

ATHIS

II^e Atelier Informatique et Histoire

L'historien, le texte et l'ordinateur

27-28 novembre 2006 – École Normale LSH de 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*, à l'Université de Florence (Andrea Zorzi) 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'École Nationale des Chartes (représenté 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.

Le second atelier est consacré aux développements des méthodes d'analyse du texte rendues possibles par l'informatique et, par voie de conséquence, aux exigences nouvelles que l'historien se doit désormais d'avoir en matière de présentation, de traitement, et d'interprétation du texte. C'est au début des années soixante – il y a donc plus de quarante ans – que les historiens ont commencé à s'intéresser à ces problèmes, en même temps qu'aux possibilités offertes par la statistique lexicale, dans le cadre des travaux réalisés à l'École Normale Supérieure de Saint-Cloud, dans le sillage de Robert-Léon Wagner et de Maurice Tournier. De son côté, Antoine Prost a été l'un des premiers historiens à réaliser le potentiel de ces méthodes. En Italie, c'est surtout au CNUCE à Pise que ces méthodes se sont développées grâce à Antonio Zampolli qui dès 1960 travaillait à Gallarate avec le Père Busa sur l'indexation de l'œuvre de Saint Thomas d'Aquin. Zampolli a ensuite joué un rôle déterminant dans la naissance du traitement automatique du langage, et dans celle du concept de ressources linguistiques. En France, des travaux importants ont ensuite été réalisés, notamment dans le domaine de l'histoire de la Révolution Française (J. Guilhaumou) et dans celui de l'histoire des mouvements syndicaux. Mais il semble que l'évolution méthodologique chez les historiens est ensuite allée dans une autre direction : alors que les politologues et les sociologues ont décidément adopté ces méthodes, les historiens, dont certains ont d'ailleurs préféré rejoindre des équipes de linguistes, n'ont guère progressé, si bien qu'ils ne sont plus totalement en phase aujourd'hui avec « l'état de l'art ». Les méthodes de lexicologie et de statistique lexicale sont en effet aujourd'hui intégrées pour nombre d'entre elles dans ce que l'on pourrait appeler la linguistique de corpus : mais il ne faudrait pas que certaines exigences apparues au fur et à mesure des nouvelles possibilités offertes par les machines, telles que le *tagging*, ou la lemmatisation, détournent les historiens de ces méthodes, alors même qu'il est maintenant infiniment plus facile de constituer des corpus riches et variés que ce n'était le cas quand l'informatique balbutiait.

Parallèlement, des recherches se sont développées dans un secteur que l'on pourrait qualifier de philologie numérique. Il ne s'agit pas tant des problèmes d'édition (qui ont fait l'objet de notre première rencontre à l'École Française de Rome) que de méthodes qui permettent de mieux comprendre et de mieux étudier les textes. Il en va ainsi de ces méthodes qui permettent de reconnaître les différentes parties du discours, pour les étudier et les

comparer plus facilement. De même, le brouillon, la variante, le remords et la rature peuvent être pris en compte. La linguistique de corpus rejoint d'autre part des traitements spécifiques pour permettre une étude sémantique des textes en prenant en compte l'ensemble du vocabulaire. De multiples possibilités sont ainsi apparues, qui sont loin d'être toutes utilisées par les historiens, dont on attendrait plutôt qu'ils manifestent leurs besoins face aux textes.

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 *Ménestrel*. Des mini-tables rondes sur certains sujets permettront des débats et seront soutenues par de brèves contributions écrites des participants. Un rapport sera présenté par les organisateurs dans les *Mélanges de l'Ecole Française de Rome*.

Programme

Lundi 27 novembre

9 h. 30 Accueil des participants

9 h. 45 Olivier Faron, directeur de l'ENS-LSH : allocation de bienvenue.
Jean-Philippe Genet, « Introduction »

La linguistique de corpus

Problèmes généraux :

10h. Christiane Marchello-Nizia : « L'apport de la linguistique de corpus aux analyses historiques : corpus, outils, problématiques ».

10 h. 30 Philippe Cibois : « Analyse de données textuelles et analyse de contenu : une perspective de sociologue ».

11h. Pause-café

11h. 15 Adeline Nazarenko : « La fouille de textes est-elle utilisable pour l'historien ? »

11h 45 Discussion.

12 h. 30 Déjeuner à l'Ecole.

14 h. Jacques Guilhaumou : « À propos du dictionnaire des usages socio-politiques ».

Exemples de corpus :

14 h. 30 J.Cl. Zancarini : « De l'hyperprince à l'hypermachiavel ».

15 h. Aude Mairey : « La construction d'un corpus (les prologues anglais et le vocabulaire de la connaissance) ».

15 h. 30 Paolo Mastandrea (Venise) : « Les archives de 'PoetriaNova' et 'PoetidItalia' comme outils pour une recherche historique ».

16 h. Isabelle Draelants (Nancy II) : « Le corpus annoté sur les encyclopédies médiévales de l'Atelier Vincent de Beauvais ».

16 h. 30 Discussion

17h. Pause-café

17 h. 15 – 19 h.15 : Table ronde

<p>Les problèmes de la lemmatisation : modérateur, Monique Goulet (LAMOP) Joseph Denooz (Liège), Sylvie Mellet (Nice), Caroline Philippart de Foy (Liège/Nice)</p>

20 h. Dîner du colloque.

Mardi 28 novembre

L'informatique et les métamorphoses du texte

9 h. Jean-Louis Lebrave : « L'informatique, les brouillons et les variantes ».

9 h. 30 Raul Mordenti (Roma Tor Vergata), « Problemi di filologia digitale a partire dallo *Zibaldone Laurenziano* di Boccaccio ».

10h. Pause-café.

10 h. 15 – 12 h. 40 : Table ronde

"Nouveaux outils et nouvelles perspectives pour l'utilisation des corpus par les historiens" (UMR 5648 CIHAM)

modérateur, Jacques Chiffolleau

Nicole Bériou, Jacques Berlioz, Marjorie Burghart, Christian Guilleré, Janine Mathieu,
Gauthier Poupeau

12 h. 45 Déjeuner à l'École

L'apport des traitements lexicographiques et lexicométriques

14 h. Antonio De Prisco et Stefano Minozzi (Vérone), « Metodi computazionali per l'indagine lessicale su testi latini medievali. Un' applicazione al Codex Diplomaticus Cavensis ».

14 h. 30 Pierre Lafon (Lyon), « Statistiques et lexicométrie : position des problèmes ».

15 h. Damon Mayaffre : « L'apport de la lexicométrie à la compréhension historique des textes politiques ».

15 h. 30 Pause-café

15 h. 45 – 17 h. 45 Table ronde

Les logiciels de traitement informatique du texte : modérateur, Alain Dallo (LAMOP)

Serge Heiden (Lyon), Pierre Lafon (Lyon), André Salem (Paris III)

17 h. 45 Gianmaria Varanini : « Conclusion des travaux du colloque ».

20 h. Dîner pour les personnes restant à Lyon.