



HAL
open science

Entretien croisé sur les “ Pratiques informatiques historiennes pionnières ” avec Jean-Philippe Genet, Jean-Luc Pinol et Philippe Rygiel

Edgar Lejeune

► To cite this version:

Edgar Lejeune. Entretien croisé sur les “ Pratiques informatiques historiques pionnières ” avec Jean-Philippe Genet, Jean-Luc Pinol et Philippe Rygiel. 2023, <https://journals.openedition.org/framespa/14176>. 10.4000/framespa.14176 . halshs-04153325

HAL Id: halshs-04153325

<https://shs.hal.science/halshs-04153325>

Submitted on 14 Apr 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0
International License

Entretien croisé sur les « Pratiques informatiques historiennes pionnières » avec Jean-Philippe Genet, Jean-Luc Pinol et Philippe Rygiel

Entretien réalisé le mardi 16 mai 2023 par Edgar Lejeune avec Jean-Philippe Genet, Jean-Luc Pinol et Philippe Rygiel.

Edgar Lejeune



Édition électronique

URL : <https://journals.openedition.org/framespa/14176>

DOI : 10.4000/framespa.14176

ISSN : 1760-4761

Éditeur

UMR 5136 – FRAMESPA

Ce document vous est offert par Université Toulouse 2 - Jean Jaurès



Référence électronique

Edgar Lejeune, « Entretien croisé sur les « Pratiques informatiques historiennes pionnières » avec Jean-Philippe Genet, Jean-Luc Pinol et Philippe Rygiel », *Les Cahiers de Framespa* [En ligne], 42 | 2023, mis en ligne le 20 juin 2023, consulté le 13 juillet 2023. URL : <http://journals.openedition.org/framespa/14176> ; DOI : <https://doi.org/10.4000/framespa.14176>

Ce document a été généré automatiquement le 13 juillet 2023.



Creative Commons - Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International
- CC BY-NC-ND 4.0

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Entretien croisé sur les « Pratiques informatiques historiennes pionnières » avec Jean-Philippe Genet, Jean-Luc Pinol et Philippe Rygiel

Entretien réalisé le mardi 16 mai 2023 par Edgar Lejeune avec Jean-Philippe Genet, Jean-Luc Pinol et Philippe Rygiel.

Edgar Lejeune

EDGAR LEJEUNE : *Jean-Philippe Genet, Jean-Luc Pinol, Philippe Rygiel, vous êtes tous les trois des historiens que l'on pourrait qualifier de « numériques », et il me semble que ces pratiques remontent pour chacun d'entre vous au début de votre carrière universitaire. Pouvez-vous nous raconter comment s'est déroulé cette rencontre avec l'informatique et les ordinateurs ?*

Jean-Luc Pinol : Avant ma thèse d'État, inscrite avant même que je ne passe l'agrégation, on m'avait conseillé de faire un troisième cycle, que j'avais décidé de faire sur les élections du Front populaire à Lyon. J'avais remarqué qu'on disposait des résultats de plus d'une centaine de bureaux de vote de l'agglomération lyonnaise, et donc mon idée c'était de développer une approche spatiale. En 1973, j'avais eu la chance de rencontrer un ingénieur informatique à l'université Lyon 2 qui m'avait montré des cartes pré-perforées à 40 colonnes. Et donc je suis allé dépouiller les listes électorales en faisant un sondage : je prenais une personne sur 10 sur les listes électorales de 1936 et j'inscrivais leur adresse sur la carte pré-perforée. Dans un deuxième temps, avec l'adresse, j'allais les rechercher dans les listes nominatives du recensement de Lyon en 1936. Donc vraiment tout de suite, pour faire ma thèse, j'ai été obligé d'utiliser l'informatique, parce que dépouiller à la main, ça n'était pas possible. J'ai appris par la suite que ce qu'on faisait pour moi au service informatique de Lyon 2, qu'on n'appelait peut-être pas encore informatique d'ailleurs, dépendait de machines qui étaient à l'autre bout de Lyon. Ça va d'ailleurs avoir des

conséquences dans mon cas puisqu'au moment de finaliser mon diplôme de 3^e cycle la foudre va couper la liaison...

Jean-Philippe Genet : Pour ma part j'ai commencé à m'intéresser à l'informatique pendant l'année universitaire 1967-1968. Mon nouveau sujet de thèse portait sur les idées sociales et politiques en Angleterre et je me suis rendu compte que j'allais en avoir besoin pour faire de la lexicométrie, sur des textes, et des bases de données sur des auteurs. Je connaissais ça à travers les travaux d'Antoine Prost, que j'ai vu à ce moment-là à la Sorbonne où j'étais moi-même assistant, et j'ai pu parler avec Édouard Cadet, qui s'occupait de l'informatique pour William Serman et Louis Girard dans le cadre de la recherche qu'ils menaient sur la chambre des Députés en 1837-1838. Dès ce moment-là, j'ai voulu me former à l'informatique. J'avais commencé de façon un peu amateur en compagnie de Guy Bois, mais en 1970-1971, j'ai pris systématiquement des cours d'informatique et de statistiques auprès de Marc Barbut pour être capable de faire moi-même le travail.

Assez rapidement, j'ai pu passer à l'œuvre. D'abord, en préparant moi-même des textes médiévaux anglais, et en les traitant au laboratoire de Saint-Cloud auquel m'avait donné accès Robert-Léon Wagner. Mais je me suis vite heurté à la question des moyens : il fallait des vacances pour saisir les textes sur cartes perforées, tâche ô combien chronophage, puis pour accéder à l'ordinateur du CIRCE à Orsay. C'est pourquoi il a fallu demander le secours du CNRS, grâce à une aide individuelle, demandée sous l'égide de Bernard Guenée. À l'époque c'était très simple : on recevait un carnet de chèque avec un montant, 100 000 francs (anciens) il me semble, ce qui n'était pas grand-chose, et on pouvait pratiquement utiliser cela que pour payer les heures calculs au CIRCE. L'aide est ensuite devenue équipe de recherche associée et enfin laboratoire associé. Mais grâce à elle, on a pu faire nos premières armes dans le domaine du traitement des textes et dans celui de l'analyse prosopographique avec SPSS qui nous permettait, à nous historiens, de nous transformer en sociologues. Il fallait seulement charger les bacs de cartes perforées pour aller le week-end à Orsay, où on attendait interminablement les résultats des traitements en batch, la moindre virgule oubliée dans une commande obligeant à un nouveau passage. C'était très rudimentaire mais enfin, ça nous a permis d'avancer. Mais on utilisait des logiciels qui n'étaient pas très utiles. Et c'est ce qui m'a poussé alors à aller plus loin, à essayer de développer un logiciel de lexicométrie adapté à l'absence totale d'orthographe dans les langues médiévales, ALINE, avec l'aide de deux brillants informaticiens, Jacques Mondelli et Jean-Paul Hucher, dont j'avais fait la connaissance grâce à Edouard Valensi, qui dirigeait le centre de calcul que nous avions à Paris 1.

Et puis comme pour la lexicométrie on a commencé à développer des logiciels spécifiques pour les bases de données. On n'a trouvé quelque chose de plus maniable que SPSS et surtout adapté à l'ordinateur qu'à partir de 1969, on a eu à disposition à Paris 1, c'est-à-dire le Philips P880. Xavier Debanne, un étudiant en médecine en rupture de ban et passionné d'informatique, développait pour le géographe Jean-Paul Trystram un logiciel de base de données au centre de calcul de Paris 1, BDP4, qui nous libérait des lourdeurs de SPSS. Et ça a inauguré une première phase qui a permis quand même de pas mal travailler, pour nous, médiévistes, et qui a duré jusqu'à ce qu'on ait accès au micro-ordinateur, non sans mal. Et l'accès au micro-ordinateur a provoqué une rupture, parce que les outils qu'on avait commencé à mettre en place pour les grosses machines se sont trouvés inutilisables, et il a fallu recommencer

toute la procédure avec les microordinateurs, puis avec internet et le travail en ligne, et peut-être bientôt encore avec l'intelligence artificielle...

Philippe Rygiel : De mon côté, le premier contact avec l'informatique n'a pas grand-chose à voir avec le métier d'historien. Au début des années 80, il y a déjà des ordinateurs chez moi, qui sont des Thompson, et j'ai aussi un certain nombre de petits camarades qui jouent avec des Apple II ou des Sinclair. Donc d'une certaine façon, ça fait partie très tôt des choses dont je connais l'existence, dont je sais qu'il faut les programmer si on veut en tirer quelque chose. Le moment où ça rentre dans ma pratique historique, c'est un peu plus tard, au milieu des années 80. Il y avait une espèce de guerre de religion à l'époque Mac-PC, et j'étais un peu par défaut dans le camp des Mac parce qu'en tant qu'élève à l'École Normale Supérieure j'avais accès librement à ces ordinateurs-là. Et donc j'ai acheté un Mac Plus parce que c'était ce qui permettait aussi de travailler sur les machines de l'École. Je me souviens encore du processeur Motorola 68000.

En fait j'ai travaillé d'abord dans une optique qui était très traditionnellement celle d'une histoire sociale quantitative. Je me souviens que dans un colloque j'avais dit que je faisais du Pinol 2.0, parce que c'était un peu la même chose, à coup de base de données, d'analyses statistiques de données sérielles. Ce qu'on appelait autrefois – alors le terme est moins utilisé aujourd'hui – des données sérielles, c'est à dire des données qui provenaient des recensements, de l'état civil, d'un certain nombre de sources décrivant beaucoup d'individus mais des informations très pauvres sur chaque, et où l'une des difficultés c'est de raccorder les individus entre eux quand ils apparaissent dans plusieurs sources. C'est un problème que Jean-Philippe connaît bien et qui est commun à tous ces usages-là.

Sur cette machine, très vite il y a un tableur. Tout ce qui est graphiques, tableaux, on le fait avec Excel, et tout ce qui est gestion de l'information, je le faisais avec une des premières SGDB qui tourne sur un micro-ordinateur qui est 4D. C'est à la fois adapté, parce que comme l'information est pauvre, on n'a pas forcément besoin d'avoir 300 000 tables. Et c'est aussi assez économe en ressources, à une époque où le gros problème c'est de pas surcharger soit la mémoire vive, soit la mémoire de masse, parce que ce sont des denrées assez rares et précieuses. Mais du coup par rapport à Jean-Philippe ou Jean-Luc, ça veut dire que je rentre dans un moment qui est déjà celui de la micro-informatique. C'est à dire que je n'ai pas le passage par les gros serveurs ou par les cartes perforées.

EDGAR LEJEUNE : Jean-Luc Pinol, Jean-Philippe Genet, justement, comment s'est fait ce passage à la micro-informatique dans vos expériences respectives ?

Jean-Philippe Genet : Alors là, c'est une nouvelle phase pour nous. Je me souviens que j'ai acheté mon propre Apple bien avant que je puisse en avoir un au laboratoire. J'avais 40 K de mémoire : pour le traitement de textes, on avait droit à seulement quatre caractères accentués ! et il fallait rentrer une disquette pour charger le logiciel, l'enlever, et en mettre une seconde pour taper le texte. J'ai ensuite acheté successivement, toujours de ma poche, deux ordinateurs PC Tandon qui avaient une forte capacité de mémoire bien avant que je puisse en avoir un au laboratoire et permettaient d'utiliser Word4 et DBase. Mais le problème des logiciels est devenu aigu. À Paris 1, Xavier Debanne a travaillé beaucoup pendant un temps à adapter les logiciels qui étaient disponibles pour les micro-ordinateurs. Alors pour les bases de données, on a commencé un logiciel qui s'appelait PROSOP, qui a donné quelques

résultats, mais malheureusement très vite, malgré le fait qu'on ait gravi les échelons du CNRS, en devenant une équipe de recherche associée, puis un laboratoire, on n'a jamais eu les moyens pour avoir véritablement un informaticien capable de développer ces logiciels. La première fois qu'on a présenté cette demande pour notre équipe c'était au début des années 90, et on a finalement obtenu un informaticien seulement au début des années 2000. Donc voilà, on a été de ce point de vue là, très en difficulté. Donc, on a quand même beaucoup travaillé, avec, essentiellement pour les bases de données prosopographique, DBase III, DBase IV et finalement FoxBase, dont la dernière version avait des performances très supérieures à celles de l'Access de Microsoft. Pour les traitements de textes, on est restés fidèles aux grosses machines pendant très longtemps, en travaillant avec notre logiciel qui s'appelait ALINE, jusqu'à ce que le programme ERC, beaucoup plus tard, en 2008, me permette de faire développer des logiciels spécifiques. D'une part, une adaptation de PROSOP, qui a permis de développer la base de données *Studium* et puis d'autre part, pour l'analyse lexicale, un logiciel qui a très bien marché jusqu'à il y a à peu près deux ans. Mais là aussi on a eu des problèmes avec le recrutement de l'informaticien par le CNRS.

Jean-Luc Pinol : Pour moi, effectivement, je travaillais au départ avec des gros serveurs, mais je ne connaissais pas les logiciels comme SPSS. C'est le service informatique de Lyon 2 qui me produisait des tableaux, des listings. Quand j'ai acheté un Macintosh en novembre 1984, avec mes deniers, ce qui m'intéressait c'était surtout la carte graphique et le fait de pouvoir enfin faire des cartes lisibles. Et donc je fais ma thèse d'état sur Macintosh. Pour les analyses factorielles, je les programme en BASIC. Au début, elles tournent doucement, puis bon ça s'accélère. Mais pour vous donner une idée, je suis à ce moment-là dans une UMR du CNRS, alors je ne sais pas si c'est UMR à l'époque, ou unité associée, enfin peu importe. On est en 1986 et il faut, pour qu'on achète des machines, que je fasse un argumentaire pour la délégation régionale du CNRS, pour leur montrer que des historiens peuvent avoir besoin d'ordinateurs. Ce n'est pas du tout, du tout évident pour eux. On n'en est pas à dire, on a besoin d'informaticien, non, non. Pour comparer, je vais enseigner après 1984 trois années de suite à Brown University dans le Rhode Island, où je fais des séminaires où j'utilise des micro-ordinateurs. Les premiers cours d'informatique, enfin, les premiers enseignements que je fais avec des ordinateurs, c'est à Brown, parce qu'à Lyon, il n'y a pas de salle informatique. À Brown, je me souviens, je passe un coup de téléphone et j'ai une salle avec 30 Macintosh pour faire de la cartographie sur ordinateur avec mes étudiants.

EDGAR LEJEUNE : Cette anecdote illustre bien les problématiques matérielles que pouvaient rencontrer les utilisateurs des ordinateurs en France avant les années 1990 et donc, d'une certaine façon, les problématiques institutionnelles. Pour tous les trois, comment ces décisions d'utiliser des ordinateurs pour vos recherches sont-elles reçues par vos collègues et vos institutions à ce moment-là ?

Jean-Philippe Genet : Je peux donner deux éléments de réponse. Tout d'abord en ce qui concerne le CNRS, où l'on a pu sentir une forme de refus de la micro-informatique, parce qu'il était évident que le CNRS voulait qu'on conserve, nous, historiens, ou « littéraires » d'une façon générale, nos crédits calcul CNRS pour alimenter l'ordinateur d'Orsay au CIRCE. C'était une sorte de chasse gardée, et donc, acheter des micro-ordinateurs, c'était priver, détruire même le circuit qui permettait à l'argent du CNRS de retourner au CNRS. Donc il y avait une forte prévention, comme l'a dit Jean-Luc, contre l'utilisation des micro-ordinateurs dans les équipes.

Au niveau de l'université par contre, c'était plutôt une ignorance complète. Tout a changé le jour où notre collègue Jean-Pierre Bardet a été nommé au cabinet du ministre des universités, en 1984 ou 1985, je crois. Il connaissait par sa propre expérience de chercheur l'intérêt de l'informatique pour les historiens et il a envoyé à Paris 1, pour l'UFR d'histoire, dix ordinateurs Bull. Robert Fossier, qui était le directeur de l'UFR, ne savait pas trop quoi en faire. Certains professeurs lui ont dit qu'il fallait en distribuer un à chacun : heureusement, ce n'était pas possible puisqu'il y avait trop peu d'ordinateurs. Fossier a hésité et j'ai insisté pour garder huit des dix machines et les mettre en bloc dans une salle que nous avons aménagée à la cave de la Sorbonne ! C'est comme ça qu'on a installé une salle d'informatique, qui existe toujours, et organisé la première formation d'informatique pour les étudiants d'histoire à Paris 1, en 1986 ou 1987. Et cet enseignement continue, aujourd'hui obligatoire alors qu'il n'était qu'optionnel. Nous avons pu recruter des PRAG, d'abord un pour les statistiques et un pour l'informatique (Giulio Romero) puis, au fur et à mesure que les besoins croissaient, cinq autres PRAG. L'un d'entre eux, Stéphane Lamassé, est devenu maître de conférences et dirige le groupe (le PIREH) qui est hébergé au sein du LAMOP. Aujourd'hui, on a à l'UFR d'Histoire à Paris 1 une équipe de six jeunes historiens qui font uniquement de la formation informatique (et aux statistiques, à la lexicométrie, etc.) pour les historiens au sein du PIREH. Presque tous ont terminé ou terminent leur thèse et certains ont été recrutés par d'autres universités.

Jean-Luc Pinol : Je me rends compte que mon expérience ressemble à celle de Jean-Philippe. Après ma thèse d'État, que je soutiens en 1989, je deviens professeur d'histoire économique et sociale à Strasbourg et je monte un centre de recherche sur la ville. Comme je suis un jeune professeur nouvellement nommé, et que Maurice Garden, que je connais, est au ministère, il m'indique : « N'oublie pas que tu peux avoir des aides comme jeune professeur qui développe la recherche ». C'est dans ce cadre-là, en lien avec le fait que je monte un centre de recherche, que j'achète des ordinateurs, des Macintosh d'ailleurs, et que je les mets à disposition des étudiants pour qu'ils puissent travailler, écrire, etc. Mais mes collègues me prennent vraiment pour un fou, parce qu'à ce moment-là on ne donne pas ce genre de machine à des étudiants.

Philippe Rygiel : Si je rebondis un peu sur cette question-là, ma position est différente, parce qu'en fait mes pratiques informatiques débutent quand je suis étudiant. C'est-à-dire que je ne suis pas du tout dans la même position institutionnelle que Jean-Philippe ou Jean-Luc dans le début de cette histoire. Je fais une maîtrise, un DEA, puis une thèse. Donc d'une certaine façon, ce que pensent l'institution et mes collègues, je n'en sais rien et ça ne m'a jamais beaucoup intéressé. Je sais que j'en ai besoin pour aller au bout de ce que j'ai envie de faire, et, du coup, dans le contexte, je ne vais même pas demander à l'institution des ressources informatiques, je vais les acheter. J'achète le premier ordinateur Mac portable, qui est un machin qui pèse 7 kilos, parce qu'à l'époque j'enseigne en Australie et que ça me permet de le balader. Je me souviens qu'il coûte 45 000 francs, ce qui à l'époque est un coût prohibitif. Mais j'en ai besoin pour travailler parce que je suis toujours sur de la manipulation de sources sérielles et que de toute façon, l'institution ne va pas me donner de machine. Elle ne va pas non plus me donner accès à un ingénieur parce qu'il n'y en a pas. En plus, je suis inscrit en thèse à Besançon avec Janine Ponty, et il

n'y a juste pas la possibilité d'avoir accès à des ressources de ce type. Donc ça veut dire que d'abord il faut se l'acheter, et que lorsque j'ai besoin de programmer, en fait, je programme moi-même.

EDGAR LEJEUNE : *Dans ce contexte où les ordinateurs sont difficiles d'accès, et où vos collègues y sont d'ailleurs parfois hostiles, quelles sont les raisons qui motivent votre choix d'avoir recours à ces instruments ?*

Jean-Philippe Genet : Je dirais d'un mot très simple, c'est pour compter, pour mesurer. Moi je suis médiéviste, alors traditionnellement, tout le monde sait que pour le Moyen Âge on ne peut pas mesurer, on ne peut pas compter, puisque les données sont forcément incomplètes et que de toute façon on n'a pas les belles sources sérielles dont disposent les modernistes ou les contemporanéistes - sauf exception bien sûr comme le *catasto* Florentin de Christiane Klapisch-Zuber et David Herlihy. Bon, en réalité, quand on construit des bases prosopographiques, par exemple, on s'aperçoit que, bien sûr, on a des données incomplètes, mais qu'on peut grâce aux traitements statistiques arriver tout de même à dégager beaucoup de choses. Et puis évidemment, la lexicométrie est un outil qui est à la disposition des médiévistes, comme il est à la disposition des autres.

Donc très vite en travaillant sur l'histoire des idées, je me suis aperçu que le langage était une chose fondamentale, et d'autre part que le statut socio-culturel des auteurs était également essentiel. Donc les deux choses étaient liées. Aujourd'hui, quand je travaille sur les étudiants et les maîtres de l'Université de Paris, bien sûr mes données ne sont pas complètes, mais je travaille quand même pour le moment sur une population qui comporte déjà 23 000 personnes. C'est quand même déjà un ensemble sur lequel on peut voir se dégager beaucoup de choses, par exemple pour les origines géographiques, l'origine sociale ou la carrière des futurs étudiants, mais aussi pour l'importance relative des différentes facultés, ou pour le poids des laïcs, des clercs ou des étudiants appartenant aux ordres religieux. C'est la même chose pour les auteurs dans le domaine de l'histoire et du politique dans le Moyen Âge anglais. J'ai quand même construit une base de données qui comporte près de 2500 personnes, dont peut-être 150 ou 200 étaient connus comme historiens. Mais quand on gratte un peu, on retrouve beaucoup d'auteurs, beaucoup d'activités.

Et c'est là aussi où j'ai découvert un autre des intérêts fondamentaux de l'informatique, qui est la nécessité de définir ces variables quand on fait de l'histoire sociale. Ce qu'on appelle aujourd'hui des ontologies, c'est un mot peut-être un peu prétentieux, parce que ça a étendu de façon extraordinaire la notion de variable, mais c'est bien ça qui est en jeu. Pour faire un travail statistique sur les historiens et les auteurs politiques au Moyen Âge, comme la base HISPOL développée pour ma thèse, on s'aperçoit qu'il faut définir ce que c'est que l'histoire pendant la période médiévale, il faut définir ce que c'est que le concept d'écrivain politique, et c'est par là que de fil en aiguille, j'en ai été amené personnellement à travailler sur la notion de champ chez Bourdieu. Là encore, c'est un élargissement des perspectives qui est considérable pour les historiens. Alors avec mes collègues, on a la même expérience que je sache.

Jean-Luc Pinol : En 1973 quand je commence à utiliser les cartes pré-perforées, ce que je veux faire sur cette centaine de bureau de vote Lyonnais en 1936, je n'envisage pas de pouvoir le faire autrement, d'autant plus que je veux faire des cartes. Je constitue un fichier par sondage de 15 000 électeurs répartis sur l'agglomération

lyonnaise et qui sont dans chacun des bureaux de vote, c'est-à-dire qu'en gros, il y a en moyenne 150 électeurs par bureau de vote. Et puis je fais des cartes avec Didier Soum, le responsable du service informatique, sans carte graphique, en empilant les lettres, ce qui se faisait à l'époque. C'est compliqué, ça demande beaucoup de travail, mais on peut automatiser la cartographie. Et j'utilise aussi des analyses factorielles pour comprendre ce qui se passe. Antoine Prost vient juste de publier son ouvrage sur les professions de foi qui est la première étude d'histoire qui utilise les analyses factorielles. Mes premières analyses, en 1977-78, sont faites avec des variables agrégées pour tester l'outil. Pour moi, ça semble impossible de le faire sans ordinateur.

Philippe Rygiel : Si j'essaye de répondre à la même question, j'aurais presque tendance à distinguer trois choses. Il y a un rapport à la quantification et le souci de la mesure qui je crois nous est commun. Et la mesure, ça suppose une instrumentation, sinon il n'y a pas de mesure. Il y a un rapport numérique à la fois aux outils et au monde numérique et là on peut tout à fait distinguer les deux. On peut tout à fait s'occuper de numérique sans s'occuper de quantification. Je pense que là on a tous des profils en tête pour lesquels c'est assez flagrant. Et puis il y a un rapport à la modélisation et un goût, ou un souci de la modélisation de la donnée, qui est un petit peu indépendant des autres éléments. Et je crois qu'on partage en fait tous les trois, et ça serait vrai d'Antoine Prost aussi par exemple, cette alliance de ces différentes dimensions, ce qui définit une position un petit peu originale dans la discipline.

Je pense qu'un élément pour moi, c'est que de toute façon, j'ai une culture qui est une culture scientifique aussi. Je ne suis pas un pur produit littéraire d'une certaine façon. Donc il y a des choses qui me semblent assez naturelles, aussi parce qu'au fond, le récit historique ne m'intéresse pas, et l'histoire comme écriture ne m'intéresse pas beaucoup non plus. Ce qui m'intéresse, c'est l'histoire comme technologie de recherche, comme méthode de production de savoir, comme travail d'un matériau à partir duquel on peut extraire des informations, reconstruire des mondes. Mais si j'ai envie d'un plaisir littéraire, je ne vais pas aller lire un historien, ils écrivent généralement comme des pieds. Donc il y a aussi des postures ou des façons de se situer implicitement dans la discipline qu'est l'histoire, qui sont en jeu là-dessous. Et dans le cas qui m'intéresse, les choses finalement sont aussi d'une simplicité biblique. Je veux reconstituer des parcours sociaux et géographiques de gens qui laissent peu de traces parce que ce sont des immigrants. Donc dans certains cas on a moins de traces de ces gens-là qu'on en a d'étudiants de la faculté de Paris vers 1520. On en trouve des mentions dans des listes, on en trouve un certain nombre de mentions dans des actes, on n'est pas toujours sûr que ce soit le même parce que si vous avez un Italien, un Espagnol ou un Polonais, les employés d'état civil ou de recensement ne vont généralement pas écrire son nom deux fois de la même façon. Donc le recours à la mesure et le recours, puisqu'il y a mesure, à une instrumentation, dérive de la question que je pose à un matériau spécifique. Je ne peux pas faire autrement, c'est-à-dire je ne peux pas obtenir de réponses aux questions que je pose sans passer par une mise en série, par une quantification, parce que je ne vais pas faire des biographies croisées de gens dont on a comme trace cinq mentions dans des sources sérielles. Ça n'est juste pas possible et ça n'est pas pensable. Donc dès le départ, c'est-à-dire si on parle finalement du début ou de la plongée en informatique ou en quanti, si on veut,

il y a pour moi quelque chose de très naturel et très simple, c'est que de toute façon, je ne peux pas faire autrement pour travailler ce matériel-là avec les questions que je lui pose.

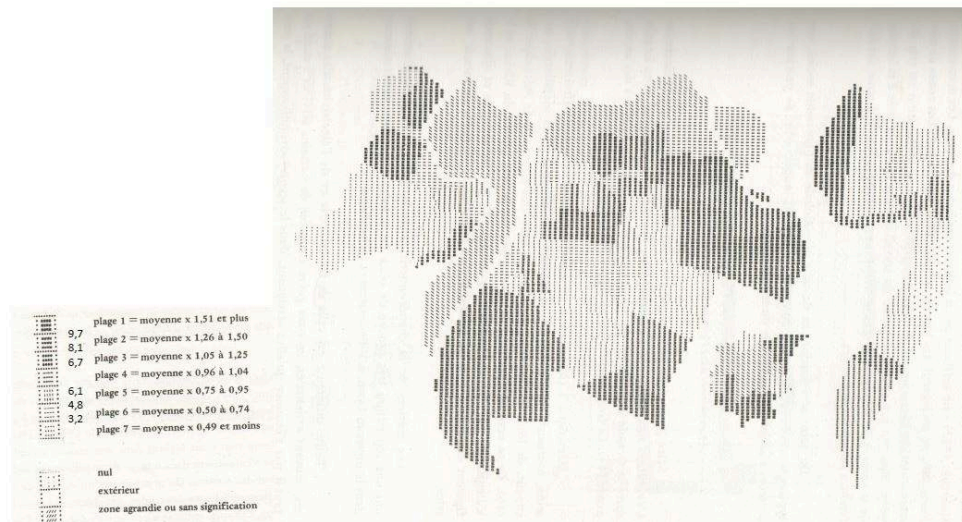
EDGAR LEJEUNE : *Merci beaucoup pour ces réponses claires. J'aimerais qu'on passe à un autre ensemble de questions qui porte plus directement sur les pratiques de l'informatique. Le premier sujet que je voulais aborder concerne la saisie des données, et j'aimerais que vous nous en parliez dans ses dimensions historiques, matérielles et épistémologiques ?*

Jean-Philippe Genet : Dans mon cas il faut distinguer analyse de texte et données prosopographiques. Pour l'analyse du texte, j'ai commencé bravement par rentrer mes premiers textes, ceux que j'étais, ceux qui m'intéressaient le plus, comme ceux de Sir John Fortescue. Pour ceux-là, c'était facile parce c'était des textes que je devais taper. Ça n'était pas une mince affaire, ça prenait beaucoup de temps, parce qu'évidemment quand il y a une erreur sur une carte perforée, comme vous le savez, on déchire la carte on en refait une autre, et ça pouvait donc doubler, voire tripler le temps. Je me suis aperçu que si je faisais ça, ma thèse allait prendre 50 ans. C'est une des raisons pour laquelle on a très vite utilisé les crédits du CNRS pour avoir des vacances pour rentrer les textes sur cartes perforées. Si on ajoute qu'il s'agissait soit de langues vulgaires médiévales, le français ou l'anglais, ou du latin, il fallait des gens qui connaissaient bien ces langues. On a fait appel comme ça, à des jeunes étudiants qui ont gagné quelques sous en faisant un travail particulièrement ennuyeux. On a continué jusqu'à ce qu'on ait le logiciel qu'on a fait avec l'ERC, qui a beaucoup simplifié les choses, parce que là, on a pu utiliser des textes que, simplement, on scannait, et on pouvait les rentrer directement sans tout retaper. Quand il a fallu tout de même faire saisir quelques textes, on a pu le faire pour des sommes relativement modiques grâce à des entreprises spécialisées installées en Inde. Ça a changé de façon considérable les choses au niveau du traitement de texte. On peut aussi traiter des ensembles beaucoup plus importants aujourd'hui. Quand on avait fait le travail pour les lettres de rémission qu'a utilisées Claude Gauvard pour sa thèse, on était obligé de faire une sélection, on prenait trente lettres par registre. Aujourd'hui, on n'a plus de limite comme ça. Et je signale que mon ami Alain Guerreau a utilisé lui directement la patrologie latine de Migne dans son entier, ce qui est une masse textuelle considérable. Ça présente d'autres problèmes parce qu'évidemment, il y a un peu de tout là-dedans, et puis les éditions sont d'assez mauvaise qualité souvent, mais néanmoins, c'est des masses de textes qui sont considérables. Donc là, c'est un gros verrou qui a sauté aussi.

Par contre, pour les bases de données, le travail de saisie, c'est entièrement personnel. C'est un travail qui est long, on le sait tous. Sans compter qu'il ne faut pas croire qu'on conçoit une base de données d'un seul coup quand les données ne sont pas des données sérielles mais des données qui sont construites de façon conceptuelle avec une construction des variables. Enfin, en tout cas, moi, je n'ai pas réussi. J'ai fait au moins deux, trois, quatre bases de données avant d'arriver à quelque chose qui m'est apparu satisfaisant. Donc, sans arrêt, il faut reprendre, refaire, ne pas s'emmêler et corriger ces données. Et voilà. Alors aujourd'hui, c'est beaucoup plus simple, parce que, évidemment, quand je fais une base comme *Studium*, je rentre mes données en clair : certes, dans une structure préétablie, mais sans avoir à me préoccuper d'un codage. Alors qu'autrefois, bien que les données soient en clair, les traitements n'intervenaient, que ce soit sur un gros système avec SPSS, ou avec FoxBase sur les microordinateurs, que sur des données qui étaient pratiquement

entièrement codées. On n'avait pas les facilités qu'on a aujourd'hui, où la mémoire est disponible en masse. Par contre, pour les bases de données de type prosopographique, il y a toujours le problème de la saisie, d'aller aux sources, d'aller comprendre. Et c'est donc un processus qui est long et qui implique beaucoup plus fortement la responsabilité personnelle de l'historien en tant que scientifique.

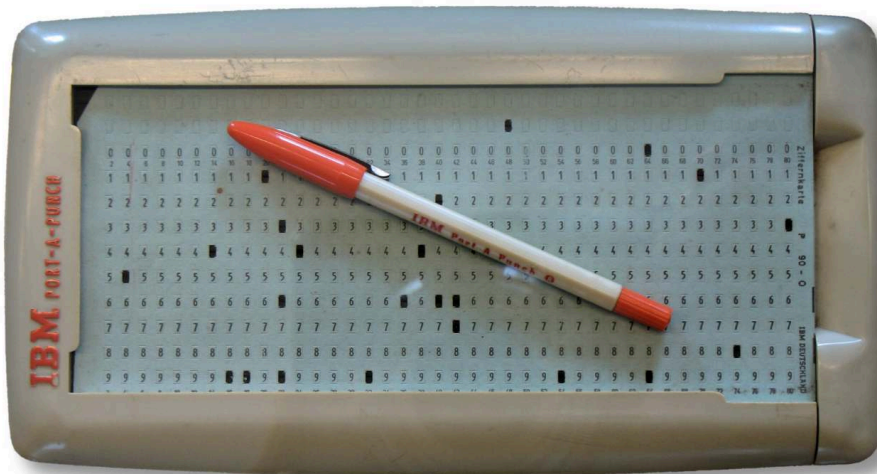
Jean-Luc Pinol : La première saisie se fait sur des cartes pré-perforées. Il n'y a que 40 colonnes et je dois rentrer un nombre limité d'informations à propos de beaucoup de gens, donc, par nécessité, je définis des nomenclatures *a posteriori*.



Carte des manœuvres, Lyon, Villeurbanne et banlieue. On saisit l'importance des manœuvres en banlieue (Vaulx-en-Verin et Vénissieux). A Lyon, leur poids est important derrière Perrache et à Gerland. Ils sont très nombreux à Villeurbanne à l'exception du quartier des Gratte-Ciel. Toutes les cartes sont construites de la même manière, les seuils dépendant de la moyenne et le l'écart type de chaque catégorie cartographiée. La plage centrale correspond à la moyenne diminuée ou augmentée de 4% de l'écart type. Pour les manœuvres, la moyenne est de 6,4%, l'écart type est de 0,3. Le seuil inférieur de la plage centrale est donc $6,4 - (0,96 \text{ de l'écart type})$ et le seuil supérieur $6,4 + (1,04 \text{ de l'écart type})$, soit 6,1% et 6,7%.

Jean-Luc Pinol

Un exemple : que le terme utilisé soit « manouvrier », manœuvre ou garçon de peine, tout ça je le mets dans la catégorie « manœuvre ». Je n'ai pas le choix, je ne peux pas l'écrire. Avec l'arrivée de la micro-informatique ça va changer ça parce que, exactement comme vient de le dire Jean-Philippe, je pourrais rentrer les dénominations, les appellations professionnelles en clair et faire mon code *a posteriori*, en fonction de ce que j'ai effectivement dans la base. Pour le cas de manœuvre, manouvrier, garçon de peine, j'ai dans ma thèse d'État pu faire la carte des différentes appellations, que je regroupe quand même sous la catégorie manœuvre, mais je peux voir que ça correspond à des espaces industriels ou manufacturiers bien différents dans la ville de Lyon. Pour ce qui est de la matérialité de la saisie, quand j'arrive aux archives départementales de Lyon avec mes cartes pré-perforées, que j'enlève mes confettis du support, je me fais remarquer d'une part, et puis une partie du personnel de salle me déteste parce que quand j'enlève les confettis, il y a de l'électricité statique et ça se colle partout sur les poubelles. J'explique, j'essaie de nettoyer...



Carte pré-perforée, support et poinçon
Jean-Luc Pinol

Philippe Rygiel : La saisie, je pense que c'est vraiment très différent selon les époques. Si je suis sur le début, c'est-à-dire les années 80/90, déjà il y a un vrai problème pratique c'est que faire la saisie directement en archive ça suppose d'avoir sa machine avec soi et de pouvoir la brancher, ce qui n'est pas toujours le cas. Du coup, c'est aussi pour ça que j'achète le premier portable qui existe parce que j'ai une autonomie de 7 heures, grâce à l'écran noir et plan 9 pouces et les grosses batteries au lithium. Je fais aussi beaucoup de saisies dans des greniers ou dans des archives de tribunaux d'instance, ou de toute façon il n'y a pas de prise, et il faut toujours calculer le temps que ça va prendre pour être sûr que l'on va réussir à optimiser sa journée. Donc il y a tout un tas de contraintes matérielles qu'on a complètement oubliées, d'autant plus qu'aujourd'hui on est sur une logique qui va être le cliché avec éventuellement une OCR derrière qui est parfaitement impossible à cette période-là. Et je crois que ce dont il faut se souvenir, c'est qu'en fait c'est extrêmement long et c'est extrêmement lent le travail de saisie. Ça représente des dizaines ou des centaines d'heures pour arriver à fabriquer quelque chose qui soit ensuite exploitable. On n'est pas du tout dans la logique d'aujourd'hui où on peut parfois récupérer des données ou extraire des données de choses qui existent déjà. On fabrique la source en fait. Jean-Philippe dirait qu'on fabrique la méta-source. Mais c'est un peu tous ce qu'on fait parce qu'en fait on agrège des informations sur des individus ou des objets qui proviennent de sources différentes en gardant la trace de la construction. Donc la saisie c'est déjà un moment extrêmement complexe, qui se fait dans des conditions matérielles qui sont compliquées, qui oblige à la penser.

profession fich état civil prov

Code moutre code CSP

absente	0.0.4	0.0.4	
agent mariposteur de poste par débiter altitudes de famille	0.0.5	0.0.4	
adjoint des	5.4.1.1.3	5.4.1	
adjoint des services économiques	5.4.1.5	4.5.1	cadre adm. moyen 44 (en)
adjoint technique	5.2	5.2	0.0
adjoint technique de laboratoire	5.2	6.4.2	0.0
adjoint	5.4.7.2	7.5.4.1.1	
adjoint chef	5.4.7.2.3	8.1.2.2	
adjoint	5.4.7.2.1	9.0.4.1.1	
agent assurance	6.4.2	10.6.4.2	
agent bureau	6.2.7	11.6.2	
agent chef	5.1.5	12.5.1	
agent chef	5.1	13.6.4.1.1	cadre moyen 44
agent chef	6.1	14.6.1	9
agent chef	1.0.4	15.1.1	
agent chef	5.1	16.6.4.2	ouv. commerce 28
agent chef	6.4.2	17.6.4.2	cadre
agent chef	7.6	18.8.0.4.2	
agent chef	5.1	19.5.2	ingénieur 33
agent chef	5.2	20.6.5.2	
agent chef	5.1	21.5.1	
agent chef	5.4.5	22.6.5.3	cadre moyen 44
agent chef	5.2	24.8.1.1	technicien 43
agent chef	6.5.7	25.6.5.3	ou 8.0.4.2
agent chef	5.1	26.6.5.3	8.0.4.2 employé bur
agent chef	5.2	27.5.2	
agent chef	5.2	28.5.2	
agent chef	5.2	29.6.1	technicien 43
agent chef	5.2	30.6.5.4	
agent chef	5.2	31.8.1.1	technicien 43
agent chef	3.5.7	32.6.1	ingénieur 33
agent chef	6.5.3	33.6.5.3	employé bur
agent chef	6.5.3	34.6.5.3	
agent chef	6.1	35.7.4	employé 51
agent chef	6.2	36.8.0.4.2	employé bur
agent chef	6.5.7	37.6.5.2	
agent chef	6.5.3	38.6.5.3	
agent chef	8.0.4.2	39.8.1.2.2	technicien 43 ou
agent chef	8.0.4.2	40.8.1.2.2	service public
agent chef	8.0.4.2	41.8.1.2.2	
agent chef	6.5.3	42.6.5.3	
agent chef	6.5.3	43.6.5.3	
agent chef	6.5.3	44.6.5.3	
agent chef	6.5.3	45.6.5.3	
agent chef	8.0.4.2	46.8.1.2.2	ouv. service public
agent chef	6.4.5	47.5.1	cadre adm. moyen 44
agent chef	5.1	48.5.1	cadre adm. moyen 44
agent chef	8.0.4.1	49.8.0.4.1.0	
agent chef	5.2	50.5.2	technicien 43
agent chef	5.2	51.5.2	
agent chef	5.2	52.5.2	
agent chef	5.2	53.5.2	technicien 43
agent chef	5.2	54.5.2	
agent chef	5.2	55.5.2	technicien 43

Page 1

Document de travail recensant (en partie) les dénominations professionnelles à partir desquelles étaient construites les classifications socio-professionnelles, qui consistaient souvent en une adaptation des diverses moutures des classifications Insee. La construction des classifications était souvent tout à fait centrale dans ces enquêtes et indispensable du fait de diverses limites techniques évoquées dans le texte

Philippe Rygiel

Et, alors je suis d'accord avec Jean-Luc. Le moment de la micro-informatique libère de certaines contraintes, par exemple celle de saisir des données déjà codées, parce qu'effectivement on n'avait pas la place d'écrire « manœuvres » ou « manouvriers ». Par contre on n'a pas la puissance de calcul qui permettrait de traiter des données qui ne soient pas des variables segmentées en modalités différentes, parce que les machines ne le permettent pas. Donc forcément on classe, on codifie, on y passe beaucoup de temps parce que c'est une opération complexe, en testant éventuellement différents codages pour voir si, en fonction de la façon dont on décrit l'information, on obtient des résultats, des significations différentes. Mais là encore, ce sont des pratiques qui ont une histoire. Par exemple, j'ai l'habitude pendant un temps de demander au centre d'archives des photocopies des choses qui m'intéressent, parce que ça va me permettre de travailler le soir et la nuit, mais c'est très coûteux.

PHOTOCOPIE DE LISTES NOMINATIVES DE RECENSEMENT

NUMERO	NOM	PROFESION	SEXE	AGE	STATUT	PROFESION
271	BOUQUIN	Com. Comm.	M	1885	Mariage	Com. Comm.
272	BOUQUIN	Com. Comm.	F	1886	Mariage	Com. Comm.
273	BOUQUIN	Com. Comm.	M	1887	Mariage	Com. Comm.
274	BOUQUIN	Com. Comm.	F	1888	Mariage	Com. Comm.
275	BOUQUIN	Com. Comm.	M	1889	Mariage	Com. Comm.
276	BOUQUIN	Com. Comm.	F	1890	Mariage	Com. Comm.
277	BOUQUIN	Com. Comm.	M	1891	Mariage	Com. Comm.
278	BOUQUIN	Com. Comm.	F	1892	Mariage	Com. Comm.
279	BOUQUIN	Com. Comm.	M	1893	Mariage	Com. Comm.
280	BOUQUIN	Com. Comm.	F	1894	Mariage	Com. Comm.
281	BOUQUIN	Com. Comm.	M	1895	Mariage	Com. Comm.
282	BOUQUIN	Com. Comm.	F	1896	Mariage	Com. Comm.
283	BOUQUIN	Com. Comm.	M	1897	Mariage	Com. Comm.
284	BOUQUIN	Com. Comm.	F	1898	Mariage	Com. Comm.
285	BOUQUIN	Com. Comm.	M	1899	Mariage	Com. Comm.
286	BOUQUIN	Com. Comm.	F	1900	Mariage	Com. Comm.
287	BOUQUIN	Com. Comm.	M	1901	Mariage	Com. Comm.
288	BOUQUIN	Com. Comm.	F	1902	Mariage	Com. Comm.
289	BOUQUIN	Com. Comm.	M	1903	Mariage	Com. Comm.
290	BOUQUIN	Com. Comm.	F	1904	Mariage	Com. Comm.
291	BOUQUIN	Com. Comm.	M	1905	Mariage	Com. Comm.
292	BOUQUIN	Com. Comm.	F	1906	Mariage	Com. Comm.
293	BOUQUIN	Com. Comm.	M	1907	Mariage	Com. Comm.
294	BOUQUIN	Com. Comm.	F	1908	Mariage	Com. Comm.
295	BOUQUIN	Com. Comm.	M	1909	Mariage	Com. Comm.
296	BOUQUIN	Com. Comm.	F	1910	Mariage	Com. Comm.
297	BOUQUIN	Com. Comm.	M	1911	Mariage	Com. Comm.
298	BOUQUIN	Com. Comm.	F	1912	Mariage	Com. Comm.
299	BOUQUIN	Com. Comm.	M	1913	Mariage	Com. Comm.
300	BOUQUIN	Com. Comm.	F	1914	Mariage	Com. Comm.

Photocopie de listes nominatives de recensement, l'une des sources sérielles les plus utilisées par les contemporanéistes faisant du quantitatif dans les années 80/90. Elle illustre le fait que la construction de bases de données consommait parfois beaucoup de papiers. Il était en effet impossible de prendre des clichés sur place, alors même que les centres d'archives étaient parfois lointains et souvent fermés. Le plus simple et le plus efficace était parfois d'y passer commander une épaisse pile de clichés afin de pouvoir saisir chez soi, surtout que les ordis n'étaient pas très portables non plus.

Philippe Rygiel

Aujourd'hui ça a totalement disparu. Parfois, je vois des collègues qui dans un centre d'archives vont passer toute la journée à cliquer un carton, mais sans jeter un œil sur ce qu'il y a dedans, d'une certaine façon. On n'est pas du tout dans la même économie. La saisie, c'est aussi le moment où on lit, et où on a eu une imprégnation de ce matériau-là, au point que je peux reconnaître certains des 32 000 individus croisés pendant ma thèse en me disant « ah tiens celui-là je croyais qu'il était marié mais il a dû divorcer ». Donc on a une connaissance très très fine en fait, du grain du matériau et on n'est pas du tout dans la même posture vis-à-vis de nos données que celle du sociologue qui va extraire du tableau INSEE la description d'un quartier et de ses habitants, et qui, d'une certaine façon, ne connaît pas les individus qui sont dedans. C'est un moment extrêmement fastidieux, c'est un travail de copiste, on est beaucoup plus proche du Moyen Âge que d'autres choses, mais c'est aussi un moment d'imprégnation et de réflexion. Heureusement parce que vu les centaines d'heures qu'on y passe.

EDGAR LEJEUNE : À l'autre bout du traitement de ces données se trouvent ce qu'on appelait autrefois les documents de sorties (graphiques, cartes, tableaux, listes, etc.). Comment avez-vous vécu le passage du listing des années 1970 aux sorties sur écrans que nous connaissons aujourd'hui ? Et qu'est-ce qui a marqué vos pratiques dans ces évolutions technologiques ?

Jean-Philippe Genet : Moi je n'ai pas grand-chose à dire là-dessus parce que je n'ai pas de sorties particulières. Je refais tous les graphiques, y compris ceux d'analyses

factorielles, et je triche honteusement, je remplace les choses, parce que je me suis aperçu que ceux qu'on sort en machine ne vont pas. Ou bien si, on peut faire des très belles choses avec la couleur, mais après quand on passe à imprimer des volumes ou des articles de revue, on n'a plus la couleur, alors ça n'arrange rien. Donc je fais des sorties, je fais en sortie plutôt des schémas que de véritables sorties venues directement de l'ordinateur. Ça, je laisse ça à Stéphane Lamassé qui est un virtuose de la chose, il adore ça, notamment pour les analyses de réseaux et les analyses factorielles, avec des couleurs extraordinaires. Mais bon, quand j'ai besoin, je fais appel à ses services, à son imagination, mais moi je suis totalement nul en sortie. Je me contente de graphiques et de tableaux. J'aime beaucoup les tableaux de chiffres, on me le reproche beaucoup, mais je trouve qu'un tableau de chiffres, ça vaut souvent des pages et des pages de développement insipides. Il suffit de savoir lire un tableau de chiffres, en général, toutes les informations sont là et ce n'est pas la peine de gloser. Voilà. Bon, c'est une réponse qui va indigner beaucoup de gens, mais je suis un peu comme Philippe Rygiel, qui disait qu'il ne faisait pas de l'histoire pour la beauté de l'écriture historique. C'est aussi mon cas.

Jean-Luc Pinol : Je ne partage pas votre opinion sur les tableaux de chiffres, parce qu'effectivement, il y a des gens qui savent lire des tableaux et puis il y a des gens qui ne savent pas les lire. D'où l'intérêt de la data visualisation et de logiciels comme Amado, directement inspiré de la Sémiologie graphique de Jacques Bertin, qui permettent de produire des analyses et des visualisations de matrices de données par exemple.

Philippe Rygiel : J'ai retrouvé plusieurs listings récemment, et je faisais pas mal de sorties papier en fait. Il y a d'abord une raison toute bête et pratique à ça, c'est qu'à l'époque on n'a pas plusieurs écrans, on n'est pas dans l'espèce de position du *trader* ou du pilote de ligne qui a trois ou quatre surfaces sur lesquelles il peut consulter quelque chose et le confronter. Vu le coût de l'équipement, on a généralement un petit écran, qui va faire 13 pouces, ou 15 si on a décidé d'investir, mais si on veut écrire quelque chose tout en consultant autre chose, en fait, le réflexe le plus logique c'est de l'imprimer. Parce qu'on peut à ce moment-là écrire en lisant ses tableaux de Burt on ses analyses factorielles.

55

situation de famille @ V G C

Profession du père :

ENTREE en FRANCE 1924 dans le dpt 1924

PROFESSIONS C. U

Lieu de résidence B V L @ LAV MEH profession actuelle C. U

rémunération 700 / m. loyer annuel 600 F dec 31

DATE RÉPONSE « nov 1932 OUI AJOURNÉ MOTIF

PASSAGE à ROSIERES / N

TRAJET

REMARQUES

au budget s @ de l'écrit des 4 enfants mineurs
 → budget total 2500 / m.
 X commissaire special 11 dec 31.

en ce qui concerne sa profess. En ce qui
 concerne de déclarer que aucun ou ne veut plus accepter
 comme gérant naturel le travail de cette catégorie
 d'ouvriers, serait en effet possible et mal propre, et
 en ce qui concerne la présence de mineurs et les
 en foyers sont heureux de recevoir des ouvriers
 étrangers qui acceptent de les ils ces condit° ?

des 4 enf mineurs dont naturalisés / in d'écrit

Feuille de pré-saisie, imprimée sur une imprimante matricielle à aiguille sur du papier listing, l'équivalent de la K7 audio, un support d'écriture que tous ceux qui ont connu cette époque reconnaissent immédiatement. Pour les mêmes raisons que plus haut (le fait qu'emmener une machine dans un centre d'archives n'était pas toujours possible), il pouvait arriver que la saisie se fasse en deux temps, une activité de copiste reportant manuellement ici sur une feuille préparée à l'avance les renseignements qu'on allait ensuite saisir une fois rentré chez soi. Les bases de données de la seconde moitié des années 80 ne permettaient généralement pas encore d'intégrer au système de gestion de la base de données des champs texte permettant d'intégrer des éléments qualitatifs dans la base elle-même. Ces informations pouvaient dans certains cas demeurer en format analogique du fait du temps nécessaire à une double saisie, manuelle puis informatique.

Philippe Rygiel

Contrairement à Jean-Philippe, j'ai quand même un rapport à l'analyse factorielle plus ambigu, parce que pour moi, elle est d'abord exploratoire, elle est d'abord un moyen d'aller visualiser l'information d'une certaine façon. Ce n'est pas forcément à partir de ça que je vais argumenter. Mais si je dois la donner au public, qu'elle accompagne le texte ou l'article, il est bien évident qu'il va falloir la retravailler sous un logiciel de dessin ou sous un logiciel de mise en page parce que ce qui sort directement des logiciels d'analyses à l'époque est peu utilisable... Je crois que c'est au début des années 90, je me mets à utiliser SPAD, qui est je crois une évolution de SPSS, et une sortie SPAD, vous ne pouvez pas coller ça dans un bouquin ou dans un article. Donc il va falloir remettre en page, retravailler. Mais généralement, je sors aussi les calculs intermédiaires. Pour l'analyse factorielle, je vais sortir le grand tableau général qui va me pointer vers les relations entre variables, que je vais stabiloter pour faire apparaître les endroits où il se joue quelque chose. On est dans un contexte où je crois l'énorme différence avec aujourd'hui, c'est qu'en fait on choisit très soigneusement le type d'analyse qu'on va mener et on n'a pas la possibilité de multiplier les angles d'approche ou de faire tourner plusieurs algos parce que c'est extrêmement coûteux en temps. Quand je veux, au début des années 90, obtenir un tri croisé, croiser deux variables sur une base de données un peu

importante, en fait je vais programmer ça le soir et puis je vais laisser la machine tourner toute la nuit en espérant que je n'ai pas oublié un point-virgule, qu'il n'y aura pas de micro coupures d'électricité et je vais récupérer les résultats le matin. C'est la puissance de calcul qui joue.

Jean-Luc Pinol : Je suis tout à fait d'accord avec le fait qu'on ne multiplie pas les analyses factorielles, on ne multiplie pas les cartes, on les prévoit. Et puis, parce que c'est coûteux. Enfin, c'est coûteux en temps. Quand je fais des analyses factorielles, je les lance le soir et je les récupère le matin si tout s'est bien passé. Alors après je vais quand même améliorer un peu les choses et j'arrive à faire des analyses factorielles sur des fichiers qui sont petits, d'une certaine manière, j'ai les résultats au bout de 45 minutes. Ce qui est déjà une grande avancée... et si ça ne passe pas, je le vois tout de suite, je relance, etc. Et quand je fais ma thèse de 3^e cycle, en plus, je suis prof au lycée. Donc voilà, il faut aller à l'essentiel.

Et au niveau des visuels moi je suis un peu comme Philippe, j'imprime beaucoup, parce que les écrans sont petits – là c'est au moment de ma thèse d'État. Pour rédiger, on ne peut pas avoir deux écrans. Même s'il y a quelques années de différence et que pour moi c'est un peu avant, c'est exactement ce que décrit Philippe. Aujourd'hui, c'est sûr qu'une partie de ce que je fais, je l'ai sur des cartes interactives, sur des sites web. Par exemple, ce que j'ai fait à la demande de Serge Klarsfeld sur la déportation des enfants, puis après sur la rafle du Vel d'Hiv, là c'est en ligne et c'est totalement différent. Je me souviens quand j'avais expliqué à des étudiants qu'avant d'utiliser un fond de carte, il faut quand même le dessiner, le vectoriser, etc., il me disait « Mais c'est très long ! ». Par rapport à ce que je faisais dans les années 70 pour réussir à faire une carte, avant qu'il y ait des cartes graphiques, ça change complètement.

EDGAR LEJEUNE : *À propos de ces pratiques concrètes de l'informatique, parlons de ce que l'on considère le plus souvent comme le cœur de cette discipline : la programmation. Avez-vous programmé vous-même ? Et dans quelle mesure était-ce nécessaire pour vous dans le but d'utiliser des ordinateurs pour vos recherches historiques ?*

Jean-Philippe Genet : Je peux répondre en disant simplement qu'aujourd'hui je ne programme plus. Voilà, j'utilise des logiciels. Je suis simple utilisateur. C'est moins gratifiant sur le plan intellectuel mais en termes de temps, il n'y a pas photo. De toutes façons, j'ai très peu programmé. Je l'ai fait au début, c'était long, il y avait des erreurs, il fallait recommencer, etc. C'est un des problèmes parce que quand même notre temps, il faut autant que possible le consacrer à notre travail d'historien. On ne peut pas à la fois être un bon informaticien et un historien à part entière. En tout cas moi c'est ce que je pense. Mais c'est important d'avoir appris à programmer pour ne pas être victime de l'effet boîte noire. Et j'ai essayé dans la mesure du possible de suivre le progrès. Puisqu'ici on avait une formation, j'ai fait des formations sur des logiciels récents, comme Python. D'ailleurs, plus que la programmation, c'est la capacité à utiliser certains langages qui compte aujourd'hui. C'est un peu ce que l'on faisait autrefois quand on écrivait des lignes de *query langage* pour FoxBase, en somme, une programmation est interne au logiciel. Voilà, c'est pas du tout la programmation qu'on faisait quand on travaillait en FORTRAN au tout début. Il y a une différence. Donc aujourd'hui, je préfère dire que je ne programme plus.

Jean-Luc Pinol : Je n'ai jamais vraiment été programmeur. Tout à l'heure, je parlais d'analyses factorielles sur Macintosh, comme elles avaient été écrites en BASIC pour des PC, ce n'était pas les mêmes processeurs, il fallait adapter, mais ce n'est pas de la

programmation, c'est plutôt de l'adaptation. Alors, je dirais que j'ai un vernis et que je suis un tout petit peu alphabétisé, mais aujourd'hui je suis tout à fait sur la même ligne que Jean-Philippe. Moi je ne programme pas, je n'ai jamais vraiment programmé. En revanche, c'est important de pouvoir dialoguer avec des gens qui savent programmer pour avancer. Par exemple, quand j'ai dirigé le très grand équipement ADONIS, il fallait que j'aie un minimum d'alphabétisation pour pouvoir parler avec les ingénieurs. Mais après, ce n'est pas moi qui mets les mains dans le cambouis au sens strict, même si je les mets pour les données, pour un certain nombre de traitements manuels, parce que très souvent on automatise une partie d'un traitement, par exemple pour affecter des coordonnées décimales à des points adresses. Donc il y a une partie qui se fait automatiquement, je parlais tout à l'heure de ce que j'avais fait avec Serge Klarsfeld pour les enfants, mais il y a toujours, on va dire, 5 ou 6 % des données qui ne passent pas et qu'il faut traiter manuellement, d'une certaine manière. Ça, je le fais, mais ce n'est pas de la programmation.

Philippe Rygiel : Oui, je crois que je vais dire à peu près la même chose, mais peut-être en précisant que la programmation c'est un mot valise qui peut désigner un ensemble de pratiques, de complexité très très variable. Je crois que je dirais que j'ai été scripteur, pas programmeur au sens où je suis incapable de concevoir une architecture logicielle. Par contre, on avait besoin d'écrire des lignes, tout simplement pour que l'ordinateur fasse ce qu'on voulait qu'il fasse. Quand vous étiez, même sous un tableur dans les années 90, si vous vouliez obtenir un test du khi-2 ou un test de Fisher ou une analyse factorielle, il fallait écrire des lignes de code... mais on ne programmait pas, on ne créait pas des logiciels. On utilisait des langages pour poser une requête. Quand on est arrivé sur les bases de données SQL, si vous vouliez que la base de données réponde à une interrogation, il fallait bien écrire le code qui permettait à la base de répondre. C'est pareil sous SPAD où de toute façon il fallait plus ou moins lui poser une question dans quelque chose qui était dérivé du FORTRAN. Donc je crois qu'on peut dire qu'on a tous écrit des lignes de code, ce qui ne veut pas dire que nous étions programmeurs, qu'on le fait sans doute moins aujourd'hui pour des raisons qui sont liées à l'évolution de l'informatique bien sûr, qui sont liées aussi au fait qu'il y a un moment où on n'a plus ni le même âge ni les mêmes fonctions. Si j'arrive aujourd'hui à faire une heure de science dans la journée, je suis content, et ça ne m'arrive pas tous les jours.

EDGAR LEJEUNE : *Merci beaucoup pour ces réponses tout à fait éclairantes sur les relations entre pratiques de l'informatique et pratiques de l'histoire. J'aimerais que l'on échange désormais sur les évolutions collectives et disciplinaires liées à l'utilisation des ordinateurs dans les champs de recherches qui vous concernent. Quels sont les espaces d'échanges scientifiques qui se sont développées (colloques, séminaires, sociétés savantes, etc.), et quel a été votre implication dans cette dynamique ?*

Jean-Luc Pinol : En ce qui me concerne, au moins jusqu'à la soutenance de mon troisième cycle, voire de ma thèse d'état, j'ai le sentiment d'être un peu isolé. Je vais commencer à faire des cours et des formations seulement quand je me retrouve professeur à Strasbourg. Je monte un centre de recherche, le centre de recherche acquiert des machines, puis on a des cours dans le cadre du DEA sur l'utilisation d'un certain nombre de logiciels, comme SPAD dont Philippe parlait tout à l'heure, mais c'est relativement tardif. J'ai eu ce sentiment. J'ai participé quand même ponctuellement à des formations, ou à des colloques à Paris, où je rencontre André Zysberg. En 1979 ou 1980, je suis une école d'été d'informatique pour les sciences

sociales à Grenoble qui est organisée par Frédéric Bon, qui est un politologue décédé très jeune. Ça me permet de rencontrer un certain nombre de gens, hors du cadre du laboratoire. J'ai aussi beaucoup d'échanges avec des géographes comme Pierre Dumolard, membre du groupe Chadule, qui m'aide beaucoup et avec qui on va d'ailleurs commenter les élections régionales à la télévision. On produisait des cartes et des commentaires en *live* de résultats des élections. Les premières élections qu'on étudie, ce sont les élections cantonales de 1986 dont on fait des cartes sur Minitel. Mais tout ça, ce n'est pas véritablement organisé dans le cadre de l'histoire contemporaine, pendant longtemps en tout cas.

Philippe Rygiel : Oui, je crois que là, pour le coup, il y a un peu une spécificité des médiévistes, et j'ai envie de dire des médiévistes parisiens, avec *Le Médiéviste et l'ordinateur*, puis toute l'aventure qui va mener à *Histoire et Mesure* et à la naissance du PIREH. Donc c'est vraiment du côté de Jean-Philippe, que l'institution est la plus forte. En contemporaine, il y a des liens d'interconnaissance et des réseaux, c'est-à-dire que là vous réunissez des gens qui se connaissent depuis 30 ou 40 ans. Les travaux de Jean-Luc, les initiatives de Jean-Philippe ou ce que fait Antoine Prost au CHS ce sont des choses avec lesquelles je suis en contact assez vite, mais c'est une poignée de gens que ça intéresse pendant très longtemps. Et l'isolement est plus fort quand on est en province, et comme je ne fais pas ma thèse à Paris, je remarque que les embryons d'institution, ça se structure quand même plutôt autour des lieux parisiens avec éventuellement un prolongement à Caen, mais plutôt pour les modernistes. Donc du coup, pendant très très longtemps, pour la contemporaine, on pourrait presque décrire ça comme des individus qui se connaissent : ils ne sont pas très nombreux, sans qu'il y ait véritablement de structure. Et puis il y a des moments de rencontres mais ça arrive assez tard. Le premier colloque sur Internet par exemple est organisé par Antoine Prost à Orléans, en 1999 ou quelque chose comme ça. À la fin des années 1990, il y a eu des colloques sur les bases de données, dont un organisé par l'École Française de Rome, mais là, j'ai oublié les dates. Mais ce sont des rencontres ponctuelles où on essaye de se tenir un peu au courant. Et puis on va chacun déployer nos propres connexions, j'ai envie de dire, avec les gens qui sont à portée et qui peuvent éventuellement nous apprendre des choses. Moi je vais longtemps travailler avec Éric Guichard par exemple. Pour résumer il y a une très faible institutionnalisation disciplinaire, et d'une certaine façon elle ne va pas se faire, en partie je crois, parce qu'il y a le mouvement et le phénomène des humanités numériques qui se met en place, avec le manifeste des humanités numériques par exemple, dont je suis, avec Jean-Luc je crois, l'un des signataires.

Jean-Luc Pinol : Oui en effet.

Philippe Rygiel : Mais là d'une certaine façon, je ne vais pas rentrer dans cette institutionnalisation-là, puisque pour moi les humanités numériques c'est une histoire d'ingénieurs et d'ingénierie, ce n'est pas une histoire de recherche et ce n'est pas une histoire disciplinaire. Donc ce ne sont pas mes enjeux d'une certaine façon, ce n'est pas ça qui m'intéresse, je suis un historien qui utilise le numérique, je n'appartiens pas au champ des humanités numériques qui est un autre champ. Il ne s'agit pas d'ailleurs de dire qu'il n'a pas sa légitimité ; simplement, ce ne sont pas mes problématiques. Et du coup, il n'y a pas institutionnalisation aussi parce que la question du numérique, ce n'est pas la même que la question du quantitatif et il y a un petit milieu de quantitativistes, mais qui n'a pas forcément un intérêt flagrant

pour le numérique, les problèmes de modélisation, d'extraction de l'information, ce n'est pas tellement sa culture ou son souci d'une certaine façon. Et du coup on va aussi, je crois Jean-Luc comme moi, trouver des interlocuteurs à l'international. Un des derniers textes que j'ai signé sur ces questions, je l'ai signé avec Jo Guldi et avec Simon Burrow, c'est-à-dire avec une professeure américaine et un professeur de la Western Sydney, parce que d'une certaine façon, le milieu français est presque trop petit dans notre champ pour que ça s'institutionnalise. Et quand il le fait, il lorgne du côté humanités numériques qui n'est pas là où se passent les choses qui m'intéressent au fond.

Jean-Philippe Genet : Pour ma part l'institutionnalisation m'a tout de suite intéressée. D'une part, comme je vous l'ai dit, parce qu'avant les micros, l'accès aux ordinateurs passait par les machines du CNRS. Donc j'étais obligé de me lancer du côté du CNRS pour aboutir à ce qui est aujourd'hui le LAMOP. Et puis aussi au niveau de la diffusion de l'informatique, il y a eu ce colloque qui avait été organisé par Lucie Fossier, André Vauchez et Cinzio Violante à l'École française de Rome en 1975, et qui a été, je crois, la première grande apparition de l'informatique chez les historiens, pas seulement d'ailleurs les médiévistes. Ça a été l'occasion avec Lucie Fossier – qui vient de disparaître et qui a été une grande pionnière qu'il faut saluer, en créant très tôt une section d'informatique à l'IRHT – et Caroline Bourlet de créer *Le Médiéviste et l'ordinateur*. C'était un petit bulletin ronéoté mais qui a été très bien diffusé de 1979 à 2003, à plus de 500 exemplaires, chez les collègues médiévistes et les archivistes et pour tous ceux qui le demandaient. Et à partir de ça, ça s'est développé et j'ai participé à une association qui s'appelait *History & Computing*, lancée par Peter Denley et Deian Hopkin, qui est d'origine anglaise, mais comme je travaille sur l'Angleterre, j'étais en contact tout de suite. J'en ai été président deux ans, je crois, et on a organisé quand même toute une série de colloques, dans les années 80. Il y en a eu en France, d'ailleurs, à Bordeaux et à Montpellier.

Et puis, *Le Médiéviste et l'ordinateur*, on a voulu le garder sur le Moyen Âge, pour un public ciblé, d'où une discussion un peu aigre-douce avec Zysberg. Au début il proposait de faire « histoire et informatique », mais nous on a tenu à l'identité médiéviste, du coup il a créé une association pour les modernistes et contemporanéistes, l'Association française pour l'histoire et l'informatique, avec une revue *Histoire et informatique* pour tous les historiens, dont le bulletin s'est *de facto* limité au public des modernistes et des contemporanéistes ; il a d'ailleurs très bien marché. Et puis, je pense que tout ça finit par s'arrêter, parce qu'avec la micro-informatique, la contrainte du regroupement obligatoire qu'impliquait l'informatique en batch, avec l'attente dans les centres de calcul, s'est allégée : quand on allait à Orsay, on travaillait avec des géographes, etc. Quand j'allais au centre de calcul de Paris 1, c'était avec des économistes, des géographes, des sociologues, c'était tout à fait pluridisciplinaire. Mais avec l'internet, l'échange pluridisciplinaire est devenu encore plus facile, si bien que les associations purement historiennes ont eu tendance à se dissoudre. En tout cas, on a fini par arrêter *Le Médiéviste et l'ordinateur* en 2003, au bout de 41 numéros. Il n'avait plus vraiment d'utilité. L'association d'André Zysberg et l'association internationale *History & Computing* se sont arrêtées au début des années 2000 de la même façon. Je crois qu'il subsiste quelque chose en Angleterre sous la forme de *IJHAC : A Journal of Digital Humanities* que publie toujours l'université d'Édimbourg.

Alors aujourd'hui effectivement, les choses se relancent du côté du numérique, alors je suis un peu comme Philippe, je ne suis pas spécialement intéressé par les humanités numériques, même si c'est important pour les jeunes ici qui font la formation des étudiants, pour leur future carrière, c'est très important d'être impliqué dans les humanités numériques, donc ils ont pris en compte cette distance-là. Au LAMOP, avec Stéphane Lamassé et son PIREH, on se tourne beaucoup plus vers la coopération avec des mathématiciens ou des informaticiens. On a une ANR justement, avec les ingénieurs et des professeurs du CNAM, des collègues qui font de l'informatique dure, et on travaille sur des concepts qui sont un peu différents des concepts auxquels on est habitué à travailler en histoire, par exemple sur la fiabilité des données ou, pour l'analyse du langage, sur des problèmes comme l'incertitude : des choses de ce genre qui n'étaient pas du tout dans l'optique des historiens il y a seulement une dizaine ou une quinzaine d'années. On fonctionne aussi avec des publications de colloques, mais c'est plutôt Stéphane Lamassé qui s'en occupe. Je me contente de bénéficier de son expérience et de ses compétences, ainsi que celles de Julien Alérini, Octave Julien, Gaëtan Bonnot, enfin tous les jeunes du PIREH qui travaillent sur ces choses-là.

Philippe Rygiel : J'ai d'ailleurs indirectement, moi aussi, profité du travail de Jean-Philippe parce que j'ai été, et je suis encore connecté aux gens du LAMOP avec qui j'ai signé un certain nombre de textes, bien que n'étant pas médiéviste... Et je crois que je partage avec Jean-Philippe le fait que mes connexions, mes collaborations sont faites directement avec des informaticiens et des mathématiciens beaucoup plus qu'avec des gens des humanités numériques. C'est là que je trouve des éléments qui nourrissent ma réflexion ou ma pratique et c'est un intérêt ancien puisque dans la conclusion même de ma thèse, il y a un certain nombre d'éléments qui pointent vers le fait que la donnée historique est susceptible d'être modélisée au mieux par des méthodes – on appelle ça à l'époque les *fuzzy set variables* – qui sont très proches du domaine de l'intelligence artificielle actuelle. De toute façon la question pour moi c'est celle du traitement de la modélisation et de l'extraction d'informations à partir de la donnée et de l'archive. Et du coup ça incite à travailler avec les gens qui partagent ces questions-là, dont les gens du LAMOP d'ailleurs, et à chercher des collaborations aujourd'hui avec les mathématiciens et les informaticiens.

Jean-Luc Pinol : Je rejoins Philippe et Jean-Philippe quant à leur sentiment sur les humanités numériques. En raison de mes fonctions au TGE ADONIS, j'étais dedans de façon très pragmatique à partir de 2010, sans pour autant adhérer forcément à tout ce qu'on dit autour des humanités numériques. Au niveau personnel, et non comme représentant d'une institution, j'ai toujours privilégié le spatial : j'avais, disons, un réseau au Royaume-Uni et aux États-Unis dans ce qu'on appelle *spatial history* en anglais, alors qu'en France, on ne sait pas ce que c'est. C'est par ce biais-là, disons, que je me retrouve au niveau des humanités numériques et aussi parce que je suis toujours impliqué dans les consortiums de l'infrastructure de recherche Huma-Num. Mais c'est vrai que je suis assez d'accord avec ce que dit Philippe. Moi, où j'ai avancé sur un certain nombre de choses, c'est en fréquentant directement des informaticiens. Et en particulier, grâce à ce séminaire un peu particulier qu'est Rochebrune, auquel est venu Stéphane Lamassé cette année. Je crois que j'étais le premier historien à y aller. On est une trentaine, il y a des informaticiens, des juristes et des gens des sciences sociales. Et comme on est dans un chalet isolé, on a des

discussions, on a le temps de discuter. Donc voilà, il y a une dissociation entre ce que j'ai pu faire dans le cadre d'institutions, comme le TGE ADONIS, ou aujourd'hui l'IR Huma-Num, et puis ce que j'ai fait en tant que chercheur.

EDGAR LEJEUNE : *Je profite du fait que vous ayez ouvert la discussion sur la numérisation de la discipline pour vous demander comment vous vous expliquez le fait qu'il n'y ait pas eu de virage numérique aussi fort en histoire que dans d'autres disciplines ? En archéologie par exemple, les questions qui nous animent aujourd'hui ne font même plus l'objet de discussions tant les chercheurs ont intégré les outils numériques et statistiques à leurs pratiques.*

Philippe Rygiel : En archéologie, il y a quand même quelque chose qui a à voir avec le matériau qu'on traite et les pratiques concrètes de recherche. Quand vous avez des tessons de poterie, ou quand vous avez des coquilles d'huîtres, vous n'allez pas en dire grand-chose si vous ne passez pas par la mise en série. Et dans le travail archéologique, il y a en plus une obligation de patrimonialisation et d'enregistrement des pièces, qui fait qu'il y a de toute façon un langage descriptif. Il y a donc une ontologie assez spontanée et partagée, alors que fabriquer une ontologie qui soit adaptée à tous les historiens de la période contemporaine, ce n'est quand même pas simple. Je ne suis pas absolument persuadé que ce soit possible d'ailleurs, même si on peut fabriquer des ontologies locales. Donc là, je pense que pour l'archéologie, c'est assez lié à la matérialité même ou à la pratique même de ce qu'est l'archéologie. Après, une de mes hypothèses, c'est qu'à partir de la micro-informatique, pour les mieux dotés de l'institution, le rapport à l'informatique n'était plus un signe de distinction suffisamment clair qui garantissait une progression rapide de carrière, alors que la maîtrise littéraire ou la connexion aux lettres permettait, en particulier au normalien, de fabriquer des choses parfois assez médiocres très vite et leur donnait un signe de distinction immédiatement repérable par l'institution. Donc je pense que dans l'économie du monde universitaire, ça n'était pas rentable en fait, cet intérêt.

Jean-Luc Pinol : À ce propos, je me souviens d'un colloque du CNRS en histoire sociale. Bien évidemment, je ne donnerai pas le nom du collègue qui a fait cette réponse, mais il nous avait expliqué tout ce qu'il avait fait grâce à la prosopographie et quelqu'un lui avait posé la question « Pourquoi est-ce que vous n'avez pas fait d'analyses factorielles ? ». On s'attendait à une réponse un peu élaborée, argumentée et ce collègue a eu cette réponse extraordinaire : « Avec une analyse factorielle, on ne peut pas faire long ! ».

Jean-Philippe Genet : Je suis absolument d'accord avec Philippe et je pense que le tropisme littéraire est largement responsable de cette attitude. Pour ma part, je citerai un nom, c'est quelqu'un que j'aime bien par ailleurs : Dominique Barthélémy qui, dans son discours de réception à l'Institut, il y a quelques semaines à peine, a encore dit à quel point il était absurde de perdre son temps à faire des bases de données, alors qu'il suffisait de prendre sa plume et d'exprimer élégamment et clairement ce que l'on pensait sur une question historique. Voilà, en tout cas, ce que j'ai compris de son discours. J'étais dans la salle, et je n'étais pas le seul à avoir fait des bases de données. Mais d'ailleurs, je ne lui en veux pas, il exprime certainement ce que pensent beaucoup de collègues...

Philippe Rygiel : Et je pense qu'en plus, il y a une autre raison qui est assez objective, c'est qu'il y a des domaines où la donnée n'est pas chère, c'est-à-dire que la production de la donnée exige des ressources assez modestes. Il y a au contraire des

domaines où l'acquisition de la donnée est extrêmement chère et coûteuse, et ça a longtemps été le cas en histoire, parce que fabriquer des bases de données ça veut d'abord dire effectivement : aller aux archives, saisir, construire, c'est extrêmement long, c'est extrêmement fastidieux. Et ce qui est intéressant je pense, c'est que là nous sommes à un moment très très particulier où le coût de l'acquisition de la donnée historique est en train de diminuer drastiquement, pour des raisons qui sont bêtement liées au progrès des algorithmes d'intelligence artificielle et des capacités de calcul, où il devient possible en fait de produire de l'information de manière semi-automatique à partir de sources primaires ou d'éléments manuscrits, et où parser un fichier devient simple. On a d'ailleurs un certain nombre d'ANR en cours autour de ça. Et ça je pense que c'est une configuration qui peut radicalement changer les choses parce qu'on va changer d'échelle dans les volumes d'informations exploitables selon ces méthodologies. On va aussi changer de gamme de coûts pour l'usage ou le travail à partir de ça. Je pense que ça va prendre un temps parce qu'il y a assez peu de gens qui sont encore formés à ce genre de choses, mais on a des API aujourd'hui qui permettent de récupérer assez facilement de la donnée depuis un certain nombre de gisements d'informations dont certains sont organisées de façon à pouvoir être utilisées par les historiens. Donc je pense qu'on pourrait avoir une bascule, et pour le coup la bascule est liée à l'intelligence artificielle, mais pas celle à laquelle on pense, au progrès des algorithmes d'apprentissage non supervisé ou supervisé en fait, qui est spectaculaire au cours de ces dernières années.

EDGAR LEJEUNE : *N'y a-t-il pas eu également, comme vous le mentionniez auparavant, un effet, ou une absence d'effet de ce type de pratiques sur les carrières, qui a pu freiner un tournant numérique plus large en histoire ?*

Philippe Rygiel : Je crois en effet que ça n'a pas été valorisé par l'institution, c'est-à-dire que ça n'a aidé personne à faire carrière ces histoires-là, pendant une longue période. Jean-Luc ou Jean-Philippe comme moi, nous n'avons pas fait carrière par, ou grâce, ou à cause, de notre présence dans le champ numérique. On nous a pardonné un violon d'Ingres parce que nous étions par ailleurs des historiens sérieux qui produisaient des textes selon les canons de la profession. Je ne sais pas si Jean-Luc serait d'accord avec moi.

Jean-Luc Pinol : Oui, oui, tout à fait.

Jean-Philippe Genet : Disons que cela a pu être positif dans la mesure où cela valait brevet d'innovation et d'originalité. Mais en termes de carrière cet avantage n'a jamais compensé le fait que faire de l'histoire avec de l'informatique a représenté un investissement en temps considérable : si au niveau du CNRS et de la création de ce qui est devenu le LAMOP cela a été essentiel, cela a représenté un facteur de retard pénalisant pour ma carrière universitaire en retardant l'achèvement de ma thèse, et ma position universitaire a plutôt été appuyée sur le fait que j'ai été longtemps l'un des seuls médiévistes, sinon le seul, à travailler sur l'Angleterre médiévale.

EDGAR LEJEUNE : *Plus largement, et en renversant donc la question précédente, quels sont, de manière plus large, les impacts qu'ont pu avoir le numérique et le recours aux ordinateurs sur la discipline historique ? En France par exemple, mais peut-être aussi dans d'autres pays que vous connaissez bien ?*

Jean-Philippe Genet : On peut évaluer cet impact à trois niveaux, collectif, individuel et, mot peut-être mal choisi, épistémologique. Au niveau collectif, l'informatique a permis le développement de bases de données accessibles aux chercheurs, aux

enseignants et aux étudiants. Pour commencer par ce que je connais le mieux, au niveau du LAMOP nous développons dans le domaine prosopographique *Studium* (Maîtres et étudiants des écoles parisiennes) et nous avons joué un rôle déterminant (Hélène Millet) dans la création des *Fasti Ecclesiae Gallicanae* aujourd'hui gérés par le CERCOR à Saint-Étienne, dans le domaine spatial le projet ALPAGE (Hélène Noizet), et pour les humanités numérique le Corpus Burgundiae Medii Aevi (Elisa Magnani, Nicolas Perreaux), MEDITEXT et le projet e-Notre Dame. Comme je l'ai déjà signalé, cela a été facilité par la présence dans l'équipe de bons ingénieurs informaticiens et du PIREH, avec une relation interactive avec nos doctorants dont beaucoup sont bien formés. Mais il y a aussi les bases documentaires produites par l'École nationale des Chartes, l'IRHT et plus ou moins regroupées dans BIBLISSIMA, celles développées ou en cours de développement dans des laboratoires associés comme le CRAHAM, FRAMESPA etc., j'en oublie certainement. La variété et l'envergure de ces entreprises dépassent largement ce que l'on désigne en général par l'appellation « humanités numériques », trop centrée sur l'édition et la diffusion numérique.

Tout cela représente incontestablement un changement majeur dans le paysage scientifique et une évolution assez spectaculaire. Tout le contraire de ce qui se passe au plan individuel : mais si l'on regarde de près le contenu des thèses récentes, il y a très souvent derrière le texte des bases de données, dont l'élaboration a été rendue possible par l'amélioration de la formation des doctorants. En revanche, on ne sait pas ce que deviennent ces bases « heuristiques », le temps manque visiblement pour en faire une évaluation critique et un « nettoyage », ce qui est nécessaire pour les mettre en ligne et permettre l'accès à leurs contenus. Du coup, au troisième niveau, celui que je qualifie d'« épistémologique », il reste beaucoup à faire : l'approche pluridisciplinaire du texte et la textométrie, d'une façon générale le recours à la mesure quand l'on n'est pas en présence de données strictement sérielles, restent à un niveau généralement insatisfaisant. Les exigences de la profession sur le plan scientifique restent en deçà de ce qu'il faudrait et de ce que l'on pourrait faire. Et il reste des questions de fonds importantes pour les médiévistes, comme celles de l'incertitude et les problèmes nés de l'incomplétude des sources sur lesquelles beaucoup reste à faire.

Philippe Rygiel : Bon là, pour le coup, en France, c'est sans doute chez les médiévistes que ça a été le plus flagrant. Épistémologiquement, ça a été extrêmement important : ça a conduit à remettre en discussion des catégories comme le texte, comme le document, comme l'archive.

Jean-Luc Pinol : Beaucoup plus qu'en contemporaine, oui. Les historiens français, en particulier en contemporaine, il y en a beaucoup qui pensent aujourd'hui qu'ils utilisent l'informatique parce qu'ils sont sur Twitter. Mais en ce qui concerne les bases de données, ils ne savent presque pas ce que c'est...

Philippe Rygiel : Beaucoup plus qu'en contemporaine oui. Tout ce que Jean-Philippe a pu écrire sur la malléabilité ou la flexibilité ou l'instabilité du texte médiéval, ça c'est un acquis très très clair, très manifeste de cette histoire-là. En contemporaine en France, je pense... Vous savez le champ historique français est très particulier parce qu'il est protégé par des barrières douanières extrêmement fortes qui permettent aux historiens français, s'ils le désirent, de fonctionner en circuit fermé sans aucun souci de ce qui peut se passer ailleurs. Donc on peut avoir une inertie

assez forte. Mais dans un certain nombre de pays étrangers, on a des effets qui commencent à être structurants par exemple avec le surgissement de collectifs ou d'institutions de recherche qui sont organisés comme des laboratoires. C'est-à-dire qu'on a une logique d'industrialisation. Quand vous commencez à regarder un petit peu comment ça se passe en Corée où vous avez des chefs de projet, avec des ingénieurs rattachés, avec des collaborateurs qui ont des rôles seconds, en fait c'est la structuration de la profession en tant qu'espace de relations sociales qui s'en trouve modifiée. Et avec Paris Time Machine on a un peu cette importation de cette logique là en fait, où on est sur une organisation quasi industrielle dont la fonction est de produire des livrables historiques.

Donc je pense qu'on a cette tendance qui est à l'œuvre, je pense qu'elle va s'accroître, je dirais que pour le moment les impacts sont quand même extrêmement peu visibles dans le champ de la contemporaine en France. Ils le sont sensiblement plus à l'étranger. Il y a par exemple les travaux de Manning ou les travaux de Jo Guldi aux États-Unis par exemple, qui sont en train de montrer qu'on peut s'attaquer dans l'état actuel des sources disponibles et des technologies existantes, à des questions, à des échelles très vastes, en déployant des processus de recherche et pas simplement des discours. Je ne suis pas absolument persuadé de l'idée de Patrick Manning, qu'on va pouvoir faire une histoire globale de l'humanité, qui soit assise sur des données quantifiées, qui soient tout à fait convaincante, mais ça a des effets sur certaines formes d'histoire économique où on arrive à avoir des travaux sur de la longue durée, sur des phénomènes à de vastes échelles qui sont gérés par un travail sur du matériau et sur de la source, alors secondaire ou primaire ou agrégée. Mais ça c'était relativement impensable il y a 20 ou 30 ans. Donc oui, il y a des effets, mais je pense que ce n'est pas en France qu'on va les voir d'abord.

Edgar Lejeune : Pour conclure, une dernière question sur les enjeux à venir pour l'histoire numérique ?

Philippe Rygiel : Les enjeux à venir sont multiples. D'abord, on a des enjeux du côté des infrastructures. Et là, je pense que je renverrai la balle vers Jean-Luc, parce que le déploiement de ces histoires numériques, ça suppose toute une infrastructure qui est assez coûteuse, qui est assez complexe, à laquelle, alors, j'ai participé aussi par le biais du CollEx Persée sur la construction, la préservation, les conditions de circulation dans les corpus. Là on est vraiment dans une organisation industrielle, c'est-à-dire quelles sont les conditions concrètes qui permettent l'usage de ces ressources et l'exploitation optimale de ces ressources. Là nous sommes dans des questions de politique scientifique et de politiques publiques d'une certaine façon.

Jean-Luc Pinol : Oui tout à fait.

Philippe Rygiel : Du point de vue des pratiques des historiens, on est en train de parler beaucoup d'intelligence artificielle, c'est le terme à la mode. Ce qui est intéressant dans Chat-GPT, ou la dizaine d'intelligence artificielle de troisième génération qui vient d'être livrée au grand public, finalement, ce n'est pas terriblement la capacité à générer des textes, parce que ça, c'est des choses qu'on connaît en labo depuis 10 ou 15 ans : ce ne sont pas des algorithmes récents, mais ce sont toutes les choses qui vont permettre de repérer et de générer des formes ou des séquences dans de très très vastes collections de sources et de données. Par exemple, on doit pouvoir, dans deux ans, cinq ans, dix ans, faire des graphes de relations à

partir des textes numérisés par Gallica qui sont en ligne gratuitement, pour reconstituer, je ne sais pas, les affinités littéraires à partir d'une revue littéraire, ou étudier des différences subtiles dans les usages de la langue dans un certain nombre de produits culturels. Là d'une certaine façon, on n'y est pas tout à fait, mais les algorithmes sont là, la puissance de calcul est là, ça frotte encore un peu sur l'accessibilité et l'usage de la donnée. Mais il y a eu une flopée de questions inédites ou de questions posées à des échelles différentes qu'on peut très très bien imaginer voir surgir.

Jean-Philippe Genet : Il y a aussi un besoin essentiel de formation, d'une part parce que la formation initiale des historiens reste encore largement déficiente, d'autre part parce que l'importance de la formation continue n'est pas reconnue. Les UFR d'histoire ou de sciences sociales ont sous-estimé les besoins et il est désolant de voir que lorsque, pour éviter que leurs postes soient attribués à d'autres disciplines par les conseils d'université, elles les « colorent » de numérique, elles laissent bien trop souvent les jurys élire des candidats sur des critères purement disciplinaires sans même prendre en considération les compétences des candidats dans le domaine du numérique ! Il y a tout de même une évolution importante grâce au financement des projets de recherche par appel à projet. La formule est très critiquée, mais elle a deux mérites : elle oblige à sortir des logiques individuelles pour concevoir des projets collectifs qui incluent très souvent des projets informatiques. Certains sont parfois purement cosmétiques, mais cela crée tout de même un effet d'entraînement, et ces projets peuvent favoriser un brassage de compétences pluridisciplinaires qui donne l'occasion aux historiens qui y participent soit de se former sur le tas, soit de se mettre à jour pour suivre le changement très rapide des technologies et des logiciels. Le gros point noir reste l'enseignement de l'histoire dans les UFR qui n'accorde pas à l'informatique (et aux méthodes linguistiques et statistiques dont elle facilite l'accès) la place que l'évolution actuelle des technologies de la communication rend nécessaire.

Philippe Rygiel : Je suis d'accord avec Jean-Philippe, mais ce n'est pas simple parce que fondamentalement le numérique, c'est des maths et de l'informatique. Et nous avons une tradition nationale qui fait qu'on est soit littéraire, soit scientifique. Et les canards à cinq pattes ne sont pas si nombreux que ça. Il y en a, mais il n'y en a pas tant que ça. Et on se heurte quand même à des difficultés qui sont réelles quand il s'agit de former des gens qui soient agiles dans cet univers-là, d'autant qu'on ne peut pas les former à des procédures logicielles parce qu'au bout de 5 ans, le logiciel est mort. Donc il faut les former d'une façon qui soit plus fondamentale et plus théorique, de façon à ce qu'ils soient adaptables. Et là il y a une vraie barrière d'entrée et il n'est pas facile de savoir ce qui serait le plus efficace. Et là pour le coup je crois que ça mériterait un entretien complet.

Jean-Luc Pinol : Oui, et puis je crois que c'est bien comme mot de la fin.

AUTEUR

EDGAR LEJEUNE

Chercheur post-doctoral (CAK – EHESS & CNRS) et chercheur associé (SPHERE – Université Paris-Cité & CNRS. edgarlejeune@hotmail.fr)