

ATHIS

IV^e Atelier Informatique et Histoire

L'informatique et l'utilisation des statistiques par les historiens

*28-29 septembre 2007 – Ecole Normale Supérieure LSH
Lyon*

Le programme ATHIS, financé par l'Agence Nationale pour la Recherche, a pour ambition de faire le bilan prospectif des apports des méthodes dont l'informatique a ouvert l'accès aux historiens et, d'une façon générale, d'évaluer l'ampleur des transformations que le recours à l'informatique introduit dans le métier d'historien. Il prend appui sur deux réseaux de médiévistes, *Reti Medievali* (Andrea Zorzi), à l'Université de Florence et le portail *Ménestrel* (représenté par Christine Ducourtieux) en France. Les laboratoires français qui en sont les opérateurs sont le CRHAM de Caen (représenté par Pierre Bauduin), le Centre d'Histoire du XX^e siècle de Paris I (représenté par Philippe Rygiel), l'Ecole Nationale des Chartes (représentée par Marc Smith), l'I.R.H.T. (représenté par Paul Bertrand) et le LAMOP de Paris I (représenté par Jean-Philippe Genet) qui coordonne l'ensemble du projet. L'École Française de Rome (représentée par Marilyn Nicoud) participe à l'ensemble du projet et le soutient scientifiquement et financièrement.

Le cinquième atelier est consacré à l'impact de l'informatique sur l'utilisation des méthodes statistiques par les historiens. Lorsque l'informatique a fait son apparition dans le champ de l'histoire, à la fin des années soixante, les moyens de calcul auxquels elle donnait enfin accès ont été l'une des incitations majeures qui ont poussé les historiens à explorer ses possibilités : de nombreux travaux américains, italiens et français en témoignent, et le plus éloquent témoin de cette période pionnière est peut-être l'étude du *Catasto* florentin de 1427 entreprise par David Herlihy et Christiane Klapisch, qui reste encore aujourd'hui un modèle pour beaucoup d'historiens. C'est d'ailleurs au cours de cette recherche que les méthodes de l'analyse factorielle (grâce à Michel Demonet) et celle de la cartographie automatique (grâce à Jacques Bertin) ont fait leur apparition dans la pratique des historiens. C'est aussi vers cette époque que les méthodes de la lexicologie statistique – que nous avons déjà partiellement explorées lors de l'atelier II à l'École Normale supérieure de Lyon – ont commencé à être développées à Saint-Cloud, Nancy et Besançon en France, et à l'École Normale Supérieure à Pise en Italie. Enfin, les nouvelles méthodes permettaient d'introduire la mesure et les méthodes quantitatives dans des domaines de l'histoire où jusqu'ici on ne les avait pas employées en traitant des données qualitatives (par exemple l'histoire culturelle).

Cet intérêt pour les méthodes quantitatives appliquées à l'histoire est resté très soutenu jusqu'aux années quatre-vingt : en témoigne encore en France la création par le C.N.R.S. de la revue *Histoire & Mesure*, toujours active aujourd'hui. Mais il est incontestable que, par la suite, cet intérêt s'est ralenti quand il ne s'est pas complètement tari, au moment où les machines (ordinateurs personnels) et les logiciels (packages de traitements statistiques ou de dépouillement d'enquête) étaient à la fois plus accessibles et plus faciles à manier.

Paradoxalement, le progrès matériel a fonctionné comme un frein au développement intellectuel : les longues périodes d'attente dans les centres de calcul, où l'on travaillait sur des *mainframes*, étaient en fait des moments privilégiés de discussion et d'échange, qui permettaient de découvrir, d'expérimenter et d'approfondir les méthodes utilisées par les chercheurs des autres disciplines ou, pour les historiens, des autres périodes. Dans le cas précis des statistiques, ces moments étaient d'autant plus importants que la formation statistique des historiens, qui n'avait jamais été très poussée régressait, les historiens de l'économie perdant peu à peu du terrain face aux économistes historiens, et la démographie ayant tendance à quitter le giron de l'histoire pour s'autonomiser ou pour devenir une science auxiliaire de la sociologie. L'accent de plus en plus résolument mis sur le raccourcissement du temps de préparation de la thèse a aussi détourné les historiens d'un investissement jugé trop lourd dans l'apprentissage méthodologique.

Ceci illustre bien l'un des thèmes conducteurs de ces ateliers : c'est moins l'informatique en elle-même que les méthodes auxquelles elle donne accès qui sont susceptibles de transformer les pratiques scientifiques des historiens. Aujourd'hui, il semble que les historiens ont pris conscience des enjeux : des filières nouvelles de formation sont apparues, les méthodes testées sur des données historiques se sont multipliées, des articles présentant ces expériences ont recommencé à paraître dans les revues. Avec l'arrivée d'internet, des outils encore plus performants et encore plus maniables, destinés aux historiens (le manuel de statistique d'Alain Guerreau sur le site de l'École Nationale des Chartes, ou celui, signalé par Alain Guerreau, de R.H. Bayen, *Analyzing linguistic data. A practical introduction to statistics*) ou visant un public beaucoup plus large (des logiciels comme TRIDEUX ou R) sont apparus et sont devenus très facilement accessibles. Et dans des disciplines proches de l'histoire (sociologie, sciences politiques etc.), des méthodes nouvelles ont fait leur apparition. Le temps paraît donc venu de faire un tour d'horizon des utilisations de la statistique par l'intermédiaire de l'informatique dans le domaine historique et de dresser un nouvel état des lieux.

Les communications doivent être très brèves, vingt à trente minutes, pour permettre d'amples discussions. Elles ne sont pas destinées à être publiées sur un support imprimé, mais à être diffusées sur internet, via *Reti Medievali* et *Méneştrel*. Un rapport synthétisant les débats sera présenté par les organisateurs dans les *Mélanges de l'École Française de Rome*.

Programme provisoire

Vendredi 28 septembre : matinée

9 h. Accueil des participants

9 h. 30 Bienvenue de la Direction de l'ENS

9 h. 35 Jean-Philippe Genet (LAMOP, Paris I – CNRS), Introduction

La pratique historique face aux méthodes statistiques

9 h. 45 Alain Guerreau (CRH – CNRS), « L'historien et les méthodes statistiques : d'un difficile démarrage à la nouvelle donne des logiciels libres ».

10 h. 30 François Djindjian, « L'analyse des données en archéologie: techniques statistiques, methods archéologiques. Un essai de bilan 35 ans après la révolution des années 70 ».

11 h. Pause café

11 h. 15 Renzo Derosas (Université Ca Foscari, Venise), « The historian and the new statistical tools: a poverty of theory ».

11 h. 45 : Discussion

12 h. 30 Repas à l'École Normale Supérieure.

Vendredi 28 septembre : après-midi

La pratique de l'analyse factorielle

14 h. Séverine Lepape, « Une application de l'analyse factorielle à l'iconographie de l'arbre de Jessé ».

14 h. 30 Discussion.

Distributions et méthodes de mesure

14 h. 45 Pascal Chareille (Tours), « Les mesures de distance en anthroponymie historique ».

15 h. 15 Philippe Rygiel (Paris I), « La mesure de la mobilité sociale : les méthodes statistiques adaptées aux petits effectifs ».

15 h 45 Laurent Lesnard (CREST, INSEE), « Décrire les rythmes sociaux avec les méthodes d'appariement optimal ».

16 h. 15 Pause café

16 h. 30 Marie Cottrell (Paris I) : « Visualisation des données de grandes dimensions au moyen de l'algorithme de Kohonen ».

17 h. Patrice Gaubert (Paris XII), « Quelques algorithmes adaptés au traitement des données historiques : réseau de neurones, modèles markoviens, tests de rupture ».

17 h. 30 Discussion

18 h. 15 Fin des travaux

20 h. Dîner de l'atelier.

Samedi 29 septembre : matinée

Espace, représentation spatiale et statistiques

9 h. Guillaume Daudin (Sc. Po), « Méthodes de classification et cartographie statistique ».

9 h. 30 Jean-Daniel Fekete (INRIA et Paris-Sud Orsay) et Nicole Dufournaud (CESR, Tours et EHESS Paris), « Visualisation exploratoire des données historiques ».

10 h. Vincent Loonis (CREST, INSEE), « Statistique spatiale et historique ».

10 h. 30 Pause café

10 h. 45 Brigitte Leroux, « Analyse géométrique des données structurées : application au champ du pouvoir norvégien ».

11 h. 15 Claire Lemercier (ENS, Paris) [et Paul-André Rosental], « Une application de l'analyse des réseaux à l'histoire des migrations ».

11 h 45. Discussion

12 h. 15 Repas à l'École Normale Supérieure.

Samedi 29 septembre : après-midi

Régression, causalité et modélisation

13 h. 30 Philippe Cibois (St.Quentin en Yvelines), « La régression logistique par le biais de l'analyse tabulaire multivariée ».

14 h. Anne-Sophie Bruno, « Analyser les différences de salaire : les vertus des modèles de régression multi-niveaux ».

14 h. 30 Noël Bonneuil : « Systèmes dynamiques appliqués à la démographie historique ».

15 h. Serge Heiden : « La modélisation des phénomènes linguistiques ».

15 h30. Discussion

16 h. Alain Dallo (coordinateur), Julien Alérini, Benjamin Deruelle, Stéphane Lamassé, Frédéric Saly : « En guise de conclusion : les réactions d'une équipe d'historiens ».

Liste provisoire des participants :

Julien Alerini (Paris I)
Noël Bonneuil (INED)
Anne-Sophie Bruno (EHESS)
Pascal Chareille (Tours)
Philippe Cibois (U. Saint-Quentin en Yvelines)
Alain Dallo (Paris I)
Guillaume Daudin (Sciences Politiques)
Renzo Derosas (Université de Venise)
François Djindjian
Nicole Dufournaud
Benjamin Deruelle (Paris I)
Jean-Daniel Fekete (Paris-Sud)
Patrick Gaubert (Paris XII)
Jean-Philippe Genet (Paris I)
Alain Guerreau (CRH – CNRS)
Serge Heiden (ENS – LSH, Lyon)
Stéphane Lamassé (Paris I)
Claire Lemerrier (ENS, Paris)
Séverine Lepape
Brigitte Leroux (Paris V)
Laurent Lesnard (CREST, INSEE)
Vincent Loonis (CREST, INSEE)
Marilyn Nicoud (Ecole Française de Rome)
Philippe Rygiel (Paris I)
Frédéric Saly (Paris I)
Claire Zalc (ENS, Paris)