat n'est pas propre à la discipline historique. Claude Kergomard le faisait il y a peu pour la géographie³⁴, dont certains secteurs sont tôt entrés à l'ère du laboratoire latourien peuplé d'étranges boîtes noires délivrant des signes dont les usagers routiniers n'interrogent plus et parfois ne connaissent plus, les principes de génération.
- 23 Ce n'est pas dire là que l'atelier de l'historien, créateur solitaire, est devenu ou a vocation à devenir pour tous et bientôt un laboratoire flambant neuf, où s'affairent des professionnels en blouse blanche. Nous savons, depuis au moins Bertrand Gille³⁵, que les systèmes techniques se recouvrent plus qu'ils ne se succèdent. La conjoncture nouvelle n'abolit pas l'ancien, elle s'y combine, l'agrège, en des installations parfois incongrues et toujours instables, car notre environnement informatique est mouvant, empêchant encore l'émergence de formes pérennes. Le plus bel exemple de ces dynamiques nous est fourni là encore par l'édition scientifique des textes anciens. Alors que les formes en apparaissaient définitivement stabilisées, elle est marquée depuis une quinzaine d'années par un bouillonnement créatif dont il n'est pas certain que le terme soit proche³⁶, de même la philologie très tôt confrontée à l'informatique et partie prenante souvent des entreprises d'édition électronique a-t-elle profondément renouvelé ses pratiques³⁷.

- 24 Simplement, pour emprunter au vocabulaire latourien³⁸, d'étranges boîtes noires trouvent leur place dans cet environnement de travail, dont certaines ont l'inquiétant aspect de très lisses monolithes. Internet foisonne aujourd'hui de traceurs, sites alléchants permettant la production automatique de graphes, de figures, d'images, offrant la possibilité à tous, et donc aussi aux historiens, de générer de fascinants objets graphiques, parfois plastiquement remarquables³⁹ et stimulant incontestablement l'imagination, qui permettent de visualiser des textes, des tableaux de données, des flux d'information, et d'accrocher à ces visualisations des discours, sans qu'il soit besoin pour cela d'aucune compétence informatique, non plus que d'aucune compréhension des opérations leur donnant naissance. Il en est beaucoup, mais le plus célèbre de ces outils au sein des communautés s'occupant d'analyse des textes est sans contexte Ngram, parce que le lancement de ce site a été accompagné d'une campagne visant à promouvoir une pratique nouvelle définie par ses promoteurs comme une science nouvelle, la *culturonomics*⁴⁰.
- 25 L'initiative a suscité rapidement des critiques justifiées⁴¹, le dispositif initial n'en demeure pas moins intéressant parce qu'il révèle certaines des dynamiques en cours et des futurs possibles.
- 26 Une entreprise privée, proposait, forte de sa position dominante, un dispositif défini comme savant qu'elle est seule, puisqu'elle contrôle l'accès à Google books, à pouvoir mettre en place. Celui-ci est au sens propre un traceur, produisant par manipulation réglée d'un vaste réservoir de données textuelles, en réponse à une requête de l'utilisateur, une trace graphique interprétable. L'instrument n'est conforme aux canons d'aucune discipline existante, ni même à certains des principes de la démarche scientifique, du fait en particulier des problèmes tenant à la définition des corpus de références. Il est proposé par une firme qui n'est pas un acteur traditionnel de la sphère savante et dont nous savons qu'elle a parfois permis ou favorisé la censure étatique, voire dont certains employés ont délibérément falsifié des données afin d'éliminer des concurrents⁴². Elle n'en a pas moins trouvé des alliés au sein des mondes savants et l'objet proposé demeure le seul mode d'interrogation possible du plus vaste corpus de données textuelles jamais rassemblé.
- 27 Nous retrouvons là une situation, par certains aspects, familière, comparable à la généralisation des tableurs qui conduisit, il y a quelques années, à la multiplication de camemberts dont les couleurs vives égayaient les pages des premiers mémoires de lettres et d'histoire entièrement composés sur ordinateur mais qui ne respectaient que rarement les règles de la sémiologie graphique, et moins encore celles de la statistique ; à ceci près, et la différence est de taille, qu'il était toujours possible de s'emparer d'autres outils ou d'en créer.
- 28 Si nous nous inscrivons dans la perspective d'une histoire attentive aux conditions d'administration de la preuve, qui se conçoit comme une entreprise de connaissance instrumentée et dont les praticiens, confrontés à un matériau opposant une résistance à leurs entreprises, mettent en œuvre des dispositifs techniques, alors la priorité du moment est celle de l'équipement de l'historien.
- 29 À cela la seule réponse possible nous paraît être l'élaboration d'une instrumentation savante par la communauté savante elle-même, qui garantisse l'inscription de ses normes au cœur des outils proposés, c'est-à-dire d'abord la possibilité de la critique et d'une appropriation créative. Pour le dire en reprenant là encore le vocabulaire latourien, il

faut que les boîtes noires puissent être ouvertes et qu'il soit possible de calibrer les instruments. Cela suppose, dans le domaine qui nous occupe, l'accès aux algorithmes et au code, la possibilité aussi de paramétrer finement les traceurs et les grapheurs appelés à se multiplier⁴³ et qui constituent de fait les outils permettant d'adapter au nouveau contexte les exigences de nos disciplines, en particulier pour l'histoire, celle de la prise en compte de la totalité du matériau pertinent disponible. Il ne s'agit pas cependant ici d'adaptation contrainte et à marche forcée. Ce contexte offre, à condition d'appropriation des outils nécessaires, outre une intensification de la circulation de l'information au sein de la communauté scientifique et de sa diffusion au-delà de celle-ci, un accroissement de nos capacités manipulatoires et donc de notre puissance cognitive⁴⁴.

- 30 La tâche est rude cependant et pour plusieurs raisons. La production, en ce contexte, d'une instrumentation adaptée aux besoins d'une communauté scientifique suppose l'alliance de compétences qui ne peuvent être réunies en une seule personne, ni même trouvées au sein des communautés scientifiques concernées. Des instruments permettant une manipulation pertinente et contrôlée de chaînes symboliques de grande taille supposent à la fois des capacités informatiques et des compétences mathématiques pointues. Aucun collectif d'historiens ne peut espérer les réunir.
- 31 Les techniques ainsi les plus prometteuses pour l'analyse exploratoire de vastes tableaux de données à forte proportion de données manquantes et comportant des données imprécises ou incertaines, ce qui est typiquement le cas pour de nombreuses tables de données produites au cours d'enquêtes historiques⁴⁵, sont aujourd'hui regroupées sous le nom de méthodes neuronales⁴⁶. Parmi celles-ci, l'algorithme de Kohonen (SOM) se révèle particulièrement adapté⁴⁷. Il offre la possibilité de visualiser un nuage de données multidimensionnelles à l'aide d'un document unique et, au prix d'un apprentissage, aisément lisible, de s'accommoder de données textuelles ou qualitatives, voire incomplètes (données manquantes ou données connues uniquement par des distances ou des indices de dissimilarité relative)⁴⁸.
- 32 De même, il existe aujourd'hui des techniques prometteuses en matière de traitement des données temporelles, mais dont l'appropriation là encore, et l'incorporation en un traitement logiciel supposent la mise en œuvre de compétences rares. Les chaînes de Markov cachées⁴⁹ se révèlent ainsi particulièrement adaptées à la production de systèmes de classification (donc de sériation et de périodisation) prenant en compte une succession d'événements ou d'observations.
- 33 Ce contexte invite donc l'historien à repenser, pour un temps du moins, le graphe de ses relations professionnelles. Le montage de chaînes d'acquisition de l'information implique une collaboration, habituelle, avec des archivistes, des bibliothécaires, mais aussi avec des physiciens, parce qu'elles supposent des compétences en matière de reconnaissance des formes et de traitement du signal. La production de grapheurs permettant des visualisations efficaces et pertinentes de vastes amas de données implique des échanges avec des mathématiciens, des spécialistes de la visualisation de données⁵⁰ et bien sûr des informaticiens. Nous écrivons ici, à dessein, collaboration et échanges et non usage d'outils ou de dispositifs produits par ces acteurs, moins encore prestation de service. Elle est en effet d'une nécessité vitale pour l'historien, plus exactement pour la communauté historique. Si le personnage collectif qu'est celui-ci veut demeurer acteur d'une chaîne de production de savoir, il ne peut accepter que la constitution des corpus, la structuration et la manipulation des données soient abandonnées à des prestataires extérieurs dont les logiques et les choix lui seraient totalement impénétrables. Ce serait

renoncer à la possibilité de toute critique procédurale et à l'appartenance aux mondes savants, pour n'être plus qu'un augure offrant une lecture incertaine de signes mystérieux dont seul le charisme, les phrases joliment tournées ou la fonction idéologique garantiraient le pouvoir de conviction.

- 34 Ces échanges sont possibles et de fait se produisent parfois, non pas seulement parce que les érudits ont besoin des scientifiques mais parce que les compétences de l'historien, et en particulier sa maîtrise des techniques de l'érudition en font un partenaire nécessaire. Le fait est particulièrement évident en matière d'élaboration de procédures en partie automatisées d'acquisition du signal et d'indexation des corpus qui présupposent une validation experte qui n'est possible qu'au prix de la maîtrise des déterminants de la matérialité de la source et de l'organisation des structures symboliques qu'elle porte.
- 35 Nous considérons qu'il en est de même en matière de structuration et d'exploration des données, fort du fonctionnement, depuis trois ans, de l'atelier analyse⁵¹, qui réunit autour d'un séminaire et d'un dispositif logiciel des compétences mathématiques, épistémologiques, historiennes et informatiques. La première raison en est que le rapport des praticiens des SHS à la donnée est en effet souvent spécifique, marqué, dans le cas des historiens, par la mise en place de boucles de rétroactions qui conduisent en permanence des résultats de l'analyse au réexamen des sources et à l'amélioration de celle-ci (par restructuration des données, tests de discrétisations ou de classifications nouvelles, production de variables supplémentaires). Une production logicielle adaptée aux recherches empiriques effectivement menées, ce qui est de fait la problématique des mathématiques appliquées aux sciences sociales, suppose donc là aussi un dévoilement des logiques d'action des chercheurs, qui ne peuvent être réduites d'ailleurs à des algorithmes, et donc des échanges entre praticiens.
- 36 D'autre part, la numérisation de nos sociétés et de nos vies quotidiennes produit une multiplication d'énormes gisements de données portant trace de comportements humains, dont s'emparent physiciens ou mathématiciens, soit qu'ils leur fournissent des objets complexes permettant d'imaginer ou de tester de nouvelles techniques de modélisation ou de description⁵², soit qu'il s'agisse, dans le but par exemple de mieux gérer un équipement, de décrire son usage, voire d'anticiper les conséquences d'une modification de certaines de ses propriétés⁵³.
- 37 Même cependant quand prime le souci de parvenir à une formalisation rigoureuse, les chemins suivis ont pour point d'origine « [...] *simple social assumptions which we show can be seen as responsible for the properties of the whole network*⁵⁴ ». En d'autres termes, même la modélisation la plus formaliste des traces d'un processus social suppose le recours à des hypothèses quant aux formes et aux déterminants de celui-ci, soit une théorie sociologique, de même que tout usage d'un dispositif comme *ngram viewer* suppose une théorie de l'histoire du livre et de la langue. Les praticiens des sciences sociales, et parmi eux les historiens, ont donc un rôle à jouer, en tant que tels, dans l'élaboration des nouveaux dispositifs de connaissance, d'autant que là encore la période est marquée par une activité intense portée par une forte demande d'intégration de la dimension temporelle et spatio-temporelle au sein des procédures d'analyses de vastes amas de données. Cela revient à incorporer à celles-ci une modélisation des systèmes de raisonnement et d'inférence géographiques et historiques⁵⁵. La collaboration entre historiens, mathématiciens et informaticiens peut en ce cas déboucher sur l'élaboration de systèmes de traitement de données temporelles ou séquentielles adaptés tant aux caractéristiques des données effectivement rencontrées qu'aux modes pertinents

d'interprétation de celles-ci, l'exploration de la série temporelle étant toujours encadrée par la connaissance, non nécessairement formalisée ou formalisable, d'un contexte général dont on cherche à déterminer la relation avec la chronique. L'étude ainsi de la production normative de l'État Piémonto-savoyard des XVI^e et XVII^e siècles menée par Julien Alerini en collaboration avec les informaticiens du CRI et les mathématiciens du Sann a permis l'écriture d'un *package* R implémentant un usage nouveau des chaînes de Markow cachées, qui s'est révélé particulièrement adapté à la description des productions normatives d'un État moderne en permettant une classification efficace autorisant à l'historien une sériation fiable de ses données⁵⁶.

- 38 La piste est prometteuse donc, mais le sentier escarpé. La participation à l'invention, en cours, de l'instrumentation historique de l'âge numérique, suppose des partenariats, qui exigent une *lingua franca*, qui, de fait, emprunte aux mathématiques et à l'informatique, parce que les dispositifs logiciels de traitement de l'information consistent en des manipulations réglées de chaînes de symboles dont l'écriture s'effectue en recourant à des langages informatiques. Cela exige, non pas que les historiens, en tant qu'individus, deviennent programmeurs ou mathématiciens, mais que la discipline fasse une place, qui ne soit pas un strapontin, à des praticiens dotés d'une solide culture en ce domaine. Cela passe également par une reconnaissance, en tant qu'acte de recherche, des contributions à l'élaboration de l'instrumentation scientifique.
- 39 Ces évolutions supposent aussi que les utilisateurs finaux des outils qui naissent aujourd'hui, et qui ne se substituent pas à ceux de l'historien d'hier, mais souvent s'y ajoutent, puissent se les approprier. Beaucoup ont la forme de traceurs et de grapheurs. Se saisir de leurs sorties et en proposer une interprétation pertinente et rigoureuse implique une familiarité avec les problématiques, travaillées depuis longtemps par les scientifiques, de la preuve graphique, mais aussi, parce que les images proposées proviennent toujours de transformations de traces au moyen d'opérateurs logiques et mathématiques, et non d'une disposition nouvelle de celles-ci, une acculturation minimale à ces domaines, et donc un enseignement adapté. C'est à ce prix que l'historien demeurera, au cours des quelques décennies qui viennent, un membre des mondes savants.

NOTES

1. *Mémoire vive*, <http://pireh.univ-paris1.fr/mv/>
2. Dans le domaine francophone, *Le Médiéviste et l'ordinateur* est un précieux témoin des réflexions et des pratiques pionnières.
3. *Revue d'histoire moderne et contemporaine*, 56/5, 2011 ; Jean-Philippe Genet et Andrea Zorzi, *Les Historiens et l'informatique un métier à réinventer*, Rome, École française de Rome, n° 444, 2011.
4. Ce sera le thème de l'un des numéros 2013 de la revue *Écrire l'histoire*.
5. Le plus connu est celui d'un historien américain, Dan Cohen, <http://www.dancohen.org/>
6. Par exemple Euroscientia : Localisation et circulation des savoirs d'État en Europe (1750-1850), projet franco-allemand (ANR-DFG), <http://ea3400.unistra.fr/index.php?id=12162>

7. Voir sur ce point Philippe Rygiel, « Écriture de l'histoire et réseaux numériques », in É. Guichard (dir.), *Regards croisés sur l'internet*, Villeurbanne, Presses de l'Ensib, 2011, p. 102-124.
8. Steven Mintz, en 2008, distinguait déjà les traits d'une histoire 3.0 et d'une histoire 4.0, coll., « Interchange : The Promise of Digital History », *The Journal of American History*, vol. 95, n° 2, 2008, en ligne.
9. James Grossman, « The Beckoning Horizon New Challenges for the AHA in the Digital Age », *Perspectives on History*, février 2012. Pour un exemple français de ces dynamiques on verra « Les boudoirs de l'historien », dispositif original de formation unissant historiens archivistes et documentaliste par leur rapport au numérique <http://www.enc.sorbonne.fr/actualite/seminaires-et-conferences/les-boudoirs-de-l-historien>
10. Christian Henriot, « Essai final », séminaire L'histoire à l'ère du numérique, 30 janvier 2012, histoirednumerique.blogspot.fr/2012/01/essai-final.html
11. Sur les transformations contemporaines des conditions d'accès à la documentation historique voir par exemple Roy Rosenzweig, « Should Historical Scholarship Be Free ? », *Perspectives on History*, 2005.
12. Alan Liu, « When Was Linearity ? : The Meaning of Graphics in the Digital Age », *Digital History Project*, août 2008.
13. Roy Rosenzweig, « Scarcity or Abundance ? Preserving the Past in a Digital Era », *American Historical Review*, n° 108, 2003, p. 735-762, également le dossier de *Matériaux pour l'histoire de notre temps*, n° 79, 2005, p. 11-13.
14. Serge Noiret, « Y'a-t-il une histoire numérique 2.0 ? », in J.-Ph. Genet et A. Zorzi (éd.), *Les Historiens et l'informatique. Un métier à réinventer*, Rome, École française de Rome, 2011, p. 235-288.
15. La Documentation française, *Les Bibliothèques numériques*, en ligne.
16. Voir, pour la France, l'action en ce domaine du réseau Quételet : <http://www.reseau-quetelet.cnrs.fr/spip/>
17. http://www.reseau-quetelet.cnrs.fr/spip/article.php3?id_article=220
18. Par exemple « Digital History and the emergence of digital historical literacies », in R. Diem et M. J. Berson (éd.), *Social Studies in the Information Age 1984-2009*, Information age publishing, 2010, p. 75-90. Pour la situation française on verra Yann Potin, « Institutions et pratiques d'archives face à la numérisation », *Revue d'histoire moderne et contemporaine*, 58-4bis, 2011, p. 57-69.
19. Archives de France, *Numérisation et mise en ligne des registres paroissiaux et d'état civil dans les services d'archives publics au 23 janvier 2012* (en ligne) et pour une cartographie des ressources <http://www.archivesenlignes.fr/>
20. Pierre Baudin et Catherine Jacquemard, « Les pratiques de l'édition en ligne. Expériences et questionnement », in J.-Ph. Genet et A. Zorzi, *Les Historiens et l'informatique*, op. cit., p. 87-108.
21. Nabel Murshed, « Automatic recording of Traffic Tickets », in D. Iopresti, J. Hu et R. Kashi (éd.), *Document analysis system V, 5th International Workshop DAS 2002*, Princeton NJ, USA, p. 66-70.
22. Céline Merlin, *Les Services publics en mutation, la poste innove*, Paris, L'Harmattan, 2006, p. 136.
23. Mickaël Coustaty, Jean-Marc Ogier, Rudolf Pareti et Nicole Vincent, « Extraction d'informations d'images de documents anciens », Congrès des jeunes chercheurs en vision par ordinateur, Trégastel, France, 2009, [hal-00435382].
24. Pour l'exemple le plus abouti à ce jour : Hélène Noizet et Éric Grosso, « The ALPAGE project : Paris and its suburban area at the intersection of history and geography (9th-19th century) », in *Digital proceedings of the 25th International Cartographic Conference (ICC'11)*, 3-8 juillet 2011, Paris, France, en ligne à http://alpage.tge-adonis.fr/documents/pub/aci_2011_ALPAGE_project.pdf
25. Nicholas Journet Jean-Yves Ramel, Véronique Eglin et Rémy Mullot, « Analyse d'images de documents anciens : une approche texture », *Traitement du signal*, vol. 24, n° 6, 2008, p. 461-479.
26. Pour un exemple du lien entre numérisation patrimoniale massive et développement logiciel on verra Mickaël Coustaty, Nicolas Sidere, Jean-Marc Ogier, Pierre Heroux et Jean-Yves Ramel,

« Content-Based Old Documents Indexing », *Eight International Workshop on Graphics Recognition*, GREC 2009, La Rochelle.

27. Jean-Marie Bertrand, Pierre Boilley, Jean-Philippe Genet et Pauline Schmitt-Pantel (dir.), *Langue et histoire*, Paris, Publications de la Sorbonne, 2012.

28. Benjamin Derruelle, « Enjeux politiques et sociaux de la culture chevaleresque XVI^e siècle : les prologues de chansons de geste imprimées », *Revue historique*, n° 655, 2010, p. 551-576.

29. L'étiqueteur Cordial pour le français ou le projet Treetagger : <http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/>

30. Le Projet SAS (signs and States), Lamop vise à créer un lemmatiseur pour le moyen français, l'anglais et le latin médiéval.

31. Damon Mayaffre, *Paroles de président. Jacques Chirac (1995-2003) et le discours présidentiel sous la V^e République*, Paris, Champion, 2004.

32. AnaLyse diachronique de l'espace urbain PARisien : approche GEomatique, <http://alpage.tge-adonis.fr/index.php/fr/>

33. La naissance même de l'informatique suppose des ruptures théoriques dans le domaine mathématique qui ne datent que des années 1930 et 1940.

34. Claude Kergomard, « Cartographier un objet variable dans le temps et dans l'espace : le climat », colloque Cartographie et épistémologie, Peyresq, 14-16 juin 2010.

35. Bertrand Gille, *Histoire des techniques*, Paris, Gallimard, « La Pléiade », 1978.

36. P. Baudin et C. Jacquemard, « Les pratiques de l'édition en ligne ... », art. cité.

37. Pour une synthèse récente en français, voir le blog d'Aurélien Berra : <http://philologia.hypotheses.org/>

38. Voir en particulier Bruno Latour, *La Science en action. Introduction à la sociologie des sciences*, Paris, La Découverte 2005, 1^{re} édition 1989.

39. Le fait n'est pas de hasard, l'expérimentation en ce domaine emprunte fréquemment à des dispositifs d'art numérique, secteur fort actif depuis déjà longtemps, voir Christiane Paul, *L'Art numérique*, Paris, Thames & Hudson, 2004, particulièrement le chapitre « Bases de données, visualisation de données et cartographie ».

40. Jean-Baptiste Michel *et al.*, « Quantitative Analysis of Culture Using Millions of Digitized Books », *Science*, DOI : 10.1126/science.1199644, en ligne le 16 décembre 2010.

41. Elise E. Morse-Gagné, « Culturomics : Statistical Traps Muddy the Data », *Science*, DOI :10.1126/science.332.6025.35-b, avril 2011.

42. « Google, moteur d'embrouilles », *Libération*, 23 janvier 2012.

43. Pour un des premiers exemples voir Éric Guichard, « L'atlasClio. Un atlas en ligne interactif de l'immigration », in Ph. Rygiel et S. Noiret, *Les Historiens, leurs revues et internet*, Paris, Publibook, 2005, p. 93-103.

44. Philippe Rygiel, « L'enquête historique à l'ère numérique », *Revue d'histoire moderne et contemporaine*, 58-4bis, 2011, p. 30-40.

45. Philippe Rygiel, *Destins immigrés, Cher 1920-1980. Trajectoires d'immigrés d'Europe*, Besançon, Presses universitaires franc-comtoises (PUFC), 2001.

46. François Blayo et Michel Verleysen, *Introduction aux réseaux de neurones artificiels*, Paris, PUF, 1996.

47. Kohonen, T., « Self-Organizing Maps », *Springer Series in Information Sciences*, vol. 30, 1995.

48. Benjamin Derruelle et Stéphane Lamassé, « Un dépouillement lexical entre analyse factorielle et cartes de kohonen : le vocabulaire mathématique des arithmétiques en langue vulgaire du bas moyen âge », *Modèles et apprentissages en sciences humaines et sociales (MASHS)*, université d'Aix-Marseille, 23-24 juin 2011, à paraître en 2013.

49. Leonard E. Baum et Ted Petrie, « Statistical inference for probabilistic functions of finite Markov chains », *Annals of Mathematical statistics*, vol. 37, n° 6, 1966, p. 1559-1563.

50. Sur ce champ en plein voir Jean-Daniel Fekete, « Visualiser l'information pour la comprendre vite et bien », in coll., *L'Usager numérique*, Paris, ADBS éditions, 2010.
51. <http://analyseshs.hypotheses.org/>
52. Remi Dorat, Matthieu Latapy, Bernard Conein et Nicolas Auray, « Multi-level analysis of an interaction network between individuals in a mailing-list », *Annals of Telecommunications*, vol. 62, n° 3-4, 2007.
53. Pierre Bornat, Céline Robardet, Jean-Baptiste Rouquier, Patrice Abry, Eric Fleury, et Patrick Flandrin, « Shared Bicycles in a City: A Signal Processing and Data Analysis Perspective », *Advances in Complex Systems*, vol. 14, n° 3, p. 415-438.
54. R. Dorat *et al.*, « Multi-level analysis of an interaction network ... », art. cité, p. 16.
55. Par exemple, Abdenour Mokrane, Oussama Laouamri, Gérard Dray, Pascal Poncelet, « Modélisation spatio-temporelle des connaissances d'un système d'information géographique », SETIT 2004, *International Conference: Sciences of Electronic, Technologies of Information and Telecommunications*, mars 2004, Tunisie.
56. Madalina Olteanu et James Ridgway, « Hidden Markov models for time series of counts with excess zeros », <http://hal.archives-ouvertes.fr/hal-00655588/fr/> preprint, décembre 2011.
-

RÉSUMÉS

L'impact de l'environnement digital sur les pratiques historiennes est généralement réduit à une transformation des conditions de diffusion des produits de l'activité historique. Nous montrons que le développement de nouvelles techniques de traitement des données a un impact sur la recherche historique qui a une certaine spécificité. Les données historiques sont rarement originellement numériques. La production de données adaptées à l'activité historique suppose la mise en place de plates-formes complexes dont l'élaboration suppose une collaboration avec des physiciens et des informaticiens. Les données produites sont souvent incomplètes et inégalement documentées, ce qui suppose un paramétrage fin des outils statistiques utilisés, ce qui implique des échanges avec des mathématiciens. Nous en concluons que cette configuration contribue à redessiner la carte des relations professionnelles des historiens.

The impact of digital technology on the research practices of historians is usually seen from the publishing end of their profession. We intend here to show that the development of new data processing methods can and does impact historical practices in specific ways. Digital historical data is rare and its production requires the creation of complex platforms with the collaboration of physicist and computer experts. The datas collected are often incomplete, and only partly documented requiring and the adjustment of statistical techniques, therefore a collaboration with statisticians. We argue that this configuration draws a new map of disciplinary alliances for historians.

INDEX

Keywords : history, digital, statistics, data mining, data acquisition, data visualization, archives

Mots-clés : histoire, numérique, épistémologie, quantification, traitement du signal

AUTEURS

STÉPHANE LAMASSÉ

Maître de conférences (Histoire, civilisation, archéologie et art des mondes anciens et médiévaux) à l'université Paris 1 Panthéon-Sorbonne.

Parmi les publications :

Avec C. Hug, C. Salinesi, et R. Deneckère, « Process modeling for Humanities : tracing and analyzing scientific processes », in *Annual Conference of Computer Applications and Quantitative Methods in Archaeology* (CAA 2011), avril 2011, Beijing, Chine, Amsterdam University Press, 2012, p. 245-255.

Avec J. Alerini, « Données et statistiques : l'avenir du travail en ligne pour l'historien », in J.-P. Genet et A. Zorzi (éd.), *Les Historiens et l'informatique : un métier à réinventer*, Rome, École française de Rome, 2011.

Avec M. Cottrel, B. Deruelle et P. Letrémy, « Lexical recount between Factor Analysis and Kohonen Map : mathematical vocabulary of arithmetic in the vernacular language of the late Middle Ages », *WSOM 2012, AISC 198*, 2012, p. 255-264.

PHILIPPE RYGIEL

Maître de conférences (Histoire et civilisations : histoire des mondes modernes, histoire du monde contemporain ; de l'art ; de la musique) à l'université Paris 1 Panthéon-Sorbonne et directeur adjoint du Centre d'histoire sociale du xx^e siècle (CNRS/Université Paris 1 Panthéon-Sorbonne).

Parmi les publications :

Avec S. Noiret (dir.), *Les Historiens, leurs revues et Internet*, Paris, Publibook Université, 2005.

« Écriture de l'histoire et réseaux numériques », in E. Guichard (dir.), *Regards croisés sur l'Internet*, Villeurbanne, Presses de l'Enssib, 2011, p. 102-124.

« L'enquête historique à l'ère numérique », *Revue d'histoire moderne et contemporaine*, 56/5, 2011.