



HAL
open science

Médiévistes et ordinateurs : organisations collectives, pratiques des sources et conséquences historiographiques (1966-1990)

Edgar Lejeune

► To cite this version:

Edgar Lejeune. Médiévistes et ordinateurs : organisations collectives, pratiques des sources et conséquences historiographiques (1966-1990). Philosophie. Université Paris Cité, 2021. Français. NNT : 2021UNIP7095 . tel-03682773v2

HAL Id: tel-03682773

<https://theses.hal.science/tel-03682773v2>

Submitted on 31 May 2022

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.

Université de Paris
École doctorale 623 – Savoirs, Sciences, Éducation
Laboratoire SPHERE (UMR 7219)
et LATTICE (UMR 8094)

Médiévistes et ordinateurs.

Organisations collectives, pratiques des sources et conséquences historiographiques (1966-1990).

Par Edgar Lejeune

Thèse de doctorat d'Histoire et de Philosophie des Sciences

Dirigée par Karine Chemla, SPHERE (CNRS – Université de Paris)
et par Shirley Carter-Thomas, LATTICE (CNRS – École Normale Supérieure
– Sorbonne Nouvelle)

Présentée et soutenue publiquement le 12 octobre 2021

Devant un jury composé de :

Rens Bod, Professor, Universiteit van Amsterdam, rapporteur
Lorraine Daston, Professor, Max Planck Institute for the History of Science, Berlin, and
The University of Chicago, rapporteuse
Jacqueline Léon, Directrice de recherche émérite, CNRS & Université de Paris (UMR 7597)
Joseph Morsel, Professeur des Universités, CNRS & Université Paris 1 Panthéon-Sorbonne
(UMR 8589)
Pierre Mounier-Kuhn, Chargé de recherche, Centre Roland Mousnier CNRS & Sorbonne
Université (UMR 8596), chercheur associé, Centre Alexandre Koyré-CRHST
Nathalie Richard, Professeur des Universités, CNRS & Le Mans Université (UMR 9016)



Titre : Médiévistes et ordinateurs. Organisations collectives, pratiques des sources et conséquences historiographiques (1966-1990).

Résumé :

Dès la fin des années 1950, l'irruption des calculateurs électroniques dans les sciences humaines et sociales a profondément affecté les pratiques des chercheurs qui se sont saisis de ces nouveaux outils de classement et de calcul. Cette thèse vise à analyser certaines transformations historiographiques que ces innovations techniques ont provoquées, en se concentrant sur un large groupe d'historiens médiévistes mettant en œuvre ces instruments pour mener leurs recherches entre 1966 et 1990.

Mes problématiques s'organisent autour de deux préoccupations principales. Tout d'abord, ce travail s'interroge sur l'apparition de nouvelles formes d'organisation collective de la recherche, à l'échelle des équipes de recherche comme de la discipline elle-même. Ma thèse montre comment ces formes s'articulent à la production, à la manipulation et à la mise en circulation de nouveaux types de textes (bordereaux de saisies, cartes perforées, programmes informatiques, manuels de codage, mais également bulletin de liaison).

De plus, cette thèse se propose d'analyser les transformations des méthodes des historiens liées à l'utilisation des moyens électroniques. J'y mets en évidence que le recours aux calculateurs électroniques et aux ensembles de savoirs associés à leurs utilisations (analyse de données, analyse automatique de textes, informatique documentaire) impliquait l'emploi de technologies intellectuelles (matrices, graphes, listes, index, inventaires, thesaurus, etc...) qui requéraient une formalisation accrue des opérations de recherche des médiévistes, tout en étant dotées, dans ce contexte de fonctions originales.

Pour travailler à ces problématiques, la méthodologie développée dans ce travail repose sur deux convictions. Tout d'abord, j'y développe l'idée qu'il faut appliquer à l'histoire des sciences humaines et sociales les méthodes développées par les historiens des sciences. Ensuite, celle qu'il est nécessaire, pour saisir les pratiques des historiens, d'étudier les textes sur lesquels reposent ces pratiques. Cette seconde direction nous a

conduit à emprunter des méthodes de recherche à la linguistique, et en particulier à l'analyse de discours.

Cette thèse est ainsi construite en trois parties. Dans la première, je propose une analyse de deux projets de recherches menés par des médiévistes français entre 1966 et 1990, aux fins de comparer leurs organisations collectives (taille des équipes, présence d'informaticien, type de financement, outils de calculs employés), les influences extra-disciplinaires qu'ils subissent dans l'élaboration de leurs méthodes (démographie, géographie, linguistique, sociologie), mais aussi les méthodes qu'ils mettent en œuvre (lexicométrie, histoire quantitative) et les conséquences historiographiques de ces travaux.

Dans la deuxième partie de cette thèse, je traite des processus par lesquels certains collectifs engagés dans l'utilisation des ordinateurs se sont organisés à une échelle méta-collective, dans le but de faire circuler les produits de leurs recherches (éditions, bases de données, programmes informatiques, etc...) et des méthodes et des techniques qu'ils avaient pu développer au cours de leurs expériences respectives. L'analyse porte sur une initiative française, à portée européenne, qui démarre avec l'organisation d'un colloque à Rome en 1975 et se poursuit par la mise en œuvre d'une publication en 1979, intitulé *Le Médiéviste et l'ordinateur*. Ce bulletin de liaison devient, dès lors, l'un des vecteurs privilégiés de la circulation des méthodes et de la construction d'une culture scientifique commune.

La troisième et dernière partie est quant à elle consacrée aux modalités des échanges qui prennent forme dans les pages de ce bulletin de liaison. Deux directions sont explorées : 1) le genre d'article qui s'y développe en relation avec la nécessité de faire circuler de nouveaux types de connaissances techniques et 2) les difficultés qu'ont pu rencontrer les utilisateurs de ces méthodes dans le partage de ce type de savoir avec certains de leurs collègues, d'un point de vue technique, mais également épistémologique.

Mots-clefs : Histoire des humanités numériques - Historiographie - Histoire des textes - Histoire des sciences - Linguistique

Title: Medievalists and computers. Collective organizations, practices of historical sources and historiographical consequences (1966-1990).

Abstract:

From the late 1950's onwards, the rapid emergence and growing use of computers in the humanities brought about many changes in the practices of researchers. My PhD dissertation sets out to analyze some of the historiographical transformations caused by these technical innovations. With this aim in mind, I focus on a small group of medievalist historians who used these tools for research purposes between 1966 and 1990.

This thesis deals with two main issues. Firstly, it looks at how, in that context, new forms of collective organization emerge at different scales, beginning inside the research team and extending to encompass the whole discipline. My dissertation shows how these forms are articulated with the production, manipulation and circulation of new types of texts (coding sheets, punched cards, computer programs, coding books, but also liaison bulletins). Secondly, it deals with the ways in which the historians' methods are transformed in correlation with the operation of the computers. In this perspective, I highlight the fact that these new technologies and the knowledge associated with them (data analysis, automatic text-processing, computer science, etc.) entailed to work with a set of intellectual technologies (matrices, graphs, lists, indexes, inventories, thesaurus, etc.) that required a strengthened formalization in each of the activities associated with the medievalists' research. Moreover, each of these intellectual technologies was, in this specific context, endowed with its own original functions.

The methodology I develop is underpinned by two firm beliefs. While I acknowledge the idea that the historian of the humanities has to apply to his objects the methods developed by historians of science, I also show that it could be worthwhile, in order to grasp the realities of the historians' research practices, to focus on the texts they produced, handled and exchanged. This second approach requires us to borrow some methods of research from linguistics, and in particular from discourse analysis.

This study is divided into three parts. In part one, I analyze two French medievalist projects in 1966-1990, so as to compare their collective organizations (manpower, computer science collaborators, funding, computing tools), the extra-disciplinary

influences exerted on them (demographic, geographical, linguistic, sociological), the methods they developed (lexicometry, quantitative history) and their historiographical consequences.

In part two, I expand on the process by which some of these entities set up a meta-collective organization, so they could exchange the fruits of their research (computer editions, databases, computer programs, etc.) as well as the methods and techniques they had developed on their specific projects. The object of the analysis is a French-based initiative forming part of a Europe-wide project, which was launched at a colloquium in Rome in 1975. A few years later, in 1979, the same organizers published the first European periodical dedicated to this subject, the liaison bulletin *Le Médiéviste et l'ordinateur*. This rapidly became one of the main media for the dissemination of methods and the construction of a common scientific culture.

The third part of my dissertation is dedicated to the modes of exchanges that took shape within the pages of the bulletin. Two different directions are explored: 1) the genre of papers born out of the newfound need to share computer-related technical knowledge and 2) the difficulties that some of the authors encountered at a technical and epistemological level in the process of sharing this knowledge with their colleagues.

Keywords: History of digital humanities - Historiography - History of Texts - History of Science – Linguistic

Remerciements.

Les remerciements que je tente de formuler ici ne sauraient à eux seuls exprimer toute ma gratitude envers tous ceux qui m'ont encouragé, découragé, soutenu, critiqué, félicité, écouté, repris, lu, relu, conseillé, invité, accueilli, contacté, écrit, posé des questions, etc.

*

**

Je tiens tout d'abord à remercier chacun des membres du jury de cette thèse pour avoir accepté d'accorder son temps et son attention à mon travail. Toute ma gratitude va ainsi à Rens Bod, Lorraine Daston, Jacqueline Léon, Joseph Morsel, Pierre Mounier-Kuhn et Nathalie Richard.

J'adresse également toute ma reconnaissance aux membres de mon comité de suivi de thèse que je n'ai pas conviés dans le jury, mais qui m'ont attentivement guidé vers les problématiques qui sont les miennes dans ce présent travail : Claire Lemerrier et Maarten Bullynck.

*

**

Si ces cinq dernières années ont pu aboutir au résultat que vous avez aujourd'hui sous les yeux, je le dois essentiellement à mes deux directrices de thèses, Karine Chemla et Shirley Carter-Thomas, qui m'ont apporté toute l'aide et tout le soutien que peut espérer un étudiant en doctorat. Je veux les remercier pour leur confiance dans mon projet de recherche, les conseils précieux qu'elles m'ont prodigués depuis le début de ce travail et leurs relectures infiniment attentives et intelligentes. Cette thèse leur doit énormément, mais il en va également de ma formation intellectuelle qui, je peux l'affirmer aujourd'hui, n'était absolument pas complète avant de travailler avec elles.

*

**

Je suis par ailleurs redevable de toutes les institutions qui ont permis à cette thèse de se réaliser dans des conditions matérielles acceptables, et de tous les personnels administratifs qui ont contribué au fonctionnement de ces institutions : feu l'université Paris-Diderot Paris 7 et son École Doctorale 400, au sein de laquelle j'ai obtenu un contrat doctoral en 2016, l'université de Paris et son École Doctorale 623, Pôle Emploi, le laboratoire SPHERE, le laboratoire LATTICE ainsi que les organismes qui m'ont accueilli ou qui ont financé mes séjours à l'étranger, riches d'enseignements et de rencontres : le Deutscher Akademischer Austauschdienst (DAAD), le Max-Planck-Institut für Wissenschaftsgeschichte et l'École Française de Rome. Je remercie également, à ce titre, Jens Høyrup pour son accueil chaleureux à Rome au début de l'année 2020.

Dans le même ordre d'idée, je remercie tous ceux qui m'ont aidé à accéder à des archives et à des documents à même de m'aider à traiter mon sujet, que ce soit à l'EHESS grâce à Christiane Klapisch-Zuber, Nicolas Veysset et Béatrice Marin, à l'université Paris 1 Panthéon Sorbonne grâce à Jean-Philippe Genet, Stéphane Lamassé et Marie-Caroline Luce ou à l'Institut de Recherche et d'Histoire des Textes grâce à Caroline Bourlet et Sandra Labouré.

Je profite également de cette section pour remercier tous les historiens, sociologues, mathématiciens, ingénieurs de recherche et archivistes qui ont accepté de me rencontrer pour discuter de leur travail durant ces cinq années, mais également ceux que j'ai eu la chance de croiser dans d'autres circonstances : Gérard Béaur, Philippe Cibois, Michel Demonet, Alain Guerreau, Jacques Guilhaumou, Jean Heffer, Béatrice Joyeux-Prunel, Gilles Kagan, Michel Kaplan, Olivier Mattéoni, Hélène Millet, Joseph Morsel, Nicolas Offenstadt, Ezio Ornato, Marie-Anne Polo de Beaulieu, Antoine Prost, Nathalie Queyroux, Jacques Revel, André Salem, Manfred Thaller et Françoise Waquet.

*
**

Au fil de ces cinq années de recherches, insistons encore sur le fait que les belles rencontres ont été nombreuses. Avec mes collègues doctorants de SPHERE tout d'abord, et plus particulièrement ceux avec lesquels nous avons créé le séminaire des doctorants et partagé de nombreuses discussions passionnantes et de nombreux repas animés : Pierre

Chaigneau, Babette Chambou-Combaz, Helmy Chekir, Mourtaza Chopra, Thomas Crane, Gautier Depambour, Benoît Duchemann, Sarah Ermann, Justin Gabriel, Sarah Hijmans, Marie Lacomme, Yang Liqiong, Nicolas Michel, Arilès Remaki, Adeline Reynaud, Miriam Rogasch, Alexis Trouillot, David Waszek et bien d'autres. J'aimerais ici adresser un remerciement plus particulier à Nicolas Michel, qui a pris le temps de relire, toujours avec une grande attention et une grande intelligence, plusieurs parties de ma thèse à des moments où cela me paraissait utile d'avoir un avis différent de ceux de mes directrices.

Dans la même direction, j'aimerais remercier tous les membres du séminaire d'histoire des historiens que nous avons initié avec Vincent Heimendinger en 2018, et dont l'aide a été précieuse à bien des égards : Benjamin Laillier, Sam Racheboeuf, Baptiste Roger-Lacan, Yuri Mourao Falcao, Christophe Araujo, Delphine Frasch et Céline Lориou.

Cette liste ne serait pas complète si je n'y ajoutais pas les jeunes chercheurs que j'ai rencontrés au cours de mes nombreuses activités académiques en dehors des murs du laboratoire SPHERE : Giovanni Colavizza à Oxford, Sjang ten Hagen, Laura Sumrall et Guillaume Yon à Berlin, Gabriel Alcaras, Constantin Brissaud, Timothée Erard et Antonin Perdoncin à Paris-Science-Lettre, Adeline Blaszkiewicz, Léonard Laborie, Nicolas Truffinet et Lucie Azens à l'université Paris 1 Panthéon-Sorbonne et Angela Cossu, Philippe Lefeuvre et Nicolas Minvielle-Larousse à Rome.

*
**

Un grand merci également aux chercheurs qui animent les séminaires des laboratoires de SPHERE pour m'avoir invité à proposer des communications et des séances dès ma première année de doctorat. Je tiens à mentionner tout particulièrement les deux organisatrices du séminaire « Approches historiques, philosophiques et anthropologiques des nombres, de la mesure et de la mesurabilité », Christine Proust et Nadine de Courtenay. Ma gratitude va également aux organisateurs et aux membres d'un autre séminaire du laboratoire SPHERE, constitué autour du groupe de recherche « Histoire des sciences, histoire du texte », parmi lesquelles Karine Chemla et Christine Proust, que j'ai déjà citées, mais aussi Agathe Keller, Martha Cecilia Bustamante,

Florence Bretelle-Establet, Julie Lefebvre, Emmylou Haffner, Stéphane Schmitt et Matthieu Husson. Ce séminaire restera pour moi à la fois l'un des plus passionnants et des plus conviviaux auxquels j'ai pu participer, et je conseille vivement à tous les lecteurs de ces remerciements de s'y faire inviter.

Un grand merci également aux équipes du second laboratoire qui m'a accueilli au cours de ces recherches, le laboratoire LATTICE, et en particulier à Frédérique Mélanie et Clément Plancq, pour l'aide très précieuse qu'ils m'ont apportée dans l'analyse de certains de mes corpus. Merci à Laure Sarda qui m'a invité à présenter mon travail dans le cadre du séminaire du laboratoire, et également à tous ceux avec lesquels j'ai pu échanger plus ou moins longuement lors de mes brefs passages dans les locaux de Montrouge : Frédéric Landragin, Benjamin Fagard, Michel Charolles, Fabien Garrido, Auphémie Ferreira, Claire Kloppman et Amine Lahouli.

*

Enfin, ces remerciements ne seraient pas complets si je n'ajoutais pas un mot pour tous ceux qui m'ont permis de garder une partie de la tête en dehors du monde académique pendant cette période : mes amis, mes parents, mon frère, ma sœur, ma grand-mère, Maria, Nonna et Gina, sans qui je n'aurais jamais eu tout le courage nécessaire pour aller jusqu'au bout de cette thèse en gardant le sourire.

Liste des principales abréviations.

ADDAD : Association pour le Développement et la Diffusion de l'Analyse des Données

Annales ESC : Annales. Économies, Sociétés, Civilisations

CETEDOC : Centre de Traitement Électronique des Documents

CNRS : Centre Nationale de la Recherche Scientifique

CRAL : Centre de Recherche et d'Application Linguistique

CRH : Centre de Recherches Historiques

EHESS : École des Hautes Études en Sciences Sociales

ENS : École Nationale Supérieure

EPHE : École Pratique des Hautes Études

ERA : Équipe de Recherche Autonome

IBM : International Business Machines Corporation

IRHT : Institut de Recherche et d'Histoire des Textes

LASLA : Laboratoire d'Analyse Statistique des Langues Anciennes

RCP : Recherche Coopérative sur Programme

Sommaire.

Remerciements	7
Liste des principales abréviations	11
Sommaire	12
Introduction générale	20
Construction de l'objet d'étude.....	20
Un terrain d'étude spécifique : l'histoire de l'histoire.....	22
Réflexions d'acteurs sur l'utilisation des ordinateurs en histoire.....	24
Observations sur les discours historiques chez les acteurs.....	30
Qu'est-ce qu'un médiéviste ? Éléments de définition.....	33
Instrumentation des recherches en histoire : des machines mécanographiques aux ordinateurs (États-Unis, France).....	35
Premières utilisations des ordinateurs dans le cadre d'enquêtes historiques en France.....	38
Trois problématiques centrales.....	42

PARTIE I

Chapitre 1 : Une étude assistée par ordinateur du <i>catasto</i> florentin de 1427 (1966-1978)	45
1. Organisations collectives et ordinateurs.....	53
1.1. Chiffrer et financer l'étude du <i>catasto</i>	54
1.1.1. Un premier programme de travail.....	54
1.1.2. Crise de confiance.....	60
1.2. Partage de compétences et choix des collaborateurs.....	62
1.2.1. Esprit d'équipe.....	62
1.2.2. Des partenaires évidents ?.....	67
1.2.3. Les ressources d'une forme d'enquête collective.....	69
1.3. Partage de matériel.....	72
1.3.1. Centres de calculs, perforatrices, lecteurs de microfilms, microfilms, programmes, cartes perforées.....	72
1.3.2. Échanges et voyages périlleux.....	76
2. Opérations de constitution de l'« édition » : collecte des données, code et codage.....	78
2.1. La collecte des données.....	80
2.1.1. Enjeux de la collecte des données.....	80
2.1.2. Quelques techniques mises en place.....	83
2.2. Le code et le codage.....	85
2.2.1. Choix des variables et interdisciplinarité.....	88
2.2.2. Construction de catégories et problèmes d'interprétations.....	93
3. Raisonnements, programmes et documents de sortie.....	98

3.1. Raison statistique : quelles innovations et comment les saisir ?.....	99
3.1.1. Puissance de calcul et multiplication des opérations statistiques.....	100
3.1.2. Conséquences de la multiplication des opérations statistiques.....	102
3.2. Une nouvelle pratique de la critique des sources.....	107
3.2.1. Trois méthodes pour l'évaluation des âges moyens au mariage.....	108
3.2.2. Une construction de la critique des sources en dialogue avec les sciences sociales.....	109
3.3. Utilisation des coefficients de corrélations et conséquences historiographiques.....	112
3.3.1. Étude de l'intensité des relations entre trois variables.....	112
3.3.2. Exploitation des résultats de la matrice de corrélations.....	115
3.4. Cartographie automatique et raisonnements historiques.....	116
3.4.1. La cartographie automatique « <i>n'a rien à voir avec la géographie</i> ».....	117
3.4.2. Apport de la cartographie automatique dans l'étude de la Toscane médiévale.....	120
4. Réutilisations de l' « édition » et des conclusions de l'étude.....	125
4.1. Réutilisation de l' « édition ».....	126
4.1.1. Réutilisations ponctuelles.....	126
4.1.2. Réutilisations programmatisées.....	129
4.2. Réception du livre du côté des historiens : un étalon pour l'histoire de l'Italie médiévale.....	130
4.2.1. Critiques élogieuses.....	131
4.2.2. Limites méthodologiques de l'étude.....	134
4.2.3. Réutilisations des conclusions.....	136
4.3. Citations et réutilisations en démographie.....	138
4.3.1. Une « <i>référence obligée de toute recherche sur les sociétés médiévales</i> ».....	138
4.3.2. Une réutilisation marquante des résultats démographiques....	139
Conclusion.....	141

Chapitre 2 : Lexicométrie et histoire médiévale : le programme ALINE de l'Équipe de Recherche Associée 713 (1971-1984).....151

1. Organisations collectives et ordinateurs.....	162
1.1. Cadre administratif du projet et ressources informatiques.....	163
1.1.1. Composition de l'Équipe de recherche sur l'Histoire politique du bas Moyen-Âge.....	163
1.1.2. Le centre de calcul de l'université Paris 1 – Panthéon Sorbonne.....	166
1.2. Informaticiens vacataires et nature du programme.....	169

1.2.1. Problèmes de recrutement et objectifs de travail.....	170
1.2.2. ALINE : un programme informatique original.....	172
1.3. Contraintes économiques et contraintes de temps.....	175
1.3.1. Rythme et coût de la perforation des cartes.....	176
1.3.2. Solutions techniques.....	177
2. Constitution des corpus et éditions des textes.....	179
2.1. Modalités de fabrication des corpus.....	180
2.1.1. Qu'est-ce qu'un texte politique ?.....	180
2.1.2. Logiques de constitution des corpus.....	183
2.2. Influences extra-disciplinaires.....	186
2.2.1. Observations sur les différences entre corpus d'historiens et corpus de linguistes.....	186
2.2.2. Fonction épistémologique des corpus.....	188
2.3. Opérations d'édition des textes sur cartes perforées IBM à 80 colonnes.	190
2.3.1. Du manuscrit à la cartes perforée.....	191
2.3.2. Quelques problèmes spécifiques aux textes étudiés à l'ERA 713.....	194
3. Raisonnements, programme et statistiques lexicales.....	198
3.1. La mesure du vocabulaire comme méthode pour l'histoire des idées politiques.....	199
3.1.1. Dépasser le « <i>contextualisme</i> » de Quentin Skinner.....	199
3.1.2. Une analyse des idées politiques de Sir John Fortescue.....	201
3.2. Comptages, tris et ordinateurs.....	204
3.2.1. Fonctionnalités du programme ALINE.....	204
3.2.2. Apports de l'automatisation de l'établissement des dictionnaires de fréquences.....	208
3.3. Automatisation de l'étude sémantique.....	211
3.3.1. Comment attribuer un sens à un mot ?.....	212
3.3.2. Analyse automatique des contextes.....	216
3.4. Approche classificatoire et analyse factorielle des correspondances....	219
3.4.1. Mise en œuvre des analyses factorielles des correspondances.	221
3.4.2. Exploitation des graphiques.....	225
4. Réutilisation et devenir de ALINE et de MEDITEXT.....	231
4.1. Circulation du programme ALINE.....	231
4.1.1. Utilisations par les médiévistes de l'ERA 713.....	233
4.1.2. Utilisation par une chercheuse « associée » à l'ERA 713.....	235
4.1.3. Utilisation dans le cadre d'activités de formation.....	236
4.1.4. Utilisation en dehors de l'ERA 713.....	237
4.2. Circulation et devenir de la cartothèque MEDITEXT.....	238
4.2.1. Un processus de transferts de données extrêmement lourd (1972- 1989)	239
4.2.2. Devenir de MEDITEXT après 1989.....	240

Conclusion	243
-------------------------	------------

PARTIE II

Chapitre 3 : Une nouvelle ambition collective pour l’histoire médiévale : les calculateurs électroniques (1970-1979)

1. Les calculateurs électroniques, nouveau vecteur de collaboration à l’échelle de la discipline (1970-1975).....	259
1.1. Réorganiser toute la discipline : la proposition de Genicot.....	260
1.1.1. Fonctions des ordinateurs dans l’élaboration de nouveaux instruments de travail.....	261
1.1.2. Une organisation collective nécessaire à l’échelle de la discipline.....	263
1.2. Mettre en commun des difficultés : l’objet initial du colloque de Rome.....	264
1.2.1. Un projet de rencontre entre deux collectifs.....	266
1.2.2. D’autres fonctions pour d’autres types de sources.....	268
1.3. Une première rencontre méta-collective : le colloque de Rome de 1975.....	269
1.3.1. Instruments de travail génériques.....	270
1.3.2. Instruments de travail spécifiques.....	273
2. Problématiques communes : les échanges du colloque de Rome.....	275
2.1. Archives en série, collections homogènes et documents répétitifs.....	276
2.1.1. Un modèle de critique des ensembles documentaires emprunté aux historiens modernistes.....	277
2.1.2. Application aux sources médiévales.....	279
2.1.3. Élaboration d’ « <i>un cadre systématique propice à une automatisation</i> ».....	282
2.2. Instruments de travail et problématiques historiques.....	285
2.2.1. Diplomatique.....	285
2.2.2. Onomastique, histoire sociale et prosopographie.....	289
2.2.3. Étude du vocabulaire juridique et institutionnel.....	293
2.3. Méthodes d’enregistrement des données en machine.....	295
2.3.1. Limites des enregistrements <i>full-text</i> et statistique.....	296
2.3.2. Une solution intermédiaire : l’indexation documentaire.....	299
3. Deux propositions d’organisation méta-collective à l’échelle européenne.....	302
3.1. Le questionnaire du groupe « Ordinateurs et Recherches Médiévales ».....	304
3.1.1. Finalité des instruments de travail et modalité d’enregistrement des données	304
3.1.2. Un « <i>modèle de description</i> » pour les projets mobilisant des ordinateurs.....	307
3.2. L’initiative de Lucie Fossier et du service informatique de l’IRHT.....	309

3.2.1. Une autre enquête et un autre questionnaire.....	310
3.2.2. Bilan de l'enquête de l'IRHT et changement de stratégie.....	313
Conclusion	315

Chapitre 4 : Objectifs et stratégies de publication du bulletin de liaison *Le Médiéviste et l'ordinateur* (1979-1989).....322

1. L'organisation du <i>Médiéviste et l'ordinateur</i>	328
1.1. Qu'est-ce que <i>Le Médiéviste et l'ordinateur</i> ?.....	329
1.1.1. Finalité et structure interne du bulletin.....	330
1.1.2. Matérialité et succès.....	331
1.2. Bien plus qu'une publication.....	333
1.2.1. Un annuaire de personnes et de lieux.....	334
1.2.2. Consultations techniques.....	336
1.2.3. Expertise.....	338
1.3. Qui écrit dans le <i>Médiéviste et l'ordinateur</i> ?.....	340
1.3.1. Trois scénarios de sélection des auteurs.....	340
1.3.2. Processus de révisions des articles.....	344
2. Coordonner la production textuelle d'une communauté discursive.....	346
2.1. Diversité des collectifs engagés et élaboration de thèmes communs.....	347
2.1.1. 70 collectifs et autant de cultures locales.....	347
2.1.2. Fonctions de l'organisation thématique.....	349
2.1.3. Évolutions des équipements : micro-informatique, lecteurs optiques, disquettes	352
2.2. Cartographie des sujets abordés dans le bulletin (1979-1989).....	355
2.2.1. Graphe de cooccurrences des termes les plus fréquents du corpus.....	356
2.2.2. Rubriques, thèmes, sujets.....	362
3. Favoriser la circulation de nouvelles pratiques.....	364
3.1. Un lectorat hétérogène aux capacités inconnues.....	364
3.1.1. Quel lectorat-cible pour le bulletin de liaison ?.....	365
3.1.2. Une stratégie de publication : le choix de l'ordre des articles...367	
3.2. Chemin d'apprentissage vers l'analyse factorielle des correspondances.....	368
3.2.1. Six articles aux approches variées.....	369
3.2.2. Six fonctions auprès des lecteurs.....	371
3.3. Publication de nouveaux types de textes et stratégies didactiques.....	373
3.3.1. Des documents de sortie hermétiques.....	374
3.3.2. Un exemple de stratégie didactique : le recours aux questions directes	376
3.3.3. Un autre exemple : l'annotation de documents de sortie.....	381
Conclusion	385

PARTIE III

Chapitre 5 : Genre et fonctions des « articles de fond » du <i>Médiéviste et l'ordinateur</i> (1979-1989).....	395
1. Une auteure, quatre contextes de publication.....	401
1.1. Étude du vocabulaire des biographies royales françaises (1982-1986)...	402
1.1.1. Un article dans les <i>Cahiers de Civilisation Médiévale</i>	402
1.1.2. Un article dans les <i>Annales. ESC</i>	406
1.2. Étude du cadastre d'Orvieto de 1292.....	408
1.2.1. Quelques points communs entre un article dans la <i>Revue du Nord</i> et un article dans le <i>Médiéviste et l'ordinateur</i>	408
1.2.2. Dimensions techniques <i>versus</i> résultats.....	409
1.3. Premières conclusions : esquisse d'une formation discursive.....	412
1.3.1. Genre et sous-genre d'articles en histoire quantitative.....	412
1.3.2. Définition et application du critère de répliquabilité	413
2. Parties méthodologiques	414
2.1. Présentation étape par étape, matérialité des opérations et expression des difficultés.....	417
2.1.1. Présentation étape par étape.....	417
2.1.2. Matérialité des opérations.....	419
2.1.3. Expression des difficultés et ressentis personnels.....	421
2.2. Variations dans un contexte de publication généraliste : les <i>Annales</i> <i>ESC</i>	423
2.2.1. Difficultés personnelles et difficultés générales.....	424
2.2.2. Considérations matérielles et instruments de calcul.....	426
2.2.3. Des marqueurs de l'enchaînement des opérations discrets.....	428
2.2.4. <i>Nous</i> ou <i>je</i> : positionnement des auteurs vis-à-vis du lectorat.....	429
2.3. Variations dans un contexte de publication spécialisé : <i>Histoire &</i> <i>Mesure</i>	431
2.3.1. Étude des séquences de composition des incunables (1987)...	431
2.3.2. Étude du vocabulaire des <i>exempla</i> (1986).....	436
2.4. Deuxièmes conclusions : quelques caractères des parties méthodologiques.....	441
2.4.1. Quelques remarques générales.....	441
2.4.2. Un genre propre aux parties méthodologiques ?.....	443
3. Les « articles de fond » du <i>Médiéviste et l'ordinateur</i>	448
3.1. Expériences pratiques des auteurs et contextes de publications.....	449
3.1.1. Finalités et stratégies rhétoriques.....	450
3.1.2. Description d'une procédure de calcul et répliquabilité.....	453
3.1.3. Une première fonction des articles du <i>Médiéviste et</i> <i>l'ordinateur</i>	456

3.2. Comparaison entre les parties méthodologiques et les « articles de fonds ».....	458
3.2.1. Présentation étape par étape des opérations.....	458
3.2.2. Mentions de la matérialité des opérations.....	459
3.2.3. Présence des auteurs dans les textes.....	460
3.2.4. Mentions des difficultés.....	462
3.3. Plan type des « articles de fond » du <i>Médiéviste et l'ordinateur</i>	464
3.3.1. Définition du plan-type.....	464
3.3.2. Variations autour du plan-type et diversité des sous-genres d'articles.....	465
Conclusion	469
Chapitre 6 : Conflits épistémologiques autour d'une utilisation des analyses factorielles des correspondances.....	477
1. Le problème ancien du développement du féodalisme à l'aune d'une procédure mathématique nouvelle.....	486
1.1. Le tableau de Banniard : un tableau pour opérer une synthèse.....	487
1.1.1. Contexte de publication du tableau.....	487
1.1.2. La légende comme indice des logiques de construction du tableau.....	489
1.2. Du tableau synoptique à la matrice d'incidence.....	493
1.2.1. Transformation des données dans le but de mettre en œuvre des analyses factorielles.....	494
1.2.2. Qu'est-ce qu'une analyse factorielle des correspondances ?...497	
1.2.3. Les graphiques publiés par Guerreau.....	504
1.3. Nouvelle visualisation synthétique des données.....	507
1.3.1. Nouvelle représentation, nouvel outil heuristique ?.....	507
1.3.2. Nouvelle synthèse ? Nouveaux résultats ?.....	510
2. Simplifier les opérations de synthèse historique.....	511
2.1. Informations « vagues », codages « <i>subjectifs</i> ».....	513
2.2. L'analyse factorielle de correspondances aplatit-elle la chronologie ?...517	
2.3. Niveaux de synthèse <i>versus</i> structure. L'histoire comme science sociale.....	519
2.4. « <i>Saint Ordinateur, priez pour nous</i> ».....	526
Conclusion.....	530
Conclusion générale.....	536
Organisations collectives de la recherche.....	536
Pratiques des sources.....	547
Conséquences historiographiques.....	557
Bibliographie.....	563

Entretiens	602
Table des matières.....	604
Table des figures.....	606

Introduction générale.

Dès le début des années 1950, l'irruption des calculateurs électroniques dans les sciences humaines et sociales a profondément affecté les *pratiques des textes* des chercheurs qui se sont saisis de ces nouveaux outils. Par cette expression nous entendons désigner à la fois les textes sources du travail et les textes produits au cours du travail. Notre thèse se propose d'examiner ce phénomène. Notre objectif est double : saisir certaines des conséquences de cette nouvelle instrumentation de la recherche, et contribuer ainsi plus largement à l'étude des effets de l'utilisation des calculateurs électroniques sur les cultures scientifiques au sein desquels le travail sur les écrits et par l'écrit joue un rôle clef.

Construction de l'objet d'étude.

Pour retracer comment s'est construit l'objet de cette thèse, il nous faut partir du livre de l'anthropologue britannique Jack Goody (1919-2015) intitulé *The Domestication of the Savage Mind*¹. Cet ouvrage, traduit en français sous le titre *La raison graphique*², accorde un rôle crucial au concept de *technologie intellectuelle*. Ce dernier avait été introduit quelques années plus tôt, sans que Goody en ait visiblement eu connaissance, par le sociologue américain Daniel Bell (1919-2011) dans un livre intitulé *The coming of post-industrial society: a venture in social forecasting*³. Bell mobilisait ce concept dans sa recherche pour désigner les dispositifs informatiques qui venaient se substituer à d'autres procédés fondés sur l'intuition. Par contraste, Goody étendait dans son livre la portée de ce concept, en l'appliquant à plusieurs autres dispositifs liés à l'écriture. Selon lui, les démarches intellectuelles des tenants de cultures qui utilisaient l'écriture avaient été transformés de façon radicale par l'emploi de technologies, spécifiques à l'écrit, comme la liste, le tableau et le diagramme.

¹ Jack Goody, *The domestication of the savage mind* (Cambridge, Royaume-Uni, 1977).

² Jack Goody, *La raison graphique. La domestication de la pensée sauvage* (Paris: Les Éditions de Minuit, 1979).

³ Daniel Bell, *The coming of post-industrial society: a venture in social forecasting* (New York, États-Unis d'Amérique: Basic books, 1973). Une traduction en français est publiée la même année : Daniel Bell, *Vers la société post-industrielle* (Paris, France: Éditions Robert Laffont, 1973).

C'est en continuité avec cette seconde approche que nous avons élaboré la problématique de cette thèse, dans la mesure où nous y explorons spécifiquement l'idée que, dans une culture scientifique accordant une place centrale à l'écriture, les raisonnements se déploient au cours de processus matériels qui s'appuient sur la production, la manipulation et l'échange de différents types de textes⁴. Partant de cette idée, il apparaissait évident, *a priori*, que le recours à des ordinateurs avaient eu des conséquences décisives sur la forme que revêtaient les *technologies intellectuelles* sur lesquelles Goody nous invitait à nous pencher, comme sur les fonctions qu'elles assumaient dans les raisonnements.

Cette hypothèse, qui a inspiré notre travail de thèse de part en part, nous a conduit à la question suivante : comment l'emploi d'ordinateurs avait-il transformé, voire fait muter les pratiques des chercheurs qui s'étaient emparés de ces nouveaux instruments à partir des années 1950 ?

Ce travail avait également une ambition plus pratique : permettre, par la connaissance poussée de leurs logiques de constitution, une meilleure critique des instruments de travail numériques produits par les chercheurs en sciences humaines. Bases de données, éditions électroniques, moteurs de recherches, archives informatisées sont aujourd'hui des ressources cruciales pour toutes les recherches conduites dans le monde académique. L'étude des contextes de production de certains de ces instruments de travail devait contribuer au déploiement de ce regard critique, à l'instar du programme de travail lancé par des historiens des sciences pour ce qui est des éditions de textes savants en mathématiques⁵ ou pour les projets encyclopédiques⁶. Pour ce faire, il nous semblait particulièrement pertinent de nous tourner non pas vers les emplois les plus récents des ordinateurs, mais au contraire vers les premières applications des outils

⁴ Cette approche n'est cependant pas une innovation de cette thèse. Elle a notamment été fortement développée par le groupe de recherche « Histoire des sciences, histoire du texte » créé en 1996 à Paris au sein du laboratoire RHESEIS par l'une des deux directrices de ce travail de thèse : Karine Chemla.

⁵ Voir, par exemple, Karine Chemla, « Commentaires, éditions et autres textes seconds : quel enjeu pour l'histoire des mathématiques ? Réflexions inspirées par la note de Reviel Netz », *Revue d'histoire des mathématiques* 5, n° 2 (1999): 127-148 ou, plus récemment, Emmylou Haffner, « L'édition des oeuvres mathématiques au XIX^e siècle en Allemagne. l'exemple des Gesammelte Werke und wissenschaftlicher Nachlass de Bernhard Riemann », *Philosophia Scientiae* 22, n° 2 (2018): 115-135.

⁶ Mathias Grote, « Total Knowledge? Encyclopedic Handbooks in the Twentieth-Century Chemical and Life Sciences », *BJHS Themes* 5 (eds 2020): 187-203.

informatiques, attestées dans les sciences humaines et sociales à partir de la fin des années 1940⁷.

Quels avaient pu être les usages envisagés par les créateurs des premières bases de données ou des éditions de sources sur ordinateur ? Quelles étaient les ambitions scientifiques attachées à ces projets ? Quelles étaient les limites de ces nouveaux types d'instruments de travail ? Et dans quelle mesure étaient-elles dictées par le contexte matériel et technique dans lequel ils avaient été développés ? Enfin, qu'est-ce que les conditions de production de ces instruments de travail pouvaient nous apprendre sur leurs limites ou sur leurs apports dans les recherches en sciences humaines et sociales ?

Un terrain d'étude spécifique : l'histoire de l'histoire en France.

Pour élaborer une approche réaliste face à des problématiques d'une telle ampleur, il est nécessaire de préciser le terrain de notre étude. Nous avons choisi de nous concentrer plus particulièrement sur le déploiement des ordinateurs et des méthodes numériques par les historiens en France. Ce choix repose sur plusieurs considérations.

La première tient à notre parcours académique et à notre spécialisation en histoire, qui fait de ce domaine une *terra* moins *incognita* que d'autres disciplines des sciences humaines et sociales. Le choix d'étudier des historiens français s'explique essentiellement par notre implantation à Paris au cours de ce parcours académique. Précisons toutefois que notre travail de thèse nous a conduit à étendre ce champ de recherche vers l'Italie, l'Allemagne ou la Belgique, comme nous le verrons au fil des six chapitres de cette thèse.

Notre deuxième motivation tient à la place que les historiens occupent dans le monde académique, puisqu'ils sont des spécialistes reconnus de l'étude des cultures écrites. À travers une étude précise de quelques-unes des manières par lesquelles ils

⁷ L'un des premiers projets de ce type, qui est également, dans l'histoire des humanités numériques, le plus largement étudié, est le projet d'Index Thomisticus de Roberto Busa. Voir par exemple : Steven Edward Jones, *Roberto Busa, S. J., and the emergence of humanities computing: the priest and the punched cards* (New York, États-Unis d'Amérique, Royaume-Uni et d'Irlande du Nord, 2016). Notons cependant que l'écriture de l'histoire des débuts des humanités numériques a connu de grands développements ces trois dernières années, et que certains historiens contestent le caractère pionnier du travail de Busa. Voir Mario Wimmer, « Josephine Miles (1911–1985): Doing Digital Humanism with and without Machines », *History of Humanities* 4, n° 2 (1 septembre 2019): 329-34.

s'étaient saisis des ordinateurs, nous pouvons ainsi nous pencher sur deux ordres de questions. En premier lieu, nous sommes en position d'examiner les mutations dans la production, la manipulation et la circulation des textes qu'ils produisaient au cours de leurs recherches (documents de travail, instruments de travail, visualisations, publications). En un second lieu, il est possible d'analyser les transformations qu'impliquait la mise en œuvre de ces nouveaux moyens techniques dans le rapport de ces acteurs aux textes qu'ils étudiaient, ou en d'autres termes, à leurs sources.

Plusieurs questions peuvent ainsi être soulevées. Comment les historiens ont-ils intégré des ordinateurs dans leurs pratiques de recherche ? Quelles répercussions ce phénomène a-t-il eu sur leur manière de constituer des corpus, de lire et de critiquer leurs sources, et d'organiser collectivement leurs recherches ? Quels nouveaux types de textes et d'instruments de travail ont-ils produits au contact des ordinateurs ? Et quels nouveaux types de rapports à leurs sources et à leurs écrits ont-ils développé dans ce contexte ?

Ce terrain d'étude présente toutefois quelques difficultés spécifiques pour aborder les problématiques générales que nous avons en vue. La principale d'entre elles a trait à l'absence d'autres travaux sur l'instrumentation de la recherche en histoire au XX^e siècle, en dépit des nombreuses innovations techniques dont les historiens ont pu profiter pour faire évoluer leurs méthodes au cours de cette période (photographie aérienne, calculatrices mécaniques, mécanographie, etc.⁸). De ce fait, il nous est très compliqué de comparer les pratiques des historiens utilisant des ordinateurs avec celles reposant sur d'autres instruments de calcul. Il est également difficile d'identifier des contrastes avec les pratiques de certains de leurs collègues engagés dans l'utilisation d'autres technologies nouvelles à la même période.

La seconde difficulté dérive de la précédente, puisqu'elle tient à l'absence de travaux d'observateurs sur l'utilisation des ordinateurs en histoire dans la deuxième moitié du XX^e siècle⁹. À ce titre, notre thèse comble indéniablement une lacune. Cependant, cet état de fait rend difficile le positionnement de notre travail vis-à-vis d'un

⁸ Nous reviendrons au fil de cette introduction et de cette thèse sur les deux derniers items de cette énumération. Pour le premier, voir par exemple : Raymond Chevallier, « Photographie aérienne et histoire économique et sociale de l'Antiquité : un livre, une exposition », *Annales* 13, n° 2 (1958): 322-26.

⁹ Notons tout de même ici qu'une thèse sur ce sujet a été entamée à la fin des années 1990 par Éric Castex. Ce dernier n'a cependant pas mené ce travail à son terme.

ensemble de réflexions formulées, depuis le début des années 1960, par des acteurs de cette histoire.

Réflexions d'acteurs sur l'utilisation des ordinateurs en histoire.

Peu après les premières utilisations des ordinateurs dans leur discipline¹⁰, les historiens ont produit des discours réflexifs sur ces nouvelles pratiques. Ils ont soulevé de nombreuses problématiques centrales, qui sont restées au cœur des réflexions des acteurs jusqu'à aujourd'hui, et ce en dépit de fulgurantes et profondes évolutions technologiques (stockage numérique, microinformatique, internet, *big data*, etc.).

Comment l'utilisation des ordinateurs vient-elle bouleverser l'accès aux archives et, plus largement, à la documentation mobilisée dans les enquêtes historiques ?¹¹ Quelles mutations dans les processus de connaissance peut-on identifier du fait recourir à des calculateurs électroniques ? Et dans quelle mesure ces transformations diffèrent-elles des changements liés à l'introduction des méthodes statistiques dans la discipline ?¹² Enfin, comment former les historiens et les étudiants à l'utilisation des ordinateurs, ou à la lecture des publications issues de recherche menées à l'aide de ces instruments ?¹³ Et

¹⁰ Nous présentons, dans la section « Instrumentation des recherches en histoire : des machines mécanographiques aux ordinateurs (États-Unis, France). » de cette introduction, ces premières utilisations.

¹¹ Voir par exemple Jerome M. Clubb et Howard Allen, « Computers and Historical Studies », *The Journal of American History* 54, n° 3 (1967): 599-607 ; André Zysberg, « Impact de l'informatique sur la recherche historique », *Matériaux pour l'histoire de notre temps* 10, n° 1 (1987): 35-37 ; Philippe Rygiel, « Les sources de l'historien à l'heure d'Internet », *Hypothèses* 7, n° 1 (2004): 341-54 ; Sonia Combe, « L'historien face à l'ordre informatique. Classification et Histoire », *Matériaux pour l'histoire de notre temps* 82, n° 2 (2006): 1-3 ; Jean-Philippe Genet et al., *Les historiens et l'informatique: un métier à réinventer* (Rome, Italie: École française de Rome, 2011) ou Philippe Rygiel, *Historien à l'âge numérique*, Papiers (Villeurbanne: Presses de l'ENSSIB, 2019).

¹² On pourra se reporter à Sheldon Hackney, « Power to the computers: a revolution in history? », in *Proceedings of the May 5-7, 1970, spring joint computer conference*, AFIPS '70 (Spring) (New York, NY, USA: Association for Computing Machinery, 1970), 275-79 ; « L'utilisation de l'informatique par les historiens d'aujourd'hui : bilan et perspectives », *Cahiers de la Méditerranée* 53, n° 1 (1996): 211-17 ; Frédéric Kaplan, Mélanie Fournier, et Marc-Antoine Nuessli, « L'historien et l'algorithme », *Le Temps des Humanités Digitales*, 2014, 49.

¹³ Quelques-unes de ces réflexions peuvent être trouvées dans Clubb et Allen, « Computers and Historical Studies » ; Edmund A. Bowles, « Towards a Computer Curriculum for the Humanities », *Computers and the Humanities* 6, n° 1 (1971): 35-38 ; Robert P. Swierenga, « Computers and American History: The Impact of the "New" Generation », *The Journal of American History* 60, n° 4 (1974): 1045-70 ; Jean-Philippe Genet, « La formation des historiens à l'informatique en France : Espoir ou désespoir ? », *Le médiéviste et l'ordinateur* 31, n° 1 (1995): 32-36 ; Jean-François Soulet et al., *L'informatique dans l'enseignement de l'histoire et la formation des historiens: actes du V^e colloque national de l'Association française pour l'histoire et l'informatique, 3 et 4 novembre 1998, université de Toulouse-le-Mirail* (Paris,

quels impacts ont pu avoir ces nouveaux instruments sur la diffusion des connaissances historiques vers des publics non-professionnels ?¹⁴

Ces questions ont été abordées dans de très nombreuses publications entre les années 1960 et aujourd'hui, et il serait impossible de produire une synthèse complète et fidèle de ces innombrables écrits. Toutefois, deux questions, centrales pour notre sujet, s'imposent : quels ont été les discours à caractère historique produits par ces auteurs quant à l'emploi des ordinateurs en histoire ? Et quelles étaient les fonctions de ces discours dans leurs réflexions ?

Nous pouvons formuler un premier constat : dans une majorité de ces textes, l'histoire des utilisations des ordinateurs par les historiens est tout bonnement absente. Cela tient au fait que leurs auteurs se concentrent sur l'étude, la présentation ou la critique de projets mobilisant des ordinateurs qui leurs sont contemporains, voire, sur les implications que pourraient avoir ces entreprises sur le futur de la discipline. Dans d'autres textes, qui nous intéresseront ici plus particulièrement, les perspectives historiques tiennent une place minime, et sont réservées à certaines parties de textes comme les introductions ou les conclusions.

Pour en saisir plus pleinement la teneur, il faut préciser que ces discours à caractère historique s'appuient, de façon constante, sur deux types de matériaux différents. Dans certains cas, les acteurs mobilisent leurs souvenirs ou leurs expériences dans le domaine¹⁵. C'est par exemple ce que propose Jean-Philippe Genet (1944-), utilisateur pionnier des ordinateurs en histoire médiévale - sur lequel nous reviendrons largement dans le chapitre 2 de cette thèse. Dans l'avant-propos d'un livre publié en 2011, intitulé *Les historiens et l'informatique : un métier à réinventer*, il présente la finalité de l'ouvrage en ces termes :

France, 2001); Roy Rosenzweig, *Clio wired: the future of the past in the digital age* (New York: Columbia University Press, 2011); Franziska Heimburger et Émilien Ruiz, « Faire de l'histoire à l'ère numérique : retours d'expériences », *Revue d'histoire moderne et contemporaine* n° 58-4bis, n° 5 (2011): 70-89.

¹⁴ L'apparition de cette problématique est beaucoup plus récente que les autres, puisqu'elle n'est attestée, à notre connaissance, que depuis une dizaine d'années. Voir Serge Noiret, « La digital history : histoire et mémoire à la portée de tous », 2011 ; Roy Rosenzweig, *Clio wired: the future of the past in the digital age* (New York: Columbia University Press, 2011).

¹⁵ Voir, par exemple, l'introduction de Jean-Philippe Genet et al., *Les historiens et l'informatique: un métier à réinventer* (Rome, Italie: École française de Rome, 2011), Éric Brian, « L'horizon nouveau de l'historiographie expérimentale », *Revue d'histoire moderne et contemporaine* n° 58-4bis, n° 5 (2011): 41-56, ou Philippe Rygiel, « Introduction », in *Historien à l'âge numérique*, Papiers (Villeurbanne: Presses de l'ENSSIB, 2019).

« faire le bilan prospectif des apports des méthodes dont l'informatique a ouvert l'accès aux historiens et, d'une façon plus générale, évaluer l'ampleur des transformations que le recours à l'informatique a introduites et continue à introduire dans la pratique du métier d'historien. »¹⁶

Dans cette optique, il dessine une périodisation des usages de l'informatique par les historiens en deux phases. Un moment au cours duquel il diagnostique un processus de création d'« *un champ spécifique* » dédié aux échanges entre utilisateurs de l'informatique entre 1970 et 1986¹⁷. Un second qu'il présente comme une phase de dislocation des modalités d'organisation, d'échanges et de travail qui constituaient ce champ, du fait de l'explosion des usages des ordinateurs personnels puis de l'amplification du recours à Internet. Au fil de cette présentation, Genet rappelle certaines des activités auxquelles il a pris part pour constituer ce « *champ spécifique* » dans les années 1970 et 1980 : publications dans des périodiques spécialisés, création d'institutions ou de revues académiques dédiées, organisation ou participation à des conférences, etc.

Les éléments historiques que développent ici Genet sont mobilisées à titre d'exemples, dans le but d'insister sur la nécessité d'anticiper les difficultés à venir pour les historiens, dans des contextes où les mutations techniques apparaissent à un rythme exponentiel. Les perspectives historiques ont dans ce cas pour fonction de contribuer au diagnostic des difficultés à venir, selon une utilisation de l'histoire déjà maintes fois commentée : se servir du passé pour trouver des solutions face à des situations semblables dans le présent et l'avenir.

Plusieurs articles publiés dans les années 2000 et 2010, le plus souvent écrits par des historiens n'ayant pas pris part à de tels travaux dans les années 1970 et 1980, proposent quant à eux des développements historiques fondés sur certaines de ces réflexions d'acteurs. Ils s'appuient plus particulièrement sur des textes parus dans les années 1970¹⁸, période au cours de laquelle nombre d'acteurs publient des panoramas

¹⁶ Genet et al., *Les historiens et l'informatique*, VIII.

¹⁷ L'étude de ce « *champ* » est l'objet des analyses que nous développons dans les chapitres 3 et 4 de notre thèse.

¹⁸ C'est notamment ce qu'on observe dans deux articles publiés en 2011 dans un numéro thématique de la *Revue d'Histoire Moderne et Contemporaine* intitulé « Le métier d'historien à l'ère numérique : nouveaux outils, nouvelle épistémologie ? » : Franziska Heimburger et Émilien Ruiz, « Faire de l'histoire à l'ère

visant à recenser des projets déjà accomplis et à définir des orientations futures pour ce genre de travaux¹⁹.

C'est par exemple le cas de l'article publié en 2011 par Franziska Heimbürger et Émilien Ruiz, dans lequel les auteurs insistent sur la responsabilité qu'ont les historiens à former et à se former aux outils informatiques²⁰. En introduction, les deux auteurs proposent un retour historique de trois paragraphes sur l'utilisation des calculateurs électroniques par les historiens. Ils s'y appuient sur deux études de référence en histoire de l'informatique²¹ ainsi que sur plusieurs textes publiés par les acteurs depuis 1967 : deux panoramas du type de ceux définis plus haut²², deux textes de Genet publiés dans le premier numéro de la revue *Histoire & Mesure* qu'il avait créée en 1986²³, et une célèbre déclaration du moderniste Emmanuel Le Roy Ladurie (1929-) datée de 1967 : « *l'historien de demain sera programmeur ou ne sera plus* »²⁴.

Sur cette base, Heimbürger et Ruiz avancent l'idée que la situation dans laquelle ils se trouvent en 2011 est tout à fait différente de celle de la première révolution informatique qui s'est produite entre la fin des années 1960 et le début des années 1990. Cette différence tient en particulier à la possibilité d'accéder aux ressources numériques *via* Internet et à l'avènement du Web 2.0, qui a permis l'élaboration d' « *outils extrêmement faciles à utiliser* » et « *ne nécessitant pas (ou peu) de compétences informatiques* »²⁵. La fonction du discours historique élaboré dans ce contexte vise ainsi à insister sur le « *danger* » que représentent ces nouveautés et, partant, à mettre l'accent sur la nécessité de repenser les formations proposées aux étudiants en histoire.

numérique : retours d'expériences », *Revue d'histoire moderne contemporaine* n° 58-4bis, n° 5 (2011): 70-89; Nicolas Delalande et Julien Vincent, « Portrait de l'historien-ne en cyborg », *Revue d'histoire moderne et contemporaine* n° 58-4bis, n° 5 (2011): 5-29.

¹⁹ Voir par exemple Clubb et Allen, « Computers and Historical Studies » ; Robert P. Swierenga, « Clio and Computers: A Survey of Computerized Research in History », *Computers and the Humanities* 5, n° 1 (septembre 1970): 1-21 ; Joel H. Silbey, « Clio and Computers: Moving into Phase II, 1970-1972 », *Computers and the Humanities* 7, n° 2 (1972): 67-79 ; Swierenga, « Computers and American History: The Impact of the "New" Generation ».

²⁰ Heimbürger et Ruiz, « Faire de l'histoire à l'ère numérique ».

²¹ Paul E. Ceruzzi, *A history of modern computing* (Cambridge, Mass., Etats-Unis d'Amérique: MIT Press, 1998); Pierre Mounier-Kuhn, *L'informatique en France de la seconde guerre mondiale au Plan Calcul: l'émergence d'une science* (Paris, France: Presses de l'université Paris-Sorbonne, 2010).

²² Clubb et Allen, « Computers and Historical Studies » ; Swierenga, « Computers and American History ».

²³ « Éditorial », *Histoire & Mesure* 1, n° 1 (1986): 5-6; Jean-Philippe Genet, « Histoire, Informatique, Mesure », *Histoire & Mesure* 1, n° 1 (1986): 7-18.

²⁴ Cette déclaration nous est parvenue du fait de sa citation dans un article bien postérieur : Theodore K. Rabb, « The Development of Quantification in Historical Research », *The Journal of Interdisciplinary History* 13, n° 4 (1983): 591-601.

²⁵ Heimbürger et Ruiz, « Faire de l'histoire à l'ère numérique ».

On relève une autre modalité selon laquelle les historiens intègrent, dans leurs réflexions sur les utilisations des ordinateurs, des discours sur le passé récent. Dans un article de 2014, un groupe de chercheurs de l'École Polytechnique Fédérale de Lausanne proposait une démonstration qui visait à établir qu'est à l'œuvre, dans les sciences historiques

*« une évolution inédite [...] et que cette transformation est différente de celle qui a caractérisé, il y a quelques décennies l'arrivée de la « cliométrie » et des méthodes quantitatives. »*²⁶

L'hypothèse posée nous semble tout à fait pertinente. Elle rejoint d'ailleurs l'une des difficultés centrales que nous avons rencontrée au cours de nos recherches : comment distinguer, dans les évolutions des pratiques des historiens, ce qui tient de l'apparition des moyens électroniques de calculs de ce qui tient d'une forme de mathématisation de la discipline, bien antérieure à l'invention des ordinateurs ?

Dans la première partie de cet article, Kaplan, Fournier et Nuessli proposent un bref retour historique sur cette question. Ce passage vise à montrer que l'une des évolutions majeures liées aux utilisations des ordinateurs ne tient pas à la mise en œuvre de méthodes quantitatives informatisées à partir de la fin des années 1950, mais, par contraste, au fait que se développent depuis le début des années 2000 de nouveaux usages tournés vers l'élaboration d'analyses spatiales et/ou géographiques. Dans ce cas précis, les perspectives historiques servent à mettre en valeur la pertinence de cette différence que les auteurs établissent entre les deux périodes.

Pour appuyer cette idée, les trois auteurs se basent exclusivement sur un article publié en 2004²⁷. Largement cité et repris depuis une vingtaine d'années dans les travaux touchant aux questions qui nous intéressent²⁸, ce texte est l'œuvre de l'historien américain William G. Thomas III (1964-), spécialiste de l'histoire de la guerre de Sécession ainsi

²⁶ Kaplan, Fournier, et Nuessli, « L'historien et l'algorithme », 1.

²⁷ William G. Thomas, « Computing and the Historical Imagination », in *A Companion to Digital Humanities* (John Wiley & Sons, Ltd, 2004), 56-68.

²⁸ Nous avons identifié 86 citations de cet article à l'aide de Google Scholar. Nous avons cependant pu constater que ce dénombrement n'était pas complet, puisque de nombreuses citations manquaient, plus particulièrement celles faites dans les articles en français.

que de l'esclavage et utilisateur d'outils informatiques depuis le début des années 1990²⁹. L'article en question paraît dans le premier *Companion to Digital Humanities*, ouvrage majeur pour les observateurs de l'histoire des *digital humanities*, puisqu'il consacre l'utilisation de ce terme pour désigner ce champ de recherche. Dans la première partie du livre, de nombreux auteurs du *Companion* cherchent notamment à établir les spécificités des *digital humanities* par rapport à un autre champ de recherche, défini par les acteurs des générations précédentes : les *humanities computing*. L'ensemble des contributeurs à l'ouvrage se fondent sur leurs expériences d'utilisateurs des ordinateurs dans leurs domaines de recherche respectifs pour proposer des remarques historiques sur la rupture entre *humanities computing* et *digital humanities*.

La contribution de Thomas s'inscrit dans ce contexte, et assume ainsi une fonction bien particulière : construire une histoire commune à même de souligner la spécificité d'un nouveau champ de recherche³⁰. De ce fait, le discours historique qu'il élabore est marqué par certaines limites. Tout d'abord, Thomas propose une approche exclusivement fondée sur son propre domaine d'étude : l'histoire de l'esclavage au XIX^e siècle aux États-Unis. De la même manière que Genet, il a aussi exclusivement recours à des exemples tirés de sa propre expérience professionnelle, voire à ses propres travaux ou ceux de ses plus proches collaborateurs.

Ensuite, l'histoire qu'il présente est uniquement concentrée sur l'utilisation des ordinateurs par des historiens américains, et passe sous silence les innombrables projets ayant mobilisés des ordinateurs en Italie, en France, en Belgique, au Royaume-Uni, au Pays-Bas, en Hongrie ou en Russie depuis la fin des années 1950³¹.

Enfin, l'article de Thomas néglige de mentionner certaines utilisations des ordinateurs en histoire. C'est plus particulièrement le cas des méthodes que l'on peut qualifier de *text-processing*, qu'il délaisse au profit d'une approche entièrement tournée

²⁹ Voir par exemple la page de présentation de l'un des projets auquel il collabore : <https://valley.lib.virginia.edu/VoS/usingvalley/valleystory.html>

³⁰ Nous ne reviendrons pas ici sur les différences entre *humanities computing* et *digital humanities*, sujet pour lequel un article complet serait, *a minima*, nécessaire. On pourra se reporter à l'introduction de la première édition du *Companion to digital humanities* publié en 2004 : Susan Schreibman, Ray Siemens, et John Unsworth, « The digital humanities and humanities computing: An introduction », *A companion to digital humanities*, 2004, 288-90.

³¹ Silbey, « Clio and Computers » ; Swierenga, « Computers and American History ».

vers les méthodes de *data-processing*, ne considérant ainsi l'utilisation des ordinateurs en histoire que dans la continuité des problématiques de l'histoire quantitative³².

Le fait que Kaplan, Fournier et Nuessli renvoient à ce seul article illustre selon nous un problème général. Les acteurs qui cherchaient, depuis le début des années 2000, à s'appuyer sur des travaux historiques pour mener leurs analyses se trouvaient confrontés aux limites de telles publications. Cet état de faits que nous avons identifié nous conduit à formuler quelques observations sur les discours à caractère historique proposés par les acteurs, qui nous permettront de préciser le positionnement et l'intérêt de notre travail dans cette vaste littérature.

Observations sur les discours historiques proposés par les acteurs.

L'analyse des énoncés par lesquels les acteurs proposent un aperçu historique des emplois des ordinateurs permet de dégager trois éléments caractéristiques, et ce quelles que soit les périodes au cours desquelles ils ont été produits. Premièrement, on remarque que lorsque les auteurs ont eux-mêmes été acteurs de cette histoire, ils fondent leur réflexion sur leur propre expérience, leur mémoire et leurs souvenirs. Ces témoignages sont par ailleurs utilisés comme des sources par d'autres chercheurs, qui n'ont pas eux-mêmes été acteurs de cette histoire.

Deuxièmement, en raison des matériaux sur lesquels ils s'appuient (expériences personnelles, souvenirs), ces témoignages proposent des histoires concentrées sur certains domaines de recherche (histoire médiévale, histoire de l'esclavage, histoire de l'immigration, etc.). Ils se focalisent par ailleurs sur certains milieux académiques, et en particulier sur le contexte américain. Ils se concentrent également sur certains usages des ordinateurs, qui supposent que l'on attribue à ces instruments certaines fonctions, et plus particulièrement celles d'une machine statistique.

Troisièmement, il semble que ces discours à caractère historique s'inscrivent dans un effort de périodisation qui a pour objet, dans la grande majorité des cas, d'éclairer les

³² Ce problème méthodologique, qui traverse toutes les études sur l'histoire des humanités numériques, a notamment été soulevée très récemment par deux chercheurs du Pratt Institute à New York : Chris Alen Sula et Heather V. Hill, « The early history of digital humanities: An analysis of *Computers and the Humanities* (1966–2004) and *Literary and Linguistic Computing* (1986–2004) », *Digital Scholarship in the Humanities* 34, n°1 (2019): 190-206. Nous y reviendrons dans le chapitre 3 de cette thèse.

raisons qui fondent les auteurs à affirmer qu'ils se trouvent à un tournant décisif de cette histoire. Le recours à des éléments tenants à l'histoire de ces pratiques n'a ainsi jamais pour finalité de faire l'histoire de l'emploi des ordinateurs par les historiens. Ce troisième trait éclaire deux principales fonctions assumées par les discours historiques dans les réflexions des acteurs.

D'une part, ces énoncés assurent une fonction d'exemple, et visent à appuyer une argumentation destinée à *diagnostiquer* les enjeux, les défis, voire les dangers et les menaces contemporains des pratiques des auteurs. Cette approche présentiste a même parfois une dimension prophétique, et les énoncés historiques s'intègrent alors à des réflexions sur le futur de la discipline³³.

D'autre part, le recours aux discours historiques peut avoir pour fonction de légitimer et de promouvoir certaines utilisations des ordinateurs par les historiens. À ce titre, l'article de Thomas présenté plus haut est particulièrement représentatif, puisque son auteur vise à contribuer à la constitution d'une histoire commune qui déterminerait l'identité des *digital humanities*. La fonction qu'assurent les parties historiques dans son travail est donc à la fois de légitimer les pratiques qui se développent au tournant des années 2000 en histoire et de promouvoir l'utilisation d'un nouveau terme pour désigner plus largement un nouveau champ de recherche qui apparaît dans les humanités aux États-Unis à cette période.

Dans notre travail, nous proposons de reprendre la question de l'impact qu'a pu avoir l'ordinateur sur la discipline historique avec d'autres méthodes, en nous appuyant sur d'autres sources et en poursuivant d'autres objectifs. En effet, comme nous l'avons affirmé au début de cette introduction, nous considérons l'histoire de l'histoire comme un terrain approprié pour poser les bases d'une histoire des pratiques des textes dans les cultures scientifiques en sciences humaines et sociales. En effet, si des travaux sur l'histoire de l'automatisation d'autres disciplines existent (linguistique, archéologie)³⁴, il

³³ Dans ces textes, on retrouve d'ailleurs de façon presque systématique l'utilisation du futur ou du subjonctif dans les conclusions, marqueur qui caractérise selon nous la dimension présentiste de ces approches du sujet. Voir par exemple Swierenga, « Computers and American History »; William G. Thomas, « Computing and the Historical Imagination », in *A Companion to Digital Humanities* (John Wiley & Sons, Ltd, 2004), 56-68 ; Genet et al., *Les historiens et l'informatique* ; Kaplan, Fournier, et Nuessli, « L'historien et l'algorithme »; Rygiel, « Introduction ».

³⁴ Jacqueline Léon, *Histoire de l'automatisation des sciences du langage*, Langages (Lyon: ENS, 2015); Sébastien Plutniak, « L'opération archéologique. Sociologie historique d'une discipline aux prises avec l'automatique et les mathématiques. France, Espagne, Italie, 2^e moitié du XX^e siècle » (Thèse de doctorat, Paris, EHESS, 2017).

ne donne pas aux pratiques des textes telles que nous les avons définies le rôle central que nous leur accorderons.

Pour inscrire notre étude dans cette direction, nous proposons de tenir compte non pas seulement des témoignages d'acteurs, mais également d'autres sources. Les archives des équipes de recherches engagés dans des projets mettant en œuvre des calculateurs électroniques, aussi bien que les périodiques dans lesquels les travaux de cette nature étaient discutés, nous permettront de nous appuyer sur des documents de travail élaborés par les historiens pour comprendre comment l'utilisation des ordinateurs a affecté leurs pratiques de recherches. Cette approche nous paraît d'autant plus pertinente qu'elle permettra de pallier à certaines limites des réflexions historiques des acteurs.

Étudier les pratiques des textes des acteurs requiert toutefois de se restreindre à un terrain encore plus circonscrit que celui de l'histoire de l'histoire. À la lumière des réflexions des acteurs que nous avons examinées, nous avons ainsi fait le choix de nous concentrer sur le cas particulier des médiévistes ayant déployé ce type de méthodes en France entre 1960 et 1990. Le choix de cet objet d'étude tenait à deux considérations principales.

Tout d'abord, ce sujet devait permettre de décentrer la perspective par rapport à l'histoire de l'utilisation des ordinateurs par les historiens, qui s'avère le plus souvent centrée sur les historiens modernistes et contemporanéistes aux États-Unis³⁵. Cette option méthodologique nous a par ailleurs permis de découvrir que la nature des sources étudiées par les médiévistes les avait confrontés à des problèmes techniques et épistémologiques différents de ceux des modernistes et des contemporanéistes, ce qui contribuait un peu plus à décentrer le regard de l'historiographie traditionnelle.

Une seconde considération rendait le sujet pertinent : les médiévistes avaient été, en Europe, parmi les premiers à s'être organisés de manière collective afin de partager, promouvoir et diffuser les méthodes de recherche élaborées au contact des ordinateurs. En France, cette organisation avait notamment entraîné la création d'un périodique d'un nouveau genre en 1979 : le bulletin de liaison *Le Médiéviste et l'ordinateur*. Cette

³⁵ Cette remarque rejoint celle, plus générale, que le philologue Aurélien Berra a avancée à propos de l'histoire des humanités numériques : Aurélien Berra, « Pour une histoire des humanités numériques », *Critique*, Des chiffres et des lettres : les humanités numériques, 819-820 (août 2015): 613-26.

documentation, la première du genre à notre connaissance en Europe³⁶, constituait une source formidable permettant non seulement d'identifier de nombreux collectifs engagés dans ce type de projets depuis la fin des années 1960 jusqu'au début des années 1990, mais également d'observer quelles stratégies les médiévistes avaient mises en œuvre pour faire circuler les savoirs et les savoir-faire associés aux usages des ordinateurs.

Une fois ce sujet défini, plusieurs questions relatives à la définition des termes doivent toutefois être soulevées, et notamment : qu'est-ce qu'un médiéviste ? Et qu'est-ce qu'un ordinateur ?

Qu'est-ce qu'un médiéviste ? Éléments de définition.

La forte cohésion de la catégorie professionnelle « médiéviste » a souvent été soulignée par les observateurs. Dans un article de 1995, l'historiographe français Olivier Dumoulin l'avait par exemple qualifié de « *tribu* »³⁷. Dans notre travail, nous avons toutefois pu observer que cette « *tribu* » est en réalité composée de spécialistes aux objets d'études très différents, employés dans des institutions très diverses (faculté de droit, faculté de lettres, faculté d'histoire, écoles nationales spécialisées, centres d'archives, etc.). Parmi les médiévistes, on trouve ainsi des historiens, des philosophes, des philologues, des juristes ou des archivistes-paléographes.

Les médiévistes partagent toutefois bien quelque chose : ils étudient tous des textes de la même période – quoique celle-ci soit très longue et sujette à de grandes discussions - et, partant, ils sont formés à certains savoirs techniques attachés à la spécificité de ces sources. Les dynamiques de collaboration et de cohésion entre les médiévistes reposent ainsi, entre autres choses sur lesquelles nous reviendrons

³⁶ Aux États-Unis, les médiévistes s'organisent dès le début des années 1970 et crée le bulletin *CAMDAP* (*Computer and Medieval Data Processing*). Pour une présentation de ce périodique dans les termes des acteurs voir par exemple René Pellen, « « CAMDAP » (« Computer and Medieval Data Processing »), revue s'intéressant à tout ce qui concerne le traitement par ordinateur des documents médiévaux. Vol. I, n° 1, février 1971 à vol. VIII. n° 1, 1978 », *Cahiers de Civilisation Médiévale* 22, n° 86 (1979): 215-16.

³⁷ Olivier Dumoulin, « La tribu des médiévistes », *Genèses. Sciences sociales et histoire* 21, n° 1 (1995): 120-33.

longuement dans notre travail, sur les difficultés qu'ils rencontrent dans l'accès et la critique de leurs sources.

Pour éclairer les configurations dans lesquelles les médiévistes échangent sur ces difficultés, il faut cependant faire un bref détour par l'histoire de la discipline. Pour le cas de la France, l'autonomisation des études médiévales s'inscrit dans un mouvement de détachement progressif par rapport aux spécialistes des autres périodes historiques. Au fil du XIX^e siècle se développent ainsi les premières institutions spécialisées, puis les premières chaires dédiées dans les facultés de Lettres³⁸.

Parmi ces institutions, l'École Nationale des Chartes fait figure d'institution centrale pour la discipline, et ce jusqu'à aujourd'hui. Créée par Louis XVIII en 1821, elle est devenue, au fil du XIX^e siècle, le creuset de la formation des futurs historiens et archivistes aux méthodes dites d'érudition, et, plus précisément, aux sciences auxiliaires associées à la critique des sources du Moyen-Âge, comme la paléographie, la philologie, la diplomatique, la codicologie, la numismatique, l'héraldique ou la sigillographie. La nature des formations mises en œuvre dans cette institution atteste, certainement plus clairement que chez les modernistes ou les contemporanéistes, de la distinction centrale qui existe au sein de l'histoire médiévale entre érudits et historiens³⁹.

Tout au long du XX^e siècle, cet état de faits a pour principale conséquence l'existence d'une frontière, quoique particulièrement poreuse, entre, d'une part, les activités des médiévistes-archivistes, chargés de l'édition critique des sources et, plus largement, de l'élaboration d'instruments de travail pour la « *tribu* », et celles des médiévistes-historiens, qui utilisent ces instruments de travail pour écrire l'histoire. Cette frontière est capitale pour comprendre que les premières utilisations des ordinateurs en histoire médiévale s'inscrivent dans des domaines d'activités différents, et poursuivent donc des objectifs scientifiques tout à fait distincts.

Le plus célèbre des projets pionniers mettant en œuvre des ordinateurs dans le champ des sciences humaines a justement pour objet la constitution d'un instrument de

³⁸ Dumoulin.

³⁹ Cette distinction est la source de l'une des tensions qui dynamise l'historiographie française de la fin du XIX^e siècle jusque, *a minima*, aux années 1970. Voir par exemple Yann Potin et Julien Théry, « L'histoire médiévale et la « nouvelle érudition ». L'exemple de la diplomatique », *Labyrinthe*, n° 4 (1 octobre 1999): 35-39; Étienne Anheim, « L'historiographie est-elle une forme d'histoire intellectuelle ? La controverse de 1934 entre Lucien Febvre et Henri Jassemin », *Revue d'histoire moderne et contemporaine* n° 59-4bis, n° 5 (2012): 105-30.

travail visant à donner accès à un ensemble de textes du Moyen-Âge. Son initiateur, le prêtre jésuite Roberto Busa (1913-2011), s'était lancé dès 1946 dans une entreprise d'édition sur cartes perforées de l'ensemble de l'œuvre de Saint Thomas d'Aquin, avec l'objectif de créer un index de tous les termes de ce corpus : l'*Index Thomisticus*⁴⁰. Pour en comprendre la spécificité, à la fois du point de vue des technologies employées et des objectifs scientifiques poursuivis, il nous faut revenir sur l'histoire de l'utilisation des instruments de calcul automatique par les historiens, et repartir des premières utilisations des machines mécanographiques dans cette discipline.

Instrumentation des recherches en histoire : des machines mécanographiques aux ordinateurs (États-Unis, France).

À la fin des années 1940, plusieurs collectifs d'historiens mettent sur pied, aux États-Unis, les premières enquêtes assistées d'instruments de calcul automatique. Ils n'utilisent cependant pas encore d'ordinateurs, mais des machines mécanographiques. Leurs objectifs sont par ailleurs tout à fait différents de celui formulée par Busa en 1946, puisque leurs projets ne visent pas à constituer des instruments de travail pour étudier l'œuvre d'un auteur, mais au contraire à analyser, à l'aide de procédures statistiques (tris, classement, dénombrement, etc.), des tableaux de données élaborés sur la base de sources électorales, démographiques ou socio-économiques.

Citons ici 1) le travail de Frank Lawrence Owsley (1890-1956) et de sa femme Harriet Owsley sur les agriculteurs *yeoman* du Sud des États-Unis⁴¹ ; 2) celui de William Osgood Aydelotte (1910-1996) sur les comportements électoraux des membres du Parlement britannique entre 1841 et 1847⁴² ; 3) l'étude des caractères démographiques et socio-économiques des américains au XIX^e siècle, mené par Merle Curti (1897-1996) et sa femme Margaret Curti (1898-1963) sur la base de trois recensements fédéraux de 1860,

⁴⁰ Jones, Roberto Busa, S. J., and the emergence of humanities computing. L'index est aujourd'hui accessible à cette adresse : <https://www.corpusthomicum.org/it/index.age>

⁴¹ Frank Lawrence Owsley, *Plain Folk of the Old South* (Baton Rouge, 1949).

⁴² William O. Aydelotte, « A Statistical Analysis of the Parliament of 1841: Some Problems of Method », *Historical Research* 27, n° 76 (1954): 141-55.

1879 et 1880⁴³ ou encore 4) l'enquête fondée sur les données économiques des archives maritimes du port de Boston pour la période 1697-1714, entrepris dans les années 1950 par Bernard Bailyn (1922-2020) et sa femme, Lotte Bailyn (1930-)⁴⁴.

Dans ces études, les machines que les historiens mettent en œuvre sont du type de celles développées par l'ingénieur américain Hermann Hollerith (1860-1929) pour le recensement américain de 1890. Elles avaient été élaborées dans le cadre d'un appel du Bureau du recensement américain, qui cherchait un moyen d'accélérer le traitement de l'immense masse d'informations collectée au cours de ses enquêtes. Hollerith avait alors proposé un modèle de machine à statistiques (*tabulating machine*) qui reposait sur le principe mécanique des métiers à tisser Jacquard. Au cœur de ce dispositif automatisé se trouvaient les fameuses cartes perforées, qui servaient d'interface privilégiée entre le chercheur et la machine⁴⁵. Dans la première moitié du XX^e siècle, la production de ces machines s'était développée, essentiellement pour des usages industriels et commerciaux, sous l'impulsion d'entreprises privées comme la Tabulating Machine Company aux États-Unis - qui deviendra IBM (pour International Business Machines) en 1924 - ou Bull en Europe.

En France, la première utilisation de ce type d'instrument par des historiens prend forme dans le cadre d'un projet d'étude sur les archives notariales de la période moderne. En 1959, deux modernistes français de la VI^e section de l'École Pratique des Hautes Études, François Furet (1927-1997) et Adeline Daumard (1924-2003), publient un article relatif à l'intérêt de la mécanographie pour traiter ce type d'archives, tout en exposant les difficultés qui résultent de la mise en œuvre de ces nouveaux instruments⁴⁶.

Tout comme les travaux américains cités ci-dessus, Furet et Daumard témoignent du fait que ces instruments rendent possible l'application de procédures statistiques à de très grands ensembles de données. La conclusion de l'article est claire :

⁴³ Merle Eugene Curti et al., *The making of an American community: a case study of democracy in a frontier country* (Stanford, États-Unis d'Amérique: Stanford University Press, 1959).

⁴⁴ Bernard Bailyn et Lotte Bailyn, *Massachusetts shipping, 1697-1714* (États-Unis d'Amérique: Belknap Press of Harvard University Press, 1959).

⁴⁵ Sur ce sujet voir notamment Lars Heide, *Punched-card systems and the early information explosion, 1880-1945* (Baltimore, États-Unis d'Amérique: Johns Hopkins University Press, 2009). Voir encore Martin Campbell-Kelly, « Punched card machinery », in *Encyclopedia of Computer Science*, 2003, 1489-91.

⁴⁶ Adeline Daumard et François Furet, « Méthodes de l'Histoire sociale : les Archives notariales et la Mécanographie », *Annales* 14, n° 4 (1959): 676-93.

« l'histoire sociale à partir des enquêtes sur les minutes notariales a tout à gagner à l'utilisation de la mécanographie »⁴⁷.

Ces recherches se présentent ainsi, en France et aux États-Unis, comme des initiatives pionnières d'utilisation d'outils de calcul automatique en histoire. Elles s'inscrivent toutefois dans la continuité des méthodes de l'histoire quantitative, et profitent essentiellement des capacités de classement et de tri des machines mécanographiques.

Dès le début des années 1940 se développent pourtant de nouvelles machines permettant de réaliser des opérations de manière automatique : les ordinateurs. Ils diffèrent des machines mécanographiques non seulement du fait de leurs capacités de calcul, plusieurs centaines de fois supérieures, mais également parce qu'ils disposent d'une mémoire interne qui permet de programmer des suites d'opérations (*stored-program computer*)⁴⁸. La plus célèbre de ces machines, l'ENIAC, est par exemple élaborée aux États-Unis à partir de 1943 dans le but de calculer des trajectoires balistiques⁴⁹.

Dans les années qui suivent, les milieux académiques commencent à se tourner vers ces ordinateurs. Aux États-Unis, des centres de calculs sont créés dans les universités à partir de la fin des années 1940⁵⁰. Il faut toutefois attendre 1963 pour voir apparaître les résultats d'une première enquête historique assistée par ordinateur sur le sol américain⁵¹. En France, c'est seulement en 1955 que s'implante à l'Institut Blaise Pascal le premier ordinateur, un Elliott 402 importé de Grande-Bretagne, auquel succéda dès l'année suivante l'installation de machines similaires dans les universités, à La Sorbonne, à Grenoble, à Nancy, à Toulouse ou à Lille⁵². Les historiens français développeront toutefois le premier projet ayant recours aux calculateurs électroniques programmables en dehors de ces infrastructures universitaires.

⁴⁷ Daumard et Furet, 693.

⁴⁸ Ceruzzi, *A history of modern computing*; Mounier-Kuhn, *L'informatique en France de la seconde guerre mondiale au Plan Calcul*.

⁴⁹ Voir par exemple Thomas Haigh, Mark Priestley, et Crispin Rope, *ENIAC in action: making and remaking the modern computer* (Cambridge (MA.), États-Unis d'Amérique: The MIT Press, 2016).

⁵⁰ Ceruzzi, *A history of modern computing*.

⁵¹ William O. Aydelotte, « Voting Patterns in the British House of Commons in the 1840s », *Comparative Studies in Society and History* 5, n° 2 (1963): 134-63.

⁵² Michel Grossetti et Pierre-Eric Mounier-Kuhn, « Les débuts de l'informatique dans les universités », *Revue française de sociologie* XXXVI-n°2 (1995): pp.295-324.

Premières utilisations des ordinateurs dans le cadre d'enquêtes historiques en France.

Dans le numéro des *Annales ESC* de 1961, l'archéologue et spécialiste de l'automatique documentaire Jean-Claude Gardin (1925-2013) et l'assyriologue Paul Garelli (1924-2006) présentent ce qui apparaît, dans le contexte français, comme la première utilisation des calculateurs électroniques programmables⁵³. Ce travail a pour objet de « *reconstituer les relations de parenté et d'affaires* »⁵⁴ d'un ensemble de 1500 personnes ayant vécu en Assyrie au XIX^e siècle avant notre ère, en se basant sur une série de tablettes en argile retrouvées sur le site de Kültepe depuis la fin du XIX^e siècle.

Pour ce faire, Gardin et Garelli obtiennent en 1959 un financement conjoint de l'association Marc Bloch et de la Communauté européenne de l'énergie atomique (EURATOM), et profitent des connections des informaticiens avec lesquels ils collaborent pour avoir accès à un ordinateur IBM 650 à l'Université Libre de Bruxelles⁵⁵. De façon tout à fait intéressante pour notre sujet, ce même numéro de la revue publie un autre travail mettant en œuvre une méthode impliquant l'utilisation de machines mécanographiques. Dans la droite ligne des travaux que Bailyn a menés aux États-Unis, le moderniste Jean Delumeau (1923-2020) y présente un traitement mécanographique de sources maritimes.

Par contraste avec le projet de Delumeau, la particularité du projet de Gardin et de Garelli saute aux yeux. Les deux auteurs affirment en effet que leur travail propose de mettre en place

*« une technique de recherche fondée sur un usage des ordinateurs différent de ce que l'on a le plus souvent à l'esprit dans les sciences historiques »*⁵⁶.

⁵³ Jean-Claude Gardin et Paul Garelli, « Étude par ordinateurs des établissements assyriens en Cappadoce », *Annales* 16, n° 5 (1961): 837-76.

⁵⁴ Gardin et Garelli, 847.

⁵⁵ Le travail de Gardin et Garelli a fait l'objet d'une analyse sur laquelle nous nous appuyons ici : Sébastien Plutniak, « Aux prémices des humanités numériques ? La première analyse automatisée d'un réseau économique ancien (Gardin & Garelli, 1961). Réalisation, conceptualisation, réception », *ARCS - Analyse de réseaux pour les sciences sociales / Network analysis for social sciences* 1 (septembre 2018).

⁵⁶ Gardin et Garelli, « Étude par ordinateurs des établissements assyriens en Cappadoce », 838.

En effet, au contraire de Delumeau, ces deux spécialistes de l'histoire ancienne ont élaboré une méthode qui repose non sur des opérations numériques, mais sur la théorie des graphes⁵⁷. Cette comparaison illustre le fait que si ces deux types d'instruments de calcul coexistent à cette période dès 1961, certains acteurs s'emparent des calculateurs électroniques programmables en rupture avec les méthodes développées au contact des machines mécanographiques.

Pour comprendre plus précisément comment les nouvelles possibilités offertes par les ordinateurs sont appréhendées en histoire au début des années 1960, il nous paraît pertinent, plutôt que de reprendre les travaux des historiens de l'informatique, de repartir d'un article sur le sujet publié dans les *Annales ESC* en 1963. Ce texte, écrit par l'historien russe Valentin Alekseevich Ustinov (1938-2015), a été porté à l'attention des éditeurs de la revue par Gardin. Ce dernier propose d'ailleurs, en amont de l'article de Ustinov, un texte de présentation, dans lequel il insiste sur la nécessité qu'il y a à éclairer les historiens sur l'intérêt de ces nouveaux instruments de calcul pour la recherche. Selon lui, le texte de Ustinov constitue à ce titre

*« une source d'information commode, sur le fonctionnement des calculateurs électroniques et sur un type d'application très général dans les sciences humaines : les travaux de classification »*⁵⁸.

Dans son article, Ustinov définit en ces termes l'intérêt de ces nouveaux instruments :

« C'est par leur aptitude à résoudre des problèmes logiques et à retenir en mémoire l'énorme volume des matériaux à analyser que ces calculateurs électroniques offrent des possibilités d'application efficace dans le domaine des sciences humaines. [...] Le développement de la science historique peut en être accéléré, car les calculateurs permettent non seulement d'élever le niveau des recherches historiques, mais encore de raccourcir les délais nécessaires à ces

⁵⁷ Plutniak, « Aux prémices des humanités numériques ? La première analyse automatisée d'un réseau économique ancien (Gardin & Garelli, 1961). Réalisation, conceptualisation, réception ».

⁵⁸ Jean-Claude Gardin, « Présentation », *Annales* 18, n° 2 (1963): 261.

*recherches. L'application des calculateurs électroniques aux humanités permet de libérer les spécialistes qualifiés du côté technique et non inventif de leur travail, qui prend cependant dans la recherche un temps proportionnellement très important. »*⁵⁹

Dans la vision de Ustinov, les ordinateurs présentent donc à la fois des apports du type de ceux offerts par les machines mécanographiques et d'autres avantages qui tiennent à leur capacité de mémoire, et à la possibilité que celle-ci offre de programmer des opérations.

Les historiens médiévistes, qui étaient, à notre connaissance, restés à l'écart des premières utilisations des machines mécanographiques, vont développer de nombreux projets mobilisant des ordinateurs à partir de la fin des années 1960. En 1966, le médiéviste américain David Herlihy (1930-1991) et la médiéviste française Klapisch-Zuber (1936-) lancent la première enquête assistée par ordinateur en histoire médiévale, dans le but d'éditer et d'étudier un recensement de la fin du Moyen-Âge, le *catasto* florentin de 1427⁶⁰.

Plusieurs projets de ce type se développent encore à la même période. Owen Gingerich (1930-) et Emmanuel Poulle (1928-2011) appliquent par exemple le calcul électronique aux tables alphonsines⁶¹. En 1972, Christian Castellani soutient à l'École Nationale des Chartes une thèse sur Carpentras, qui repose sur une exploitation des sources assistée par ordinateur⁶². À la même période, Élisabeth Carpentier (1933-) utilise,

⁵⁹ Valentin A. Ustinov, « Les calculateurs électroniques appliqués à la science historique », *Annales* 18, n° 2 (1963): 263-64.

⁶⁰ Notre deuxième chapitre est consacré à l'analyse de ce projet, qui apparaît comme l'un des plus anciens dans notre contexte. On trouvera les résultats de cette étude de grande ampleur dans David Herlihy et Christiane Klapisch-Zuber, *Les Toscans et leurs familles: Une étude du catasto florentin de 1427*. (Paris: Fondation nationale des sciences politiques, 1978). On pourra encore se reporter à Christiane Klapisch-Zuber, Béatrice Marin, et Nicolas Veysset, « Le *catasto* florentin de 1427-1430 : présentation générale de l'enquête et du code », *L'Atelier du Centre de recherches historiques. Revue électronique du CRH*, 30 novembre 2016.

⁶¹ Les résultats de cette étude sont publiés en 1967 dans les *Comptes rendus de l'Académie des inscriptions et des belles lettres* : Emmanuel Poulle et Owen Gingerich, « Les positions des planètes au Moyen Age : application du calcul électronique aux tables alphonsines », *Comptes rendus des séances de l'Académie des Inscriptions et Belles-Lettres* 111, n° 4 (1967): 531-48.

⁶² La référence de cette thèse n'a pas pu être retrouvée, et nous nous basons ici sur l'article suivant : Lucie Fossier et al., « Vingt ans d'informatique en histoire médiévale », *Actes des congrès de la Société des historiens médiévistes de l'enseignement supérieur public* 20, n° 1 (1989): 501-25. On pourra cependant trouver certaines des conclusions du travail de Castellani dans un article publié la même année dans la revue

elle aussi, des ordinateurs électroniques pour mener ses recherches sur le *contado* d'Orvieto, en vue d'une thèse qu'elle soutiendra en 1975⁶³.

D'autres institutions se dotent au même moment de centres dédiés à l'informatique. En 1966 est créé à Nancy le Centre de Recherche et d'Applications Linguistiques (CRAL) où l'on prévoit de traiter des sources médiévales, et en particulier les actes originaux antérieurs à 1200⁶⁴. En 1970, l'Institut de Recherche et d'Histoire des Textes (IRHT) se dote d'une section informatique. Cette initiative s'inscrit en continuité avec la vocation du laboratoire depuis sa création en 1936 : produire des instruments de travail permettant de donner un accès aux sources de la recherche en histoire. Ces deux institutions collaborent d'ailleurs dès le début des années 1970, sous la forme d'un projet qui implique notamment l'historienne Lucie Fossier (1924-) et l'informaticienne Marion Créhange (1937-), première étudiante à avoir soutenu un doctorat en informatique en France en 1961 à Nancy⁶⁵. Cette entreprise a notamment pour objet de développer une méthode d'exploitation de données extraites de chartes médiévales dont le dépouillement a été opéré à l'IRHT depuis sa création⁶⁶.

Ces quelques projets pour lesquels les médiévistes mobilisent des ordinateurs nous permettent de poser les contours d'un champ de recherche constitué de l'ensemble des projets d'histoire médiévale mettant en œuvre des ordinateurs en France. Nous verrons cependant au fil de cette thèse que ces collectifs sont en lien avec d'autres chercheurs en Italie, aux États-Unis, en Belgique, au Royaume-Uni ou en Allemagne. C'est sur cet ensemble de recherches que nous travaillerons pour examiner la problématique qui nous intéresse, à savoir, les conséquences de l'utilisation des ordinateurs sur les cultures écrites des historiens.

les *Annales ESC* : Christian Castellani, « Le rôle économique de la communauté juive de Carpentras au début du XV^e siècle », *Annales* 27, n° 3 (1972): 583-611.

⁶³ Les traitements informatiques sont donc antérieurs. La publication de ce travail est cependant bien postérieure : Élisabeth Carpentier, *Orvieto à la fin du XIII^e siècle: ville et campagne dans le cadastre de 1292* (France, 1986).

⁶⁴ Voir notamment Hélène Naïs, « Présentation des travaux du Centre de recherche et d'applications linguistiques (CRAL) », *Revue du LASLA*, n° 1 (1970): 71-79.

⁶⁵ Notons ici que Hala Khassiba vient d'entamer un travail de doctorat sur les débuts de l'informatique à Nancy.

⁶⁶ Créhange Marion, Fossier Lucie. « Essai d'exploitation sur ordinateur des sources diplomatiques médiévales. » In: *Annales. Economies, sociétés, civilisations*. 25^e année, N. 1, 1970. pp. 249-284.

Trois problématiques centrales.

Pour mener à bien notre travail, nous avons basé notre étude sur trois axes principaux, qui guideront notre analyse des usages des ordinateurs par les médiévistes.

Le premier concerne les conséquences de l'utilisation des ordinateurs sur les organisations collectives de la recherche. Nous y explorons l'idée centrale que les collectifs de travail, qui se constituent dans ce contexte à différentes échelles, peuvent être étudiés sur la base des textes que les acteurs produisent, manipulent et échangent. C'est ainsi en nous intéressant à l'apparition de nouvelles formes d'éléments textuels ainsi qu'aux transformations de l'usage et de la fonction des éléments textuels préexistants à l'usage d'ordinateur, mais également à l'évolution des modalités de circulations de ces éléments qu'il nous sera possible de comprendre comment l'utilisation des ordinateurs a des conséquences profondes sur les pratiques des médiévistes.

Le deuxième axe de recherche qui structurera notre travail concerne ce que nous qualifions de *pratiques des sources*. Par cette expression, nous entendons mettre l'accent sur l'ensemble des opérations intellectuelles et matérielles qui fonde la démarche des historiens : l'identification des sources dans les collections d'archives, leurs critiques interne et externe, la constitution de corpus et les problématiques qui sont attachées à ces différentes opérations. Là encore, nous nous appuierons, pour analyser l'évolution de ces pratiques, sur les documents de travail produits par les historiens au cours de leurs recherches (bordereaux de perforation, masque de saisie, tableaux, graphiques, listes, etc...).

Le dernier axe de recherche au cœur de notre thèse concerne l'analyse des conséquences historiographiques de ces évolutions sur l'épistémologie de l'histoire et sur les résultats produits par les collectifs qui se sont emparés des ordinateurs. Comment les recours aux ordinateurs ont-ils contribué à l'évolution des méthodes d'analyse des historiens ? Quelles nouvelles approches de problèmes anciens se sont développées grâce à la mise en œuvre des moyens électroniques de calculs ? Et, plus largement, comment ces évolutions se sont-elles transmises des utilisateurs pionniers des ordinateurs à leurs collègues ?

Chacun de ces axes de recherche sera exploré dans chacune des trois parties de cette thèse. Dans la première partie, nous proposons une analyse et une comparaison de

deux projets de recherches menés par des médiévistes français entre 1966 et 1990, aux fins de comparer leurs organisations collectives, les influences extra-disciplinaires qu'ils subissent, ainsi que les méthodes qu'ils mettent en œuvre et les conséquences historiographiques de ces travaux. Ces chantiers ont été choisis, au sein du corpus des premières études utilisant des ordinateurs, pour la diversité de leurs approches historiques (histoire politique, démographie, histoire sociale), de leurs méthodes (lexicométrie, histoire quantitative), de leurs outils de calcul (informatique lourde, micro-ordinateur) et de leurs organisations collectives (taille des équipes, présence d'informaticien, type de financement).

Dans la deuxième partie de cette thèse, nous traitons des processus par lesquels certains collectifs engagés dans l'utilisation des ordinateurs se sont organisés à une échelle méta-collective, dans le but de faire circuler les produits de leurs recherches (éditions de textes sur support informatique, bases de données, programmes informatiques, etc.) ainsi que les méthodes et les techniques qu'ils avaient pu développer au cours de leurs expériences respectives. L'analyse repose essentiellement sur une initiative française, mais à portée européenne, qui démarre avec l'organisation d'un colloque à Rome en 1975 et se poursuit par la mise en œuvre d'une publication en 1979, intitulée *Le Médiéviste et l'ordinateur*. Ce bulletin de liaison devient, dès lors, l'un des vecteurs privilégiés de la circulation des méthodes aussi bien que de la construction d'une culture scientifique commune.

La troisième et dernière partie est, quant à elle, consacrée aux modalités des échanges qui prennent forme dans les pages de ce bulletin de liaison. Deux directions sont explorées : 1) le genre d'article qui s'y développe en relation avec la nécessité de faire circuler de nouveaux types de connaissances techniques et 2) les difficultés qu'ont pu rencontrer les utilisateurs de ces méthodes dans le partage de ces savoirs auprès de certains de leurs collègues, d'un point de vue technique, mais également épistémologique.

PARTIE I

Chapitre 1 : Une étude assistée par ordinateur du *catasto* florentin de 1427 (1966-1978).

En 1964, David Herlihy (1930-1991), jeune historien médiéviste américain, spécialiste de l'histoire économique et sociale de la Toscane médiévale, est nommé professeur à l'université du Wisconsin à Madison. C'est vraisemblablement l'année suivante qu'Emmanuel Le Roy Ladurie (1929-), alors directeur d'étude à la VI^e section de l'École Pratique des Hautes Études (EPHE), lui présente Fernand Braudel (1902-1985) au cours d'un colloque international en Italie⁶⁷. Braudel est alors au faîte de sa carrière académique, et cumule les prestigieuses fonctions de professeur au Collège de France, de président de la VI^e section de l'EPHE et de directeur du laboratoire d'histoire de la VI^e section : le Centre de Recherches Historiques (CRH). Braudel et Le Roy Ladurie proposent à Herlihy de diriger un projet de recherche, dont les objectifs sont précisés par Herlihy dans une lettre du 20 novembre 1966 adressée à Le Roy Ladurie⁶⁸. Il s'agit de réaliser une enquête, mais aussi, selon les mots du médiéviste américain, une « *edition* [...] *for the computer* »⁶⁹ d'un monument d'archive pour l'histoire de l'Italie à la fin du Moyen-Âge : le *catasto* florentin de 1427.

Pour mener à bien cette étude, et la réalisation de cette « édition », Braudel propose une collaboration entre l'université du Wisconsin et le CRH. Il fournira notamment à Herlihy « *des hommes* » pour l'aider dans cette entreprise⁷⁰. Plutôt que des

⁶⁷ Cette rencontre est documentée par une série d'entretiens menée dans le cadre de mon doctorat auprès de Christiane Klapisch-Zuber entre 2016 et 2020.

⁶⁸ Ce chapitre s'appuie en partie, et nous y reviendrons, sur la correspondance des acteurs du projet, conservée dans les archives du Centre de Recherches Historiques de l'École des Hautes Études en Sciences Sociales. Il nous faut préciser ici qu'aucun inventaire ni classement n'a été dressé par les archivistes du CRH pour le fonds qui nous intéresse ici, certainement en raison de ses modestes dimensions (seulement deux cartons) et d'une faible demande pour ce type d'archives. Il nous est donc impossible de rendre compte aux lecteurs d'une cote d'archive pour les documents cités. Pour garantir la possibilité de retrouver ces documents dans ce fonds, nous indiquerons dans le corps du texte les informations permettant de les identifier (type de document, date, titre, auteur, destinataire, etc.). Je remercie Nicolas Veyssset et Patricia Breton, archivistes du laboratoire, pour leur aide et leur sollicitude, sans lesquelles ce travail n'aurait pas pu être mené à bien.

⁶⁹ Ce mot est utilisé par Herlihy dans une lettre du 20 novembre 1966, sur laquelle nous reviendrons. Pendant la période que dure le projet (1966-1978), les acteurs utilisent cependant de nombreux termes pour désigner cet objet : « *edition* », « *data bank* », « *édition simplifiée* », « *édition abrégée* ». Le choix fait dans ce chapitre est d'utiliser un seul de ces termes d'acteurs pour se référer à cet objet : « édition ».

⁷⁰ Nous nous appuyons ici encore sur les entretiens menés auprès de Klapisch-Zuber entre 2016 et 2020.

hommes, ce sera surtout une femme qui sera choisie : Christiane Klapisch-Zuber (1936-). Cette jeune médiéviste est, depuis 1961 et une rencontre décisive avec Braudel ⁷¹, cheffe de travaux à la VI^e section de l'EPHE ⁷². En 1966, elle vient tout juste de soutenir une thèse de 3^e cycle sur la circulation des marbres de Carrare entre 1300 et 1600 ⁷³. Ce travail, applaudi et remarqué dans le champ des études italiennes⁷⁴, fait d'elle l'une des spécialistes de la Toscane de la fin du Moyen-Âge dans les rangs du CRH. En décembre 1966, elle s'engage aux côtés de Herlihy dans cette aventure, et assurera dès lors la direction scientifique de la partie française de l'enquête de cette source monumentale.

Résultat d'un recensement mené entre 1427 et 1430 par l'administration de la ville de Florence dans le cadre d'une loi fiscale, le *catasto* florentin de 1427 est un ensemble documentaire composé de plus de 300 volumes, conservés aux *Archivi di Stato* de Florence et de Pise. Cette source contient des informations démographiques et économiques sur 60 000 foyers de Florence et de son territoire (que les médiévistes désignent sous le nom de *contado*), soit près de 264 000 personnes. La minutie des agents du fisc florentin en fait par ailleurs une source d'une précision et d'une richesse sans pareilles pour la période. Toutes les familles du territoire, même les plus pauvres, ou celles habitant dans les coins les plus reculés, ont rempli à ce moment-là une déclaration qui prenait la forme d'un inventaire complet de toutes leurs formes de richesses⁷⁵.

En raison de ces dimensions, ce matériau n'a cependant été, jusque-là, exploité que dans le cadre de recherches ponctuelles. Dans les années 1960 par exemple, Enrico

⁷¹ Cet épisode, sur lequel nous ne reviendrons pas, est notamment raconté par Klapisch-Zuber dans un entretien récent : Didier Lett et Christiane Klapisch-Zuber, « Entretien avec Christiane Klapisch-Zuber », *Genre & Histoire*, n° 19 (25 août 2017). Ajoutons cependant que pour une raison que nous ignorons, son nom n'apparaît pas dans les rapports d'activité du CRH avant sa nomination comme maître-assistante en 1968.

⁷² Cette fonction, qui correspond au premier échelon de la hiérarchie des titulaires, est octroyée à l'initiative et à la discrétion du président du CRH Fernand Braudel.

⁷³ Christine Klapisch, *Carrare et ses marbres, 1300-1600*, 1966.

⁷⁴ Son travail est remarqué à la fois par les historiens de l'économie et par les historiens de l'art. Voir François Souchal, « Christiane Klapisch-Zuber. Les maîtres du marbre. Carrare 1300-1600. Paris, 1969. (École Pratique des Hautes Études, VI^e section. Ports, routes, trafics, XXV.) », *Bibliothèque de l'École des chartes* 130, n° 1 (1972): 306-8. Ou encore Alberto Tenenti, « Christiane Klapisch-Zuber, Les Maîtres du marbre. Carrare, 1300-1600 », *Annales* 25, n° 5 (1970): 1390-92. Ou encore pour la réception de ce travail aux États-Unis : Richard A. Goldthwaite, « Les Maîtres Du Marbre: Carrare, 1300-1600. By Christiane Klapisch-Zuber. Paris, S.E.V.P.E.N., 1969. Pp. 364. 43F. », *Business History Review* 45, n° 1 (ed 1971): 93-94.

⁷⁵ Pour une présentation de la source, le lecteur pourra se reporter à la première partie du livre publié en 1978 par David Herlihy et Christiane Klapisch-Zuber : David Herlihy et Christiane Klapisch-Zuber, *Les Toscans et leurs familles : Une étude du catasto florentin de 1427*. (Paris: Fondation nationale des sciences politiques, 1978).

Fiumi (1908-1976), historien et démographe italien, a utilisé le *catasto* comme chaînon manquant pour retracer la courbe de la population toscane à la fin du Moyen-Âge⁷⁶. Des historiens basés aux États-Unis, comme Lauro Martines (1927-) ou Raymond De Roover (1904-1972), ont eux aussi puisé dans ces riches informations pour « *dessiner le milieu social des familles ou des individus qu'ils étudiaient* »⁷⁷. Herlihy a lui-même étudié une partie de cette source, et achève en 1966 un livre sur la ville de Pistoia⁷⁸. Enfin, l'historien italien Elio Conti (1925-) vient d'écrire une histoire agraire de certaines zones-témoins du *contado* sur la base du *catasto*⁷⁹.

Le projet proposé par Herlihy à Braudel et Le Roy Ladurie est beaucoup plus ambitieux. Klapisch-Zuber le présente en ces termes dans un article de 1969 :

*« l'ampleur même du document a découragé l'analyse comparative et la difficulté à rapporter constamment les informations démographiques aux données économiques a presque toujours interdit d'exploiter le côté le plus intéressant du document : la simultanéité de l'enregistrement des biens et des personnes. Seul l'ordinateur permet de mettre en corrélation systématiquement une telle masse d'informations »*⁸⁰.

Plusieurs dimensions apparaissent ici. Tout d'abord, à la différence des travaux précédents, Herlihy et Klapisch-Zuber entendent non seulement traiter la source de manière exhaustive, mais aussi enregistrer simultanément ces deux types d'informations pour rechercher systématiquement leurs corrélations. Dans cette optique, l'ordinateur apparaît alors comme un instrument indispensable.

L'idée d'utiliser les ordinateurs pour poursuivre ce projet est en fait proposée par Herlihy dès sa lettre de novembre 1966 à Le Roy Ladurie. Cette initiative n'est pas tout

⁷⁶ Enrico Fiumi, « Stato di popolazione e distribuzione della ricchezza in Prato secondo il catasto del 1428-29 », *Archivio Storico Italiano* 123, n° 3 (447) (1965): 277-303.

⁷⁷ La citation est tirée de Christiane Klapisch-Zuber, « Fiscalité et démographie en Toscane (1427-1430) », *Annales* 24, n° 6 (1969): 1316. Pour les travaux de Martines, voir Lauro Martines, *The social world of the Florentine humanists, 1390-1460* (Princeton (N. J.), États-Unis d'Amérique: Princeton University Press, 1963). Pour ceux de De Roover, voir : Raymond De Roover, *The rise and decline of the Medici bank: 1397-1494* (Cambridge (Mass.), États-Unis d'Amérique: Harvard University Press, 1963).

⁷⁸ David Herlihy, *Medieval and Renaissance Pistoia : The Social History of an Italian Town, 1200-1430*. (New Haven London: Yale University Press, 1967).

⁷⁹ Elio Conti, *La formazione della struttura agraria moderna nel contado fiorentino* (Istituto storico italiano per il Medio Evo, 1965).

⁸⁰ Klapisch-Zuber, « Fiscalité et démographie en Toscane (1427-1430) », 1316.

à fait originale dans le contexte des études historiques aux États-Unis à la fin des années 1960, puisque plusieurs enquêtes d'histoire politique, et quelques projets à caractère prosopographique ont déjà été menés à terme. Elle l'est cependant pour les études médiévales, puisque la totalité de ces entreprises concernait alors l'histoire moderne⁸¹. Dans le contexte français, plusieurs études assistées par ordinateur ou moyens mécanographiques ont aussi été menées avant 1966 : en histoire moderne et contemporaine surtout⁸², en histoire ancienne et archéologie également⁸³, ainsi qu'en démographie historique⁸⁴. En histoire médiévale, l'enquête de Klapisch-Zuber et Herlihy apparait, aux côtés du projet de traitement des actes diplomatiques antérieurs à 1200 lancé la même année par Lucie Fossier au Centre de Recherche et d'Applications Linguistiques (CRAL), comme une initiative pionnière en histoire médiévale⁸⁵.

La singularité du projet dans le champ des études médiévales conduit Herlihy à formuler plusieurs arguments pour motiver cette décision de recourir aux ordinateurs. Tout d'abord, la nature de la source répond parfaitement aux critères dégagés par les historiens modernistes pour justifier ce type d'exploitation⁸⁶. Dans une lettre du 3 mars 1967, adressée à Klapisch-Zuber, il affirme :

⁸¹ Voir l'article suivant, qui liste les projets menés avec l'aide des ordinateurs aux États-Unis avant 1968 : George G. S. Murphy, « Historical Investigation and Automatic Data Processing Equipment », *Computers and the Humanities* 3, n° 1 (1 septembre 1968): 1-13.

⁸² On trouve une première publication concernant une étude assistée par des moyens mécanographiques dès 1959, mais elle ne propose qu'un retour sur expérience : Adeline Daumard et François Furet, « Méthodes de l'Histoire sociale : les Archives notariales et la Mécanographie », *Annales* 14, n° 4 (1959): 676-93. La première publication d'une étude assistée par ordinateur est souvent considérée comme étant : Jean Delumeau, « Méthode mécanographique et trafic maritime : les terre-neuviers malouins à la fin du XVIIe siècle », *Annales* 16, n° 4 (1961): 665-85.

⁸³ On trouve dans une livraison des *Annales* de la même année une autre étude de ce type : Jean-Claude Gardin et Paul Garelli, « Étude par ordinateurs des établissements assyriens en Cappadoce », *Annales* 16, n° 5 (1961): 837-76. Pour une chronologie plus précise de ces premiers travaux, nous renvoyons ici à l'introduction de cette thèse.

⁸⁴ Voir par exemple Etienne Gauthier et Louis Henry, « Crulai. Démographie d'une paroisse normande au XVII^e et au XVIII^e siècles. Présentation d'un cahier de l'I.N.E.D », *Population* 13, n° 2 (1958): 283-86 ; ou encore M. Couturier, « Démographie historique et mécanographie électronique », *Annales de Démographie Historique* 1966, n° 1 (1967): 57-78.

⁸⁵ Voir Hélène Naïs, « Présentation des travaux du Centre de recherche et d'applications linguistiques (CRAL) », *Revue du LASLA*, n° 1 (1970): 71-79. Pour un exposé, plus tardif, et par des acteurs en France, des raisons qui auraient poussé les médiévistes à s'engager plus tardivement que d'autres dans l'utilisation des ordinateurs, voir notamment les deux premières pages de Lucie Fossier et al., « Vingt ans d'informatique en histoire médiévale », *Actes des congrès de la Société des historiens médiévistes de l'enseignement supérieur public* 20, n° 1 (1989): 501-25.

⁸⁶ Le troisième chapitre de cette thèse s'attache, entre autres choses, à présenter l'évolution de la réflexion sur ces critères en histoire, prenant pour cas d'étude une communauté de médiévistes européens entre 1973 et 1979. Leurs utilisations comme les modes de justification du choix d'utiliser l'informatique sont par ailleurs abordés plus précisément dans le chapitre 5. Pour les discussions qui animent, sur ce sujet, la communauté des historiens américains, on pourra, par exemple, partir du livre un peu plus tardif d'Edward

« It really is the kind of massive, satisfactorily accurate survey which computers were designed to handle ».

Mais une autre raison justifie encore l'emploi des ordinateurs pour le traitement de cette source : la possibilité d' « éditer » l'ensemble de la source sous la forme d'un outil documentaire d'un nouveau type. Dans une lettre déjà mentionnée, datée du 20 novembre 1966, Herlihy définit cet aspect du travail qu'il entend mener :

« The project I have been working on is to prepare an « edition » of the Catasto for the computer. This edition, once made, could be cheaply reproduced and utilized by scholars in any part of the world. They could, in other words, « program » the data in any way they wanted, according to their own research goals. Eventually, it seems to me, many such great statistical sources ought to be edited for the computer, so that entire libraries of such data could be accumulated. This project would be an initial step in achieving such a goal and would convey invaluable experience as to what should be included in such computer editions. »

Dans l'optique de Herlihy, l'utilisation d'ordinateurs pour l'étude du *catasto* constitue aussi une expérience à même de poser les bases d'une nouvelle forme d'édition pour les grandes sources statistiques. Elle pallierait, dans ce cas, à l'absence d'une publication exhaustive des documents, qui n'ont été, en 1966, édités qu'en partie, et de façon abrégée, par Bruno Casini⁸⁷, et que l'*Istituto* de Florence se refuse d'engager en raison du coût trop élevé de l'entreprise⁸⁸.

Shorter, dont l'introduction comporte notamment une partie intitulée « When to use the computer », in Edward Shorter, *Historian and the Computer: A Practical Guide* (New York: W W Norton & Co Inc, 1975). Pour aborder plus précisément ces questions dans le contexte américain, nous avons pour projet, après ce travail de doctorat, de consulter les archives de Herlihy conservées à l'université Brown, et dont l'inventaire est disponible à cette adresse : <https://www.riamco.org/render?eadid=US-RPB-of1UF.H3&view=all>

⁸⁷ Bruno Casini, *Il catasto di Pisa del 1428-29* (Giardini, 1964).

⁸⁸ Cette indication provient d'une lettre de Herlihy à Klapisch-Zuber, datée du 23 mars 1967, dans laquelle il explique que l'*Istituto* refuse de publier de longues et complexes tables en raison du rapport entre le coût de cette opération et le nombre de chercheurs intéressés par ces données (« *hight cost and poor market* »).

Deux ambitions croisées motivent donc l'utilisation des ordinateurs dans ce contexte. D'un côté, une ambition heuristique. Les dimensions de la source et la richesse des informations qu'elle contient justifient le recours aux ordinateurs. Ils doivent permettre une nouvelle forme d'exploration de la source, à travers des analyses statistiques qui alimenteront l'étude d'un instantané de l'ensemble de la société florentine au milieu du *Quattrocento*. De l'autre, Herlihy montre une ambition que nous appellerons « documentaire ». Elle se manifeste par le fait que l'« édition » de la source pour l'ordinateur aura pour objet, non seulement de permettre la réalisation de l'ambition heuristique de l'étude, mais devra encore être réutilisable par d'autres historiens⁸⁹. Comme nous le verrons, Herlihy imagine même ce projet comme la première pierre d'une bibliothèque rassemblant un nouveau genre d'éditions sur supports informatiques. Si l'on ne sait rien de l'adhésion de Klapisch-Zuber à cette seconde ambition en 1966, la médiéviste française rejoint en tout cas le dessein de son aîné en cours de route, puisqu'elle affirme dans un article de 1970 :

*« Les données codées et préparées actuellement, exploitées encore de façon limitée, pourront alors être reprises et intégrées à une recherche plus large. Notre souhait est que l'ordinateur en permette non seulement une première analyse, mais aussi la conservation et la réutilisation ultérieures. Historiens et démographes y trouveront sans doute leur compte. »*⁹⁰

L'intérêt de cette étude de cas pour notre sujet dérive de cette double aspiration des acteurs : mener une recherche historique sur l'histoire de Florence, autant qu'offrir à la communauté des médiévistes un instrument d'érudition. Par ailleurs, il faut ajouter que deux autres raisons justifient le choix de ce cas pour notre étude. Tout d'abord, la configuration institutionnelle dans laquelle se développe cette entreprise apparaît comme

⁸⁹ Dans une série d'entretiens menée entre 2016 et 2019, Jean-Philippe Genet développe l'idée qu'il existe une distinction entre projets heuristiques et projets documentaires, dans le contexte plus large du recours aux ordinateurs en histoire. Le cas choisi ici, comme celui analysé dans le chapitre suivant, qui porte justement sur l'un des projets informatiques de Genet, nous permettra de discuter cette distinction formulée par certains acteurs. Nous y reviendrons par ailleurs plus précisément dans le troisième chapitre de cette thèse, qui s'intéresse plus particulièrement aux débats sur les enjeux disciplinaires de l'utilisation des ordinateurs en histoire médiévale en Europe.

⁹⁰ Christiane Klapisch-Zuber, « Sources et méthodes de la démographie médiévale, le Catasto Florentin de 1427-30 », *Actes des congrès de la Société des historiens médiévistes de l'enseignement supérieur public* 1, n° 1 (1970): 61.

l'une des formes paradigmatiques des organisations collectives en histoire médiévale assistée par ordinateur. Nous y reviendrons longuement dans la première partie de ce chapitre, et par contraste dans le deuxième chapitre de cette thèse. Mais c'est encore l'importance de ce projet dans le paysage disciplinaire qui en fait un objet de choix. L'étude du *catasto* de 1427 est non seulement l'un des premiers cas d'utilisation des ordinateurs en histoire médiévale en France, mais elle demeure aussi l'expérience de ce type la plus commentée depuis la fin des années 1970.

De très nombreuses recensions permettent en effet d'analyser la réception du livre *Les Toscans et leurs familles. Une étude du catasto florentin de 1427* en France, aux États-Unis, en Italie et au Canada. De nombreuses communautés scientifiques s'emparent, dès 1975, des conclusions de l'enquête, mais aussi de l'« édition », pour se pencher sur de nouvelles questions démographiques, sociales, économiques, ou pour réviser des théories existantes. Malgré de nombreuses critiques, sur lesquelles nous reviendrons, l'enquête du *catasto* florentin de 1427 se positionne selon tous les spécialistes comme un élément incontournable des recherches sur l'Italie médiévale ou sur la démographie de la fin du Moyen-Âge en Europe. Cet objet d'étude apparaît donc comme un exemple emblématique, peut-être même incontournable, pour étudier les usages des ordinateurs par les médiévistes.

Comment les acteurs s'organisent-ils pour étudier cette immense source ? Quelles opérations mettent-ils en œuvre pour produire cette « édition » informatique du *catasto* ? Et qu'est-ce-que ce nouvel objet éditorial ? Quels raisonnements sont développés pour étudier simultanément les données économiques et démographiques de la source sur la base de cette « édition » ? Comment cette enquête est-elle reçue par les pairs ? Et comment ces derniers réutilisent-ils les produits du travail de Klapisch-Zuber, de Herlihy et de leurs équipes ? Les réponses à ces questions, qui structureront ce chapitre, permettront d'aborder une problématique plus large, celle de l'évolution des *pratiques des sources* qu'ont développées les médiévistes au contact des ordinateurs.

Cette dernière problématique, qui a déjà été régulièrement soulevée pour ce qui est de l'histoire de l'historiographie quantitative⁹¹, ou de la *microstoria*⁹², se pose, dans notre contexte, en des termes renouvelés. Il s'agit en effet de saisir comment une innovation technologique bouleverse les *pratiques des sources*, et, donc, les transformations des rapports que les historiens entretiennent avec leurs documentations. Cette seconde question a, elle aussi, été abondamment posée depuis le début du XXI^e siècle⁹³, mais rarement pour cette période, et rarement sur la base d'une approche empirique appuyée, comme la nôtre, sur les archives qui documentent le travail des historiens. La démarche mise en œuvre ici permettra d'éclairer les relations qu'entretiennent les raisonnements des historiens avec les élaborations d'autres sciences sociales, et de comprendre comment ces raisonnements s'inscrivent dans des contextes matériels étrangers à ceux des humanités numériques telles qu'elles sont pratiquées aujourd'hui.

Pour aborder ces questions, nous nous appuyerons sur le fonds d'archives de l'enquête collective du *catasto*, conservé au CRH à Paris, sur les rapports d'activité du CRH pour la période concernée (1966-1978), mais aussi sur la publication de l'étude en français et sur les recensions précédemment mentionnées. Il faut cependant noter dès

⁹¹ Parmi l'abondante littérature qui porte sur les relations entre l'historien et ses sources en histoire quantitative, nous ne citerons ici que deux exemples d'articles d'acteurs : François Furet, « Histoire quantitative et construction du fait historique », *Annales* 26, n° 1 (1971): 63-75 ; et, pour un regard légèrement plus récent sur cette question : Bernard Lepetit, « L'histoire quantitative : deux ou trois choses que je sais d'elle », *Histoire & Mesure* 4, n° 3 (1989): 191-99.

⁹² On pourra notamment regarder : Jacques Revel et École des Hautes Études en Sciences Sociales, éd., *Jeux d'échelles: la micro-analyse à l'expérience*, Hautes études (Paris: Gallimard: Seuil, 1996) ; ou plus récemment Aurélien Le Coq, « Compte-rendu de la journée d'études « Microhistoire et pratiques historiennes. Échelles, acteurs, formes narratives », Marne la Vallée, 11 mai 2012 », *Memini. Travaux et documents*, n° 15 (31 décembre 2011).

⁹³ Deux ensembles de questions ont particulièrement intéressé les observateurs : les nouvelles modalités d'accès aux ressources et les questions relatives à l'écriture de l'histoire. Certains historiens français et italiens se sont particulièrement intéressés aux impacts de la numérisation des ressources sur le travail des historiens. Voir à ce titre : « Matériaux pour l'histoire de notre temps 2006/2 (N° 82) ». Ces débats se sont par ailleurs articulés avec de nombreuses réflexions sur la place de l'histoire dans le débat public à l'ère digitale : un exemple avec le numéro 37 de la revue *Memoria e Ricerca* : « Public History. Pratiche nazionali e identità globale », dirigé par Serge Noiret et publié en 2011, dans lequel on peut notamment lire Kiran Klaus Patel, « La scienza al tempo di Wikileaks. Riflessioni sulla storia contemporanea nel ventesimo secolo », *Memoria e ricerca*, n° 37 (2011): 155-71; ou Marie-Pierre Besnard, « Notre-Dame de Saint-Lô nella Manica. Una restituzione virtuale in 3D dopo la distruzione durante la Seconda guerra mondiale », *Memoria e ricerca*, n° 37 (2011): 193-212. On pourra encore regarder sur cette question les derniers chapitres de Roy Rosenzweig, *Clio wired: the future of the past in the digital age* (New York: Columbia University Press, 2011). Autour des questions relatives à l'écriture de l'histoire on pourra, par exemple, consulter : Jack Dougherty et Kristen Nawrotzki, éd., *Writing history in the digital age*, Digital humanities (Ann Arbor: University of Michigan Press, 2013).

maintenant que, du fait des conditions matérielles dans lesquelles cette thèse a été rédigée, il ne nous a pas été permis d'entamer un travail sur les sources américaines ou italiennes. De ce fait, de nombreux éléments de cette enquête restent dans l'ombre. Nous avons cependant eu la chance d'avoir accès à un ensemble documentaire très riche, composé d'une correspondance fournie, qui s'étend entre 1966 et 1978, mais aussi de fiches manuscrites produites par Klapisch-Zuber, de cartes, d'extraits de listings, de graphiques, de tableaux, de listes et même d'un cahier de laboratoire sur lequel nous reviendrons.

Ces sources nous permettront d'examiner, dans un premier temps, les conditions de la mise en œuvre d'un tel projet (coût, équipe, équipement). Nous analyserons ensuite les opérations qui ont conduit à la réalisation de cette forme originale d'« édition » du *catasto*, au regard notamment de l'éclatement géographique des équipes (France, Italie, États-Unis) et des injonctions méthodologiques qui dérivent de la décision d'utiliser des ordinateurs. Dans un troisième temps, nous rendrons compte des raisonnements menés par les historiens sur la base de cette « édition », tout en observant les relations entre ces raisonnements et les contextes matériel et historiographique dans lesquels ils s'inscrivent. Nous proposerons enfin une analyse des réutilisations de cette étude, aussi bien de ses conclusions, que de ses données et de sa méthodologie, afin de mieux saisir l'importance et la postérité de cette entreprise.

1. Organisations collectives et ordinateurs.

En 1968, nous y reviendrons, la collaboration entre le CRH et l'université du Wisconsin bat de l'aile. Dans une lettre du 14 novembre à Le Roy Ladurie, Herlihy motive en des termes très clairs le caractère nécessairement collectif du projet mis en place avec Klapisch-Zuber depuis 1966 :

« I [...] believe that the magnitude of such enterprise, demanding so much money, effort, and sharing of skills, ought to promote some sort of international cooperation ».

Ces différents éléments mis en lumière par Herlihy pour appuyer la nécessité d'une collaboration avec le CRH (dimensions de la source, moyens humains et financiers

requis), nous permettent de soulever de nombreuses questions centrales pour notre sujet. Elles concernent les difficultés qu'engendre l'analyse par des moyens informatiques d'une source des dimensions du *catasto*, du point de vue de son financement comme des moyens humains et matériels qu'elle suppose. Nous verrons que cette configuration de travail caractérise une évolution de l'organisation des collectifs de la recherche en histoire médiévale.

Trois ensembles de questions guideront notre analyse de l'organisation collective de cette enquête : 1) Quel est le coût d'une telle entreprise ? Comment a-t-elle été financée ? Et quels sont les impacts de ces problématiques pécuniaires sur les choix scientifiques faits par les acteurs ? 2) Comment sont choisis les collaborateurs ? Et quelles sont les compétences qui sont mises en commun au sein de ces collectifs ? 3) Enfin, comment les acteurs organisent-ils l'enquête d'un point de vue matériel ? De quels équipements ont-ils besoin ? Et comment partagent-ils les données et les programmes ?

Pour répondre à ces questions nous nous appuierons essentiellement sur la correspondance entre Herlihy et Klapisch-Zuber, conservée aux archives du CRH. Nous regarderons aussi un ensemble plus vaste de documents, constitué des rapports d'activité du CRH pour la période 1966-1978 et de la littérature sur les enquêtes collectives et l'histoire institutionnelle de la recherche à la VI^e section de l'EPHE.

1.1. Chiffrer et financer l'étude du *catasto*.

1.1.1. Un premier programme de travail.

Quelques mois après l'entrevue avec Braudel et les premiers échanges de lettres avec Le Roy Ladurie, Klapisch-Zuber rencontre Herlihy à Florence en décembre 1966. L'objectif de ce séjour est de mettre en place un programme de travail. Trois points cruciaux sont discutés pendant ces quelques jours⁹⁴. Le premier enjeu est pour Herlihy de présenter à Klapisch-Zuber la méthode d'enregistrement des données sur laquelle il travaille depuis quelques mois. Le deuxième est de déterminer le coût du projet. Le

⁹⁴ La teneur de cette rencontre est renseignée par les lettres échangées entre Herlihy et Klapisch-Zuber en janvier, février et mars 1967.

troisième objectif de cette rencontre, que nous aborderons dans un second temps, concerne les modalités de la répartition du travail entre la France et les États-Unis.

Dès 1966, Herlihy commence à concevoir le code qui présidera à l'établissement de règles pour l'enregistrement des données. Les détails de ce travail nous sont parvenus par une lettre, précédemment citée, qu'il adresse à Le Roy Ladurie le 20 novembre 1966. Herlihy y détaille notamment les variables qu'il projette d'étudier et d'inclure dans cette « *computer edition* » : noms et statuts des chefs de familles, âges et sexes des membres des familles, statut des femmes (mariées, fiancées, veuves, célibataires, esclaves), type de résidences (grande ville, petite ville, montagnes, plaines, collines), propriétés, statut migratoire, nombre d'individus par feu, valeur des investissements monétaires, valeur de la participation à la dette publique, valeur totale des biens taxables et total des dettes.

Pour permettre l'analyse de ces variables par ordinateur, il propose d'éditer les données du *catasto* sur support informatique. Dans ce même courrier de novembre 1966, il définit la forme de l'édition pour l'ordinateur qu'il entend constituer en ces termes :

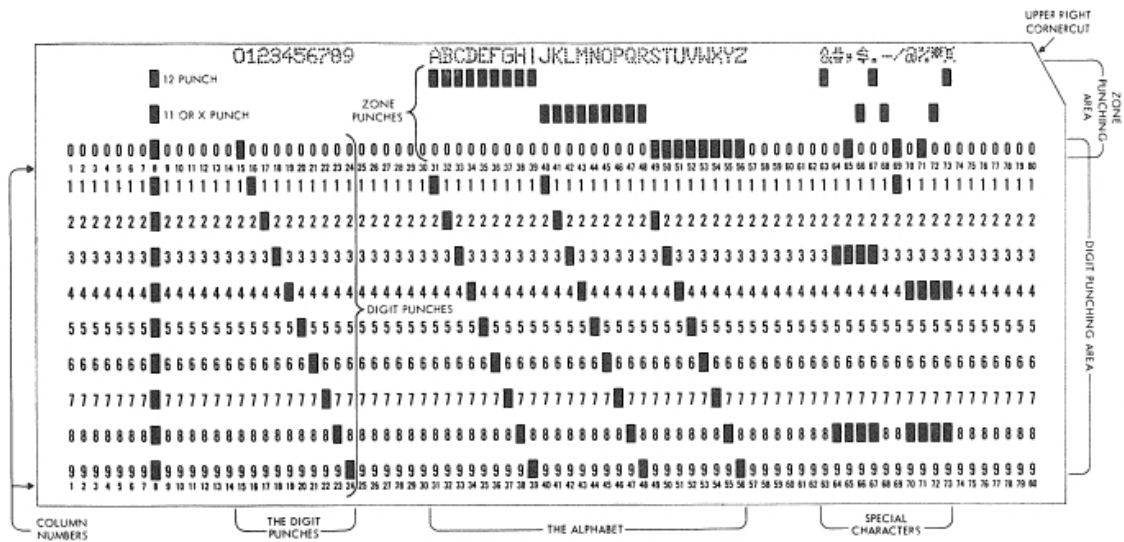
« Once punched, the data will then be read onto tape and this will be the final form of the edition. »

Herlihy mentionne ici deux supports de stockage différents mais complémentaires. Tout d'abord les cartes perforées, inventées par l'ingénieur américain Herman Hollerith (1860-1929) pour traiter à l'aide de machines mécanographiques les données du recensement américain de 1890. Ce support évolue au début du XX^e siècle, avant d'être breveté en 1929 par IBM sous la forme de cartes à 80 colonnes d'environ 20 centimètres sur 8 centimètres, permettant de stocker chacune 80 caractères alphanumériques⁹⁵. Elles servent notamment, selon des modalités sur lesquelles nous reviendrons dans le contexte particulier de l'étude du *catasto*, à entrer en machine des données ou des programmes, pour, notamment, les enregistrer sur un autre type de support, les bandes magnétiques⁹⁶.

⁹⁵ Voir Charles J. Bashe, *IBM's early computers*, 1986.

⁹⁶ Voir par exemple Martin Campbell-Kelly, « Punched card machinery », in *Encyclopedia of Computer Science*, 2003, 1489-91; Margaret Adams, « Punch Card Records: Precursors of Electronic Records », *The American Archivist* 58, n° 2 (1 avril 1995): 182-201.

Figure 1. Carte perforée IBM à 80 colonnes.



Source : Edward Shorter, *Historian and the Computer: A Practical Guide* (New York: W. W. Norton & Co Inc, 1975), 31.

Les bandes magnétiques sont des supports de stockage plus récents dans le domaine informatique. Dans les années 1970, leurs formes industrielles standard est une bobine de 27 centimètres de diamètre, permettant de stocker plusieurs millions de caractères⁹⁷. Elles comportent encore deux autres avantages techniques par rapport aux cartes perforées : leurs vitesses de lecture, excessivement plus rapide, et la possibilité de réécrire les données qui y sont enregistrées depuis l'ordinateur.

Sans être inscrits dans le marbre, ni le code établi, ni la procédure d'enregistrement des données imaginés par Herlihy ne sont remis en cause au moment de la rencontre avec Klapisch-Zuber en décembre 1966. La principale décision sur ce point est cependant le choix de constituer un fichier de cartes perforées en deux exemplaires, le premier conservé à Madison dans le Wisconsin et le second à Paris. Mais comme l'avait souligné Herlihy dès sa lettre de novembre à Le Roy Ladurie :

« *The fly in the ointment is money.* »

⁹⁷ Voir L. D. Stevens, « The Evolution of Magnetic Storage », *IBM Journal of Research and Development* 25, n° 5 (septembre 1981): 663-76.

Pour estimer le budget, les deux médiévistes procèdent d'abord à une évaluation du nombre de cartes perforées nécessaires pour enregistrer toutes les données identifiées par Herlihy. Il est décidé qu'au moins deux cartes IBM seront réalisées par foyer fiscal. La première contiendra des données sur le feu fiscal, tandis que les suivantes concerneront les renseignements sur les membres du feu⁹⁸. Dans une lettre du 2 janvier 1967, Klapisch-Zuber fait le bilan de cette estimation : le travail d'établissement et de perforation de 50 000 cartes, soit au maximum 25 000 familles, est évalué à environ 27 000 francs (soit 0,55 francs la carte). Pour le dépouillement et le codage de l'ensemble de la source, composée de 60 000 familles, les fonds nécessaires sont donc évalués à environ 60 000 francs (soit l'équivalent d'environ 13 000 euros)⁹⁹.

Cependant, ce budget reste encore théorique sans une rencontre avec les acteurs de l'enregistrement de ces données : les entreprises de perforation des cartes et les centres de calcul. Dans une lettre du 19 janvier 1967, Herlihy, qui est resté à Florence, informe Klapisch-Zuber de ses discussions avec son université aux États-Unis :

« I asked the computer center at the University of Wisconsin some months ago for an estimate. But programmers have had little experience with this kind of data analysis and it was difficult for them to judge by the amount of information I was able to provide them at long distance. »

Herlihy devra donc rentrer aux États-Unis pour rencontrer le personnel du centre de calcul et fournir une estimation plus précise. Klapisch-Zuber, quant à elle, a déjà visité l'entreprise de mécanographie, certainement dès son retour de Florence à la fin du moins de décembre 1966. Dans sa lettre du 2 janvier, précédemment mentionnée, elle informe Herlihy de la possibilité d'utiliser 8 000 francs, via un crédit de son laboratoire auprès de cette entreprise de mécanographie. Herlihy a quant à lui déjà proposé de contribuer à hauteur de 1 000 dollars (soit environ 5 000 francs¹⁰⁰). Sur la base de leur estimation, il

⁹⁸ Le contenu de chacune de ces cartes est présenté et analysé dans la partie 2.2. de ce chapitre.

⁹⁹ La conversion a été faite par rapport à la valeur du franc au premier janvier 1967, via le site internet fxtop.com

¹⁰⁰ La conversion a été faite au taux de change de l'année 1966, via le site internet fxtop.com. Les conversions qui suivent ont été obtenues à l'aide du même outil.

leur manque donc environ 14 000 francs pour financer la première phase du projet (soit l'équivalent d'environ 3 000 euros).

Le premier défi pour ces deux chercheurs est donc de trouver des fonds pour lancer leur étude. Sur les conseils de Le Roy Ladurie et de Jacques Le Goff (1924-2014), médiéviste influent et directeur d'étude à la VI^e section de l'EPHE, une demande de Recherche Coopérative sur Programme (RCP) est formulée dès 1967 auprès du Centre National de la Recherche Scientifique (CNRS). Ce type de dispositif, mis en place depuis 1958 par le CNRS, a fait ses preuves pour d'autres enquêtes à partir de la fin des années 1960, notamment dans d'autres disciplines. Comme nous l'apprennent Olivier Orain et Marie-Pierre Sol dans un article de 2007, les RCP devaient être portées par une commission permanente, composée d'au moins un professeur des universités associé à des pairs. D'une durée de trois ans, reconductible une fois, ce cadre administratif avait l'avantage de permettre la fédération de groupes dispersés dans plusieurs institutions, au contraire d'autres cadres, comme les Équipes de Recherche Associée (ERA), qui autorisaient seulement le regroupement de chercheurs au sein d'une même université¹⁰¹.

Un premier ensemble de lettres est envoyé dès 1967 à différents chercheurs dans le but de constituer la commission permanente, qui aura pour rôle d'encadrer le projet. L'une d'entre elles est adressée, le 14 juin 1967, à Michel Mollat du Jourdin (1911-1996)¹⁰², futur président de la commission d'attribution des fonds au CNRS, avec mention d'une demande de 17 500 francs pour l'année 1968 (soit l'équivalent d'environ 3 700 euros). Suivront des courriers à différents spécialistes de l'histoire des villes ou de la Toscane médiévale. Philippe Wolff (1913-2001), professeur à l'université de Toulouse et directeur d'étude à l'EPHE, reconnu pour son rôle dans le renouvellement des études régionales¹⁰³, sera choisi comme directeur de la RCP. Il préfacera le livre de Klapisch-Zuber et Herlihy en 1978. Trois spécialistes de l'histoire de la Toscane sont aussi

¹⁰¹ Voir notamment Olivier Orain et Marie-Pierre Sol, « Les géographes et le travail collectif. La recherche coopérative sur programme à l'œuvre », *La revue pour l'histoire du CNRS*, n° 18 (3 octobre 2007). Nous reviendrons sur ces problématiques de configuration institutionnelle dans le deuxième chapitre de cette thèse, qui traite justement d'un collectif organisé sous la forme d'une ERA.

¹⁰² Il semble pertinent de noter ici pour le futur que Michel Mollat du Jourdin est visiblement un personnage important pour le développement des méthodes informatiques en histoire en France. Il est notamment remercié pour son aide par Jean Delumeau, en note d'un article déjà cité, publié en 1961, qui présente la première étude d'histoire moderne assistée par ordinateur en France : Delumeau, « Méthode mécanographique et trafic maritime ».

¹⁰³ Voir par exemple : Maurice Berthe et Pierre Bonnassie, « In memoriam Philippe Wolff (1913-2001) », *Annales du Midi* 114, n° 237 (2002): 5-24. Il publie notamment en 1958 une monographie remarquée sur la ville de Toulouse : Philippe Wolff, *Histoire de Toulouse*, 1958.

contactés. Charles-Marie de La Roncière (1929-), jeune professeur de 37 ans à l'université de Dakar, travaille sur l'histoire économique de la Toscane¹⁰⁴. Federigo Melis (1914-1973), 53 ans, est titulaire de la chaire d'histoire économique de l'université de Florence. Elio Conti enfin, figure sur laquelle nous reviendrons, a publié en 1965, sur la base du *catasto* de 1427, une étude en trois volumes des structures agraires de Toscane déjà mentionnée en introduction¹⁰⁵.

Des contacts sont aussi pris avec deux spécialistes de la démographie historique : Jacques Dupâquier (1922-2010), maître-assistant à l'EPHE, spécialiste de la période moderne en France, et l'un de ses plus fameux collègues, Louis Henry (1911-1991), chef de file de la démographie historique en France¹⁰⁶. Sur les conseils de ces deux démographes, que nous analyserons précisément dans la suite de ce chapitre, l'enquête va prendre une nouvelle orientation méthodologique, avec notamment l'application des méthodes formelles de Henry pour l'étude de la composition des familles toscanes. La constitution du dossier et la stratégie mise en place pour obtenir un financement additionnel entraînent ainsi la fabrication d'un projet scientifique où vont se mêler méthodes historiques et démographiques.

En novembre 1967, une première bonne nouvelle arrive : la demande de RCP est soutenue à dix-sept voix pour, une voix contre, et quatre abstentions. Le dossier du *catasto* se classe en deuxième position, et le financement du projet est en bonne voie. L'acceptation officielle de la RCP sera finalement annoncée à Klapisch-Zuber par une lettre de Wolff le 8 février 1968. On ne connaît pas, d'après les archives du CRH, le financement versé par le CNRS pour l'année 1968. Pour l'année 1969, le crédit accordé est de 16 000 francs (soit l'équivalent d'environ 3 400 euros). Ce budget permet d'abattre une grande partie du travail de collecte de données. Dans une lettre du 3 avril 1968 à de La Roncière, Klapisch-Zuber impute d'ailleurs cette réussite à l'orientation démographique prise par l'enquête grâce à Dupâquier et Henry.

¹⁰⁴ Il publiera notamment en 1973 : Charles-Marie de La Roncière, *Un changeur florentin du trecento* (S.E.V.P.E.N., 1973).

¹⁰⁵ Conti, *La formazione della struttura agraria moderna nel contado fiorentino..*

¹⁰⁶ Quoiqu'une controverse existe, Louis Henry est considéré par certains comme le père de la démographie historique en France. Voir par exemple Alfred Sauvy, « Vues, travaux, cheminement de Louis Henry », *Population* 32, n° 1 (1977): 7-11.

1.1.2. Crise de confiance.

Malgré la confirmation de cet apport financier, des tensions apparaissent entre le CRH et Herlihy dès l'automne 1968. Elles concernent notamment la répartition du travail entre les équipes de Paris et du Wisconsin, mais aussi l'engagement du CRH dans le projet. Comme en témoigne Klapisch-Zuber dans un document intitulé « *mémorandum* », rédigé à l'intention de Le Roy Ladurie à l'automne 1968, différents éléments de la collaboration entre Herlihy et le CRH ont évolué depuis décembre 1966.

« [Herlihy] *estime disproportionnées les parts entre la France et l'Amérique* ».

On apprend, à la lecture de ce document, mais sans pouvoir en expliquer les raisons, qu'une incompréhension est survenue entre les différents protagonistes de ce projet. En dépit de l'apport financier obtenu auprès du CNRS, Herlihy est convaincu que la majeure partie de l'exploitation informatique sera faite aux États-Unis, mais que le livre sera publié en France. Or, pour Herlihy, la collaboration avait avant tout pour objet le partage des tâches d'enregistrement des données :

« *the real work involved is coding, which is long, hard and supremely dull, but indispensable* ».

Pour lui, la tournure que prennent les événements atteste d'un déséquilibre manifeste dans la répartition du travail : la majeure partie sera faite aux États-Unis, mais la rétribution symbolique, à savoir la publication, sera française. Comme Klapisch-Zuber le note dans ce même « *mémorandum* » :

« [Herlihy] *a la conviction aujourd'hui bien ancrée qu'il est le dindon de la farce* ».

De plus, un autre élément renforce ces tensions : le manque d'engagement des chercheuses que les directeurs d'études Le Roy Ladurie et Braudel ont choisies pour mener à bien cette collaboration. C'est du moins l'avis de Herlihy, qui les met en cause en ces termes dans une lettre du 14 novembre 1968 adressée à Le Roy Ladurie :

« *Mme Klapisch and Mme Triantafillydou are apparently handicapped by financial and family reasons* ».

Derrière ces accusations personnelles, c'est en réalité toute la collaboration avec la VI^e section qui est mise en cause par Herlihy.

Un troisième élément sous-tend en effet les deux autres. Il nous est connu à travers ce même « *mémoire* » de Klapisch-Zuber. Il s'agit d'une promesse d'invitation de Herlihy au CRH pour une durée d'un an, au titre de directeur d'étude, dont on ne sait pas si elle est directement liée à la nécessité de travailler avec Klapisch-Zuber, ou à d'autres considérations. Or, cette invitation tarde à être officiellement formulée. Dans ce même courrier du 14 novembre 1968, Herlihy exige, après trois lettres sans aucune réponse, une invitation « *in frank term and direct language* » de la part de Le Roy Ladurie.

Le 19 novembre 1968, Le Roy Ladurie répond à Herlihy en lui réaffirmant son soutien au projet, en lui rappelant aussi les efforts consentis par la VI^e section de l'EPHE et le CNRS pour assurer sa réalisation. Rassuré, Herlihy répond le 3 décembre 1968 :

« *Thank you for your letter [...] which had helped to clear up the confusion clouding the project.* »

Une promesse d'accueil comme chercheur invité, assortie d'une charge de cours et d'un salaire, est par ailleurs officiellement adressée à Herlihy le 12 décembre 1968. Il passera finalement un an à Paris pendant l'année scolaire 1969-1970. Au début de l'année 1969, Klapisch-Zuber part quant à elle pour un séjour de quatre mois aux États-Unis, à l'université Princeton, avec pour objectif de travailler avec Herlihy. Un programme général de travail, plus équilibré, est mis en place : l'équipe américaine enregistrera les données de la ville de Florence, et l'équipe française celles du *distretto*¹⁰⁷.

Cette crise, qui a failli conduire à l'abandon du projet, montre toute la difficulté de mettre en place un climat de confiance entre des équipes très éparpillées géographiquement. Une fois la collaboration officiellement lancée et les premiers

¹⁰⁷ Le *distretto* réunit « un ensemble de petites « pays » qui avaient préservé leur autonomie jusqu'au début du 14^e siècle », in Herlihy et Klapisch-Zuber, *Les Toscans et leurs familles*, 124. Les auteurs montrent cependant que la distinction *distretto/contado* n'est pas dénuée d'ambiguïté dans le document lui-même.

financements trouvés, les difficultés ne disparaissent pas pour autant, mais redoublent. Elles tiennent, entre autres choses, à l'extraordinaire amas de compétences que nécessite une telle enquête collective.

1.2. Partage de compétences et choix des collaborateurs.

1.2.1. Esprit d'équipe.

Derrière les problèmes qui se jouent autour de la répartition du travail entre les équipes américaine et française en 1967 et 1968, l'un des enjeux majeurs de ces premiers moments de l'enquête est de constituer un collectif de travail, tout en considérant l'éparpillement géographique des chercheurs et des sources (Paris, Madison, Toscane). Les logiques de constitution de cette équipe sont par ailleurs non seulement liées aux compétences scientifiques des acteurs, mais apparaissent aussi à l'observateur comme recouvrant des intentions stratégiques. Quant à ce dernier aspect, il s'agit par exemple pour Herlihy et Klapisch-Zuber de se mettre en relation avec des chercheurs de Florence, dans le but d'avoir des relais, sur place, pour la collecte des données du *catasto* et, dans le même temps, de convaincre certains spécialistes de collaborer à l'enquête en partageant leurs connaissances de cette source.

La première vague de prises de contact est menée par Herlihy, qui depuis son travail sur Pistoia¹⁰⁸, a pu identifier certains spécialistes susceptibles de leur prêter main-forte¹⁰⁹. En 1967, Herlihy rencontre Conti à Florence pour le convaincre de travailler avec eux. Herlihy attend deux types de résultats de cette rencontre. Tout d'abord, il recherche des indications sur la précision des données de la source, afin de décider sur quelles informations s'appuyer pour mettre en œuvre des analyses pertinentes. Cet objectif est révélé dans une lettre adressée à Klapisch-Zuber le 12 mars 1967 : les données qu'ils veulent éditer doivent représenter « *the highest possible level of accuracy* ». Conti lui

¹⁰⁸ Herlihy, *Medieval and Renaissance Pistoia*.

¹⁰⁹ Parmi ces spécialistes on peut notamment nommer Martines et de Roover, mais aussi Conti. Il faut en effet préciser ici que les logiques de conservation du *catasto* font de la constitution du corpus l'un des premiers défis pratiques pour les auteurs. En effet, non seulement l'ensemble documentaire présente de nombreuses lacunes, mais ses différentes parties sont conservées dans différents fonds d'archives de Toscane, en particulier à Pise et à Florence. Nous reviendrons sur les logiques de constitution du corpus dans la deuxième partie de ce chapitre.

répond, comme il l'avait affirmé dans son livre de 1966, que les données du *catasto* sont très justes. Il met cependant en garde Klapisch-Zuber et Herlihy contre une interprétation brute de certaines données d'ordre économique, élément sur lequel les archives du CRH ne donnent pas plus de précision¹¹⁰.

Ensuite, Herlihy propose à Conti d'être le garant de l'enregistrement des données pour la ville de Florence, en lui laissant le dernier mot quant à la vérification des informations indiquées sur les cartes perforées IBM. Malheureusement, cette idée déplait à Conti, qui avance plusieurs éléments pour refuser cette partie de la collaboration. Premièrement, il considère qu'une étude de tout le *catasto* ne donnera pas des résultats différents de son travail sur un échantillon de quinze petits villages ou domaines (*ville*). Ce premier point de désaccord permet de dégager l'un des caractères centraux de ce projet dans le contexte de la discipline dans les années 1960 : le recours aux ordinateurs va de pair avec l'ambition d'exploiter le *catasto* de 1427 de manière exhaustive.

À ce propos, on peut envisager que Conti rejette l'intérêt d'une telle étude par crainte de voir son travail supplanté. Il est cependant tout aussi légitime de penser qu'il considère sincèrement qu'un échantillonnage permet d'obtenir des résultats tout à fait significatifs. Ce « *vieux démon de l'exhaustivité* »¹¹¹, qui se manifeste ici à travers l'enquête du *catasto*, est d'ailleurs attaqué, avec le recul, par plusieurs historiens dans les années 1990 et 2000. D'une part, par des acteurs ayant choisi l'étude exhaustive de leurs sources dans les années 1970 et 1980 mais regrettant cette décision, comme le moderniste André Zysberg¹¹², et d'autre part, par des observateurs, comme plus récemment Claire Zalc et Claire Lemerrier dans un manuel à destination d'étudiants¹¹³.

Cependant, Herlihy avance une autre raison qui motive ce refus. Dans une lettre du 23 mars 1967 qu'il adresse à Klapisch-Zuber, il explique la position de Conti de la sorte :

« He has some doubts about machine processing, apparently for the reason that one cannot express doubts, or add comments, upon IBM cards. »

¹¹⁰ Cet échange est documenté par une lettre adressée par Klapisch à de La Roncière du 3 avril 1968, dans laquelle l'historienne dresse un bilan de l'avancée des travaux.

¹¹¹ André Zysberg, « Impact de l'informatique sur la recherche historique », *Matériaux pour l'histoire de notre temps* 10, n° 1 (1987): 36.

¹¹² Zysberg, « Impact de l'informatique sur la recherche historique ».

¹¹³ Claire Lemerrier et Claire Zalc, « II. Face aux sources : corpus et échantillons », *Repères*, 2008, 19-33.

Ici une raison plus incontestablement liée aux moyens d'enregistrement des données choisis par Herlihy et Klapisch-Zuber apparaît. En effet, on comprend, sur la base de cette citation, que Conti critique l'utilisation des cartes IBM, apparemment en raison du fait qu'elles contraignent l'historien à transformer sa méthode de critique des sources. Contrairement à d'autres pratiques plus « traditionnelles », cette façon de faire force l'historien à prendre, en amont de l'analyse, tout un ensemble de décisions tenant à l'interprétation des documents, sans par ailleurs pouvoir enregistrer sur ces cartes les doutes et les hésitations qui peuvent lui apparaître à la lecture des sources.

Nous découvrons ici un élément fondamental qui caractérise les projets informatiques en histoire. Il repose sur les transformations de la méthode critique qu'implique le processus d'enregistrement des données sur cartes perforées IBM à 80 colonnes, sur lesquelles nous reviendrons plus précisément dans la deuxième partie de ce chapitre. Deux choix scientifiques s'opposent ici, et chacun présente des avantages. D'un côté, Conti entend rester proche des sources, mais ne peut de ce fait traiter de manière exhaustive et simultanée les différents types de données d'un ensemble documentaire des dimensions du *catasto*. De l'autre, Herlihy et Klapisch-Zuber proposent une exploitation exhaustive de la source, au détriment d'une forme plus souple et plus fine de critique du matériau d'analyse.

Par ailleurs, le refus de Conti de participer au codage des données extraites du *catasto* dépasse la simple critique méthodologique : elle engage une vision différente de la collaboration entre historiens. Un premier élément de cette différence tient aux formes du travail collectif. Au contraire de Klapisch-Zuber et Herlihy, qui s'engagent dans une enquête internationale, structurée autour de deux équipes de recherche sous l'égide d'une structure administrative commune (la RCP), Conti travaille, pour la collecte et les vérifications des informations aux archives, en s'appuyant sur des étudiants, ou des membres de sa famille. Il tire ainsi profit du fait qu'en Italie les étudiants avancés préparaient alors une *tesi di laurea*, qui consistait à transcrire et présenter des parties d'archives manuscrites¹¹⁴. On voit ici comment l'organisation sociale de la recherche et l'institutionnalisation de la formation des historiens façonnent les formes de collectifs

¹¹⁴ Cette précision provient d'un entretien mené avec Klapisch-Zuber en 2020. Pour une étude plus approfondie de ce système voir : Stuart Woolf, « On University Reform in Italy: Contradictions and Power Relations in Structure and Function », *Minerva* 41, n° 4 (2003): 347-63.

dans lesquels travaillent les chercheurs, à travers la mise en œuvre de dispositifs de collecte des informations et de vérification de celles-ci.

Le deuxième élément qui illustre la distance entre ces deux conceptions du travail collectif concerne les modes de publication des outils de travail. À la différence de Klapisch-Zuber et Herlihy, qui entendent créer une « édition » de la source enregistrée sur bande magnétique, Conti préfère laisser les fiches comportant les dénombrements qu'il a réalisés sur la base du *catasto* à l'*Istituto* de Florence, où, comme Herlihy le note dans une lettre du 23 mars 1967 : « *scholars could go and check, to make sure his affirmations were right* ». Pour Herlihy, cette méthode de partage des dénombrements n'est rien d'autre que limitée et primitive (« *A singularly primitive method* »), et les moyens informatiques doivent au contraire en permettre un partage plus large.

Dans une autre lettre du 20 mars 1967, Herlihy avait justement rendu compte à Klapisch-Zuber de sa consultation, à l'*Istituto*, des documents de Conti. Il y avait découvert un certain nombre d'erreurs dans les dénombrements effectués par le médiéviste italien sur la base du *catasto*. Il remarquait alors :

« *The moral of this is that, in handling a large and complex document such as the catasto, the best scholarship is cooperative scholarship.* »

Cette remarque de Herlihy permet de caractériser l'évolution des dispositifs de vérifications à l'œuvre en histoire médiévale. L'un des apports épistémologiques de l'organisation collective mise en œuvre par Herlihy et Klapisch-Zuber, rarement documenté ailleurs, apparaît ici. La mise en œuvre de cette forme de travail collectif permet une vérification multilatérale des données plus rigoureuse, au moins aux yeux de ses initiateurs. Cette lettre de Herlihy révèle encore l'existence de deux conceptions différentes de la rigueur, qui reposent sur des formes distinctes de coopération entre les collectifs. Dans sa lettre du 23 mars 1967, Herlihy conclut : « *he is not a team man* ». Klapisch-Zuber est du même avis. Elle lui répond quelques jours plus tard, le 7 avril :

« *On ne peut en effet rêver tempérament plus allergique à nos méthodes* ».

Le cas Conti met en lumière des difficultés inattendues pour Klapisch-Zuber et Herlihy dans la conduite de leur projet. Leur choix scientifique, autant que l'organisation collective qu'ils sont en train de mettre en place, supposent que les collaborateurs modifient certains aspects de leur *ethos* scientifique (place pour la critique des sources, forme du travail collectif, conception de la rigueur). Cette organisation collective est corrélée à la spécificité du projet de Klapisch-Zuber et Herlihy, qui repose essentiellement sur le choix d'utiliser des ordinateurs. Cette analyse permet donc d'affirmer que l'usage de l'informatique a des conséquences indirectes en provoquant, par exemple, des mutations dans la constitution des collectifs de travail. Le recours à cet instrument provoque encore des évolutions quant aux valeurs épistémiques défendues par les acteurs - ici la rigueur - mais nous en découvrirons d'autres, notamment dans les deuxième et troisième parties de cette thèse.

En dépit de la réaction de Conti, qui décline finalement la proposition de Herlihy, l'originalité du projet dans ce contexte ne condamne pas l'avancée de l'entreprise. D'autres historiens italiens acceptent d'y participer. C'est notamment le cas de Michele Luzzati (1939-2014), assistant du grand médiéviste italien Cinzio Violante (1921-2001)¹¹⁵ à l'université de Pise, qui accepte de coder cinq volumes du *catasto*, ou encore de Federigo Melis, déjà mentionné, qui offre généreusement l'aide de l'une de ses étudiantes, Madame Stampi.

Ces considérations sur l'esprit d'équipe requis pour devenir collaborateur soulèvent de fait une autre question : pourquoi cette coopération s'est-elle nouée justement entre Herlihy et le CRH, et non avec d'autres institutions géographiquement plus proches des sources, comme l'université de Pise ou de Florence par exemple ? Sur ce point, nous en sommes réduits aux conjectures. La correspondance conservée au CRH ne permet pas d'éclairer les raisons de ce choix du côté français, et l'absence de consultation des archives du côté américain nous empêche d'affirmer si Herlihy a contacté, ou non, d'autres institutions en Europe ou ailleurs, par exemple, pendant la période de crise que nous avons décrite. On peut cependant présumer que certains facteurs tenant à la réputation, aux méthodes et aux relations de la VI^e section de l'EPHE et du CRH avec les États-Unis, ont favorisé la mise en œuvre de cette collaboration.

¹¹⁵ Pour comprendre l'importance de Violante pour l'historiographie italienne, voir par exemple Pierre Toubert, « Nécrologie Cinzio Violante (1921-2001) », *Le Moyen Age* Tome CVII, n° 2 (2001) : 395-97.

1.2.2. Des partenaires évidents ?

Rappelons dans un premier temps les relations scientifiques des travaux des historiens de l' « école des Annales »¹¹⁶ avec les États-Unis. Les spécialistes de l'historiographie américaine s'accordent à dire que les méthodes et les thèmes de recherche des Annales circulent surtout à partir des années 1960 chez les historiens outre-Atlantique¹¹⁷. Ces deux éléments deviennent, pour certains d'entre eux, parties d'un nouveau modèle de l'histoire sociale, qui se caractérise par un attachement à l'étude de toutes les strates de la population, et non plus seulement aux élites ou aux personnages illustres¹¹⁸. Cette histoire « par le bas » s'appuie par ailleurs largement sur les méthodes quantitatives développées dans d'autres sciences sociales¹¹⁹.

Herlihy est l'un de ces historiens intéressés par ces nouveaux thèmes de recherche, et notamment par l'étude de la famille, des femmes ou des jeunes. Dans son livre publié en 1967¹²⁰, il développe une étude de l'ensemble de la population de Pistoia et de sa campagne, en s'appuyant surtout sur des analyses statistiques, mais aussi sur les conclusions de travaux utilisant ce type de méthodes quantitatives en histoire ou en démographie historique. Il y cite notamment les travaux novateurs de Le Roy Ladurie sur l'histoire du climat¹²¹, ou encore les conclusions du travail de certains historiens proches

¹¹⁶ Pour rappel, l'expression « école des Annales » désigne un réseau d'historiens francophones, notamment articulée autour d'une revue, les *Annales ESC*, et d'une institution : la VI^e section de l'EPHE (depuis 1948). Pour une définition rapide voir l'entrée « Annales » du dictionnaire de l'historiographie : Christian Delacroix et al., *Historiographies: concepts et débats* (Paris, France: Gallimard, 2010). Pour une étude plus approfondie voir par exemple André Burguière, *L'école des annales: une histoire intellectuelle*, Histoire (Paris: O. Jacob, 2006) ; ou Peter Burke, *The French historical revolution: the Annales school, 1929 - 89*, Key contemporary thinkers (Cambridge: Polity Press, 1990).

¹¹⁷ Voir Markus Bodler, « Les historiens français et les États-Unis dans les années 1950 et 1960 », *Nuevo Mundo Mundos Nuevos. Nouveaux mondes mondes nouveaux - Novo Mundo Mundos Novos - New world New worlds*, 29 janvier 2010.

¹¹⁸ Sur ce point voir par exemple l'analyse d'un moment clef de discussions autour de ces méthodes en histoire moderne : le colloque de Saint-Cloud de 1965. Marcel David, « Aux sources de l'histoire sociale », *Revue historique de droit français et étranger (1922-)* 46 (1968): 32-62. Pour saisir la réception de ces problématiques aux États-Unis voir par exemple : Philip Benedict, « Between Whig Traditions and New Histories: American Historical Writing about Reformation and Early Modern Europe », in *Imagine Histories* (Princeton University Press, 1999). Pour un regard plus critique sur l'apport des historiens des Annales aux méthodes et aux objets de l'histoire américaine voir encore : Peter Burke, « The "Annales" in global context », *International Review of Social History* 35, n° 3 (1990): 421-32.

¹¹⁹ Voir par exemple : François Dosse, « À l'école des Annales, une règle : l'ouverture disciplinaire », *Hermès, La Revue* n° 67, n° 3 (2013): 106-12.

¹²⁰ Herlihy, *Medieval and Renaissance Pistoia*.

¹²¹ Herlihy, 108.

des Annales comme Yves Renouard (1908-1965)¹²². Mais l'intérêt est, dans ce cas, réciproque : en France, le travail de Herlihy sur Pise, publié en 1958, a, par exemple, fait l'objet, dès 1962, d'une recension dans les *Annales ESC* sous la plume, justement, de Renouard¹²³.

Cet intérêt des historiens des Annales pour les États-Unis tient aussi au mécénat, notamment celui de la fondation Rockfeller, qui avait été l'une des sources de financement à la base de la création de la VI^e section de l'EPHE¹²⁴. Ce type de coopération franco-états-unienne se poursuit dans les années 1950 et 1960, essentiellement à travers les relations de Braudel et Clemens Heller (1917-2002), citoyen américain et « *grand interlocuteur des fondations états-uniennes qui appuient les initiatives de Braudel* »¹²⁵. Elle prend même la forme, à partir de 1966, d'une collaboration entre le CRH et l'Inter-University Consortium for Political Research (IUCPR) de Ann Arbor, créée en 1962 dans l'optique d'archiver les données des activités informatiques d'une vingtaine d'universités américaines¹²⁶. Le principal objet de cette coopération internationale était l'exploitation par ordinateur d'une source contenant une incroyable masse d'informations démographiques, sociales et économiques sur la population française au XIX^e siècle : la *Statistique générale de la France*¹²⁷.

Herlihy trouve donc avec Braudel et le CRH un partenaire non seulement intéressé par les méthodes quantitatives, mais aussi par l'utilisation des ordinateurs pour la recherche historique. La VI^e section de l'EPHE se caractérise même par une tradition mécanographique bien installée au milieu des années 1960. Dix ans plus tôt, les premières enquêtes mécanographiques sur les archives notariales modernes sont lancées. Le célèbre

¹²² Herlihy, 107. Pour une présentation brève du travail de Renouard, voir Guy Fourquin, « Yves Renouard (1908-1965) », *Revue du Nord* 47, n° 184 (1965): 167-69.

¹²³ Yves Renouard, « Destin d'une grande métropole médiévale : Pise », *Annales* 17, n° 1 (1962): 137-45.

¹²⁴ Voir Brigitte Mazon, *Aux origines de l'École des hautes études en sciences sociales : le rôle du mécénat américain (1920-1960)* (Paris: Cerf, 1988).

¹²⁵ Philippe Minard, « Éditorial. Les passeurs de frontières », *Le Mouvement Social* n° 259, n° 2 (21 juillet 2017): 3-11. On pourra encore consulter Hinnerk Bruhns, Joachim Nettelbeck, et Maurice Aymard, éd., *Clemens Heller: imprésario des sciences humaines*, Collection 54 Histoire (Paris: Maison des sciences de l'homme, 2017).

¹²⁶ Voir Jerome M. Clubb, « The Inter-University Consortium for Political Research: Progress and Prospects », *Historical Methods Newsletter* 2, n° 3 (1 juin 1969): 1-5. Pour une présentation des données disponibles en 1979, voir Erik W. Austin, « The Historical Data Resources of the Inter-University Consortium for Political and Social Research », *Historical Social Research / Historische Sozialforschung*, n° 12 (1979): 43-45.

¹²⁷ Gérard Béaur et Béatrice Marin, « La Statistique Générale de la France – Présentation », *L'Atelier du Centre de recherches historiques. Revue électronique du CRH*, 17 mars 2011.

article de Furet et Daumard, publié en 1959, consacre l'intérêt de ce nouvel outil pour l'histoire quantitative¹²⁸. Le rapport d'activité du CRH des années 1966-1967 renseigne même sur l'installation d'une section mécanographique propre à la VI^e section, qui occupe alors le premier étage des locaux du CRH. Marcel Couturier y développe, entre autres, une nouvelle méthode de saisie des données¹²⁹, sur laquelle nous reviendrons.

Enfin, c'est encore son fort attachement aux méthodes collectives qui semble faire du CRH un partenaire idéal pour la conduite de cette étude, puisque les grandes enquêtes collectives sont nombreuses au sein de l'institution pendant les années 1960. Nous allons le voir, la composition de l'équipe du côté français renseigne sur cette culture locale, autant que sur les connaissances et savoir-faire de certains acteurs concernant l'informatisation des méthodes en histoire.

1.2.3. Les ressources d'une forme d'enquête collective.

Pour mieux comprendre la nature des compétences mobilisées dans le cadre d'un tel projet, mais aussi la division du travail qu'il suppose, commençons par regarder brièvement qui sont les acteurs et les laboratoires engagés. L'équipe qui se met en place du côté français entre 1966 et 1967 est avant tout composée d'historiens déjà familiers de ce type d'organisation collective. Le rapport d'activité du CRH de 1966-1967 révèle, par exemple, que Klapisch-Zuber est déjà impliquée, en 1966, dans deux autres enquêtes collectives du CRH : l'une sur les villages désertés, dirigée par Le Roy Ladurie, et une seconde sur la distribution des ordres mendiants en France, menée par Le Goff.

Deux autres historiens sont par ailleurs engagés, du côté français, dans l'entreprise. La première, Yolande Triantafyllidou, sera la collaboratrice principale de Klapisch-Zuber pour la collecte des données entre 1967 et 1976. Au début de l'enquête, Triantafyllidou est vacataire, et participe aux enquêtes collectives lancées par le CRH au titre de spécialiste de la période médiévale. Elle sera, en 1968, nommée cheffe de travaux, et dès lors rémunérée à ce titre par l'EPHE. Un deuxième historien est rétribué par vacations pour les activités de collecte des données. Il s'agit de Spyros Asdrachas (1933-2017), qui prépare alors un doctorat en histoire économique au sein de l'EPHE. Après

¹²⁸ Daumard et Furet, « Méthodes de l'Histoire sociale ».

¹²⁹ Pour une présentation de la méthode, voir par exemple : Marcel Couturier, « Vers une nouvelle méthodologie mécanographique. La préparation des données », *Annales* 21, n° 4 (1966): 769-78.

avoir obtenu une bourse pour venir étudier à Paris en 1965 via la Fondation des Bourses de l'État grec, il obtient son diplôme de docteur en histoire sociale et économique à l'EPHE en 1972, avant d'enseigner à l'École entre 1974 et 1984¹³⁰.

Il faut encore ajouter un acteur majeur de cette entreprise : Michel Demonet, titulaire d'un DEA de statistiques mathématiques obtenu auprès de Jean-Paul Benzécri (1932-2019) à la Faculté des Sciences de Paris. Il s'engage dès 1964 dans un projet d'utilisation des ordinateurs en histoire aux côtés d'Antoine Prost (1933-), alors jeune spécialiste de l'histoire politique contemporaine française et assistant de recherche de Louis Girard (1911-2003), professeur à la Sorbonne¹³¹. Il travaille ensemble à la fois sur des votes et sur des textes, à partir notamment du Barodet, recueil des professions de foi des élus conservé à l'Assemblée Nationale. Figure incontournable de l'utilisation des ordinateurs en histoire en France, Demonet s'associera notamment au laboratoire de lexicométrie de l'ENS de Saint-Cloud, dont il sera question dans le chapitre 2 de cette thèse, avant de devenir un collaborateur régulier du CRH au cours de l'étude internationale, déjà mentionnée, sur la *Statistique générale de la France*.

En 1967, Le Roy Ladurie organise une rencontre avec Klapisch-Zuber, et Demonet devient dès lors l'ingénieur informatique en chef du projet d'étude du *catasto* du côté français, et ce jusqu'à la publication du livre en 1978. Demonet est, en 1967, non seulement un jeune spécialiste de l'analyse de données, qui par conséquent est capable de programmer et d'utiliser un ordinateur, mais il possède aussi un solide bagage historique au regard de ses expériences de recherches précédentes. Fait curieux pour l'observateur que nous sommes, il lui sera demandé, à la fin des années 1970, de préparer une thèse de 3^e cycle d'histoire en son nom propre, en vue d'être recruté définitivement au CRH au titre de maître de conférences¹³².

En plus de ces acteurs clairement identifiés dans les archives françaises, il faut ajouter un autre ensemble de petites mains qui ont contribué de manière décisive à la

¹³⁰ Ces renseignements sont tirés de l'article suivant : Alexandra Sfoini, « In Memoriam: Spyros I. Asdrachas (1933-2017) », *The Historical Review/La Revue Historique* 14, n° 0 (2017): 9-12.

¹³¹ Raymond Huard et Philippe Gut, « Louis Girard 1911-2003 », *Revue d'histoire du XIX^e siècle. Société d'histoire de la révolution de 1848 et des révolutions du XIX^e siècle*, n° 26/27 (1 décembre 2003): 445-46.

¹³² Demonet soutiendra cette thèse en 1985, sous la direction d'Emmanuel Le Roy Ladurie : Michel Demonet, « Tableau de l'agriculture française au milieu du XIX^e siècle : l'enquête de 1852 » (Paris 1 Panthéon-Sorbonne, 1985). Pour une recension de ce travail, voir Rolande Bonnain, « Michel Demonet, Tableau de l'agriculture française au milieu du XIX^e siècle, l'enquête de 1852. », *Études rurales* 121, n° 1 (1991): 275-275.

constitution de l'édition pour l'ordinateur du *catasto*. Parmi les noms mentionnés dans les sources, on peut citer, pour l'équipe américaine, Mahlon Jones, John Muendel ou Barbara Croken, « *étudiants avancés ou jeunes chercheurs* »¹³³ sous la direction de Herlihy, ainsi que, pour les collaborateurs en Italie, Enzo Settesoldi, Michele Luzzati, Mme Stampi et Lilianna Fantacci-Bambagiotti. La mobilisation de cette « *main d'œuvre* », pour reprendre l'expression de Klapisch-Zuber et Marie-Laurence Netter, illustre l'inscription de cette enquête du *catasto* dans un modèle de division du travail qui préexiste, au moins du côté du CRH, et qui est à même d'intégrer des collaborateurs extérieurs¹³⁴.

En effet, en 1966, une dizaine d'enquêtes collectives de ce type, n'ayant pour la plupart pas recours aux ordinateurs, sont conduites au CRH. Le rapport d'activité de 1966-1967 nous apprend que le laboratoire n'emploie pas moins de dix-sept vacataires et dix collaborateurs techniques, essentiellement des femmes, qui participent, à divers titres, au fonctionnement de ce modèle de travail collectif. Il faut encore ajouter à ces « ressources » humaines, un ensemble de ressources techniques mis à disposition par la VI^e section de l'EPHE. En plus du département de mécanographie déjà mentionné, Klapisch-Zuber et Herlihy pourront encore s'appuyer sur le laboratoire de cartographie automatique dirigé par le géographe Jacques Bertin (1918-2010), qui contribuera de manière décisive, nous le verrons, à l'analyse des données extraites du *catasto*.

Cependant, cette formation des collectifs de travail ne s'arrête pas aux portes des milieux académiques. Entre Madison, Paris et Florence, entre 1966 et 1978, des équipes immenses se constituent. En Toscane, deux photographes, fonctionnaires à l'*Archivio di Stato*, sont chargés de produire des microfilms sur la base de la source, afin d'en permettre la consultation à Paris et à Madison. Par ailleurs, des deux côtés de l'Atlantique, les historiens s'attèlent à la saisie des données sur des bordereaux de perforation qui sont ensuite confiés à des entreprises de perforations privées, chargées de produire les cartes perforées IBM. L'attribution des fonds du CNRS en 1968 correspond ainsi au moment de la mise en place d'une équipe monumentale, à cheval sur deux continents, trois pays et quatre institutions. Cette organisation collective se caractérise, parmi d'autres éléments,

¹³³ Herlihy et Klapisch-Zuber, *Les Toscans et leurs familles*, 8-9.

¹³⁴ Marie-Laurence Netter et Christiane Klapisch-Zuber, « Cinquante ans d'histoire au Centre de Recherches Historiques », *Les Cahiers du Centre de Recherches Historiques. Archives*, n° 36 (30 juin 2005).

par les modalités de partage d'un ensemble d'équipements de toutes sortes, indispensable à la mise en œuvre du caractère informatique de l'enquête.

1.3. Partage de matériel.

Le choix d'une méthode qui emploie, entre autres choses, l'informatique, et l'éparpillement géographique des équipes de recherches, posent de nombreuses difficultés matérielles aux acteurs. Cette configuration met en lumière le poids et le coût d'une telle organisation au début des années 1970, bien avant l'apparition des e-mails et des possibilités de partage de données à distance.

1.3.1. Centres de calcul, perforatrices, lecteurs de microfilm, microfilms, programmes, cartes perforées...

Deux problèmes matériels majeurs se posent dès le départ de l'entreprise pour l'équipe menée par Klapisch-Zuber à Paris. Le premier consiste dans la fabrication d'un fichier informatique, stocké sur bandes magnétiques, constitué à partir de sources situées en Toscane. Pour en comprendre les modalités, il faut ici préciser rapidement les contextes matériels dans lesquels s'inscrit cet ensemble d'opérations. Nous reviendrons dans le détail sur les autres aspects de cette dimension du projet dans la deuxième partie de ce chapitre.

Schématiquement, la fabrication du fichier de cartes perforées se fait en quatre étapes, qui correspondent chacune à un changement de support des données. La première étape consiste en la réalisation de microfilms de certains volumes du *catasto*. En effet, si Luzzati et Stampi, s'occupent de la saisie de quelques-uns de ces volumes depuis l'Italie, la plupart des données seront saisies à Paris et à Madison. Le 20 avril 1967, une lettre de Klapisch-Zuber à Herlihy annonce que les premiers microfilms sont arrivés à Paris et que la deuxième étape du travail peut commencer. Elle consiste en l'inscription des données de la source sur un bordereau de perforation, élaboré par Klapisch-Zuber et Herlihy, que les acteurs appellent aussi « *feuille de codage* » ou « *fiche* » dans leur correspondance. Un document de ce type, correspondant à deux cartes perforées IBM à 80 colonnes, doit être rempli pour chaque déclaration du *catasto*.

Figure 2. Bordereau de perforation établi dans le cadre de l'enquête.

Sér.	Famille	Réf.	Nom			Père	Famille				
1	3	7	12		22	32					
Source :											
Vol.	Pp.	C	M	A	I	Oc.	Inv.	Public	Total	Déduct.	Tax.
42	45	48				52	55	60	65	71	76
Sér. & Famille		Membres									
(1-6)		Cte									
Comme ci-dessus		7	9			16		23		30	37
44		51			58		65		72		

Source : Archives de l'enquête du *catasto* conservées au CRH, Paris.

Ces documents sont ensuite transmis aux entreprises de perforations, qui sont chargées, à leur tour, de transformer ces bordereaux en cartes perforées, à l'aide de perforatrices du type de celle reproduite ci-dessous¹³⁵. Ces cartes, une fois retournées au CRH, doivent alors être conservées, et surtout classées, afin qu'elles ne soient pas mélangées les unes avec les autres. Pour ce faire, un meuble de classement d'une capacité de 30 000 cartes est installé dans les locaux du CRH.

La dernière étape consiste en un dernier changement de support : les cartes perforées sont lues par un ordinateur et enregistrées sur bandes magnétiques. L'enjeu de cette dernière opération est d'inscrire les données sur un support qui permette de corriger de façon plus souple les erreurs qui y sont repérées, puisque, en effet, le stockage sur cartes perforées obligeait, pour chaque erreur identifiée, à perforer à nouveau une carte entière.

¹³⁵ Pour une présentation technique des méthodes de perforation des cartes, voir Campbell-Kelly, « Punched card machinery ». Pour une approche centrée sur les acteurs de ces opérations et de leurs compétences, voir par exemple : Julianne Nyhan et Melissa Terras, « Father Busa's Female Punch Card Operatives », in *Debates in the Digital Humanities* (University of Minnesota Press, 2016).

Figure 3. Exemple de perforatrice utilisée à la fin des années 1960 (ici le modèle IBM 029 introduit sur le marché en 1964)



Source : Columbia University Computing History : www.columbia.edu/cu/computinghistory

Dans ce contexte matériel extrêmement lourd, la migration des données des cartes aux bandes magnétiques rend compte d'une volonté d'alléger le dispositif d'enregistrement et de stockage des données. Elle demande cependant une attention toute particulière pour éviter les erreurs. Un cahier, que nous avons retrouvé dans les archives, permettait notamment à Klapisch-Zuber et Demonet de suivre cette migration d'un support vers un autre. Cet objet, si rare dans les fonds d'archives d'historiens, atteste de la lourdeur matérielle qu'impliquent ces différentes opérations de stockage et de migration des données, dans un contexte où une carte d'environ 8 centimètres sur 19 contenait, au maximum, l'équivalent de 960 octets.

Ces quatre étapes entraînent déjà des dépenses importantes : un voyage en Italie pour Klapisch-Zuber et Triantafyllidou en mars 1968, dans le but de définir les volumes qu'il reste à microfilmer, l'emploi des photographes, la réalisation et les envois de microfilms, des achats de lampes pour les lecteurs de microfilms et les vacations de Asdrachas. La perforation entraîne elle-aussi de nombreux frais et nécessite des

investissements matériels adaptés. Une entreprise de perforation, CENPAME, est par exemple engagée pour transformer les bordereaux de saisie en cartes perforées. Les fonds obtenus n'en permettent pas moins d'assurer un rythme de codage soutenu, avec près de 5 000 familles par mois, sans compter les procédures de vérifications des erreurs, soit environ 25 000 familles pour l'année 1967. Malheureusement, en 1968, les crédits vacataires alloués à la rémunération de Asdrachas sont épuisés et le rythme baisse, avec la réalisation de cartes perforées pour seulement 15 000 familles.

Un autre problème matériel majeur pour les acteurs tient à la recherche d'un ordinateur, dans un contexte où ces instruments sont très difficilement accessibles. En 1968, au moment où Klapisch-Zuber rencontre Demonet, ce dernier fait son service militaire au sein de la Direction des Recherches et des Moyens d'Essai. Dans ce cadre, Demonet s'occupe notamment du centre de calcul de l'unité. Les deux chercheurs en profitent et utilisent l'ordinateur de ce service, ainsi que son système de traitement des fichiers nommé OSIRIS, pour vérifier que le fichier informatique en cours de constitution ne comporte pas d'erreurs. Cependant, une fois le service militaire de Demonet terminé, l'accès à cet ordinateur lui est refusé et les deux chercheurs doivent trouver un autre centre de calcul. Le mari de Klapisch-Zuber, Robert Klapisch, physicien, obtient alors du directeur de son laboratoire à l'université d'Orsay, qu'il mette à disposition des historiens un nouvel ordinateur, en attendant que les aides de la RCP mettent fin à cette instabilité. Autour de l'année 1970, sans qu'il soit possible de préciser la date exacte, l'exploitation informatique est finalement transférée au sein d'un des plus grands centres de calcul scientifique d'Europe, géré par le CNRS, et lui aussi établi à Orsay : le Centre Inter-Régional de Calcul Électronique (CIRCE).

Ces pérégrinations de centre de calcul en centre de calcul illustrent bien les difficultés qui résultent de la décision de travailler sur ordinateur pour des médiévistes en France à la fin des années 1960. Difficulté d'accès tout d'abord, dans un contexte où ces infrastructures sont rares, et le terrain privilégié des chercheurs en sciences dites « exactes ». L'organisation du temps de calcul au sein de ces centres redouble par ailleurs cette première difficulté. En effet, dès 1969, les centres de calcul en France fonctionnaient sur la base d'un « ticket modérateur », qui permettait au CNRS de facturer les heures de

calculs aux équipes qui utilisaient ses services¹³⁶. Dans le cadre de l'enquête du *catasto*, le cadre administratif mis en place avec la RCP a ainsi joué un rôle fondamental pour faciliter l'accès aux instruments de calcul électronique.

Il faut aussi insister ici sur l'importance de Demonet dans ce processus de recherche d'un ordinateur pour lancer l'enregistrement des données et la programmation des analyses statistiques. Ses compétences en informatique font de lui l'interlocuteur privilégié entre Klapisch-Zuber et les opérateurs et ingénieurs des centres de calcul. Si aucune archive ne nous permet de saisir comment se passaient ces opérations de traduction interdisciplinaire, nous mettrons en lumière dans le chapitre suivant les impacts de la présence d'un informaticien sur les objectifs des projets des médiévistes, les moyens mis en œuvre et les méthodes d'enregistrement des données.

Une fois ces difficultés surmontées, il faut encore prendre en compte les problèmes de compatibilité entre les environnements de travail français et américains, tout comme les obstacles aux échanges de données et de programmes entre les équipes française et américaine.

1.3.2. Échanges et voyages périlleux.

Pour réaliser l'ensemble des opérations qui aboutiront à la constitution des bandes magnétiques et à la publication d'un livre en 1978, les différents acteurs mobilisés dans cette collaboration échangent massivement pendant ces douze ans de travail : des lettres, comme nous l'avons vu, mais aussi des livres, des feuilles de codage, des cartes perforées, des microfilms, des cartes, des listes de noms de villes, des listes de références, des adresses, des programmes de traitement des données et des résultats d'analyse des données. Les modalités de certains de ces échanges nous paraissent aujourd'hui rocambolesques. Nous présenterons ici deux cas de ce type, découverts dans la correspondance entre Herlihy et Klapisch-Zuber.

Entre 1969 et 1972, Le Roy Ladurie voyage plusieurs fois aux États-Unis. Profitant de ces occasions, Klapisch-Zuber le charge, dans les deux sens du mot, d'acheminer des bordereaux de perforation, des cartes perforées ou des bandes

¹³⁶ Pour mieux comprendre la configuration institutionnelle de la période et le rôle du CIRCE au sein de cette configuration, voir Wladimir Mercouff, « Témoignage : l'informatique dans l'enseignement et au CNRS au début des années soixante-dix. », *Histoire de la recherche contemporaine. La revue du Comité pour l'histoire du CNRS*, n° Tome IV-N°2 (15 décembre 2015): 180-85.

magnétiques à l'équipe de Herlihy, puis d'en ramener dans ses bagages à son retour vers Paris. Pour des raisons pratiques, rappelons-le, il est en effet décidé que deux fichiers complets doivent être créés, l'un à Paris et l'autre dans le Wisconsin. L'échange de données se fait ainsi, contrairement à ce que l'on connaît aujourd'hui, par avion, sur les genoux, d'un directeur d'étude de l'EPHE.

Ce mode d'échange renseigne sur la lourdeur matérielle des technologies informatiques au début des années 1970. Sans possibilité de transfert de fichiers à distance, ces derniers voyagent physiquement, comme tout autre type de marchandises. Une fois les fichiers ou les programmes transportés à bon port, il faut encore s'assurer que ceux-ci sont lisibles par d'autres équipements informatiques.

Le 12 novembre 1971, Herlihy envoie à Klapisch-Zuber deux programmes élaborés aux États-Unis : HHLD et MARSTA. Il y détaille les opérations qu'ils rendent possible, sur lesquels nous reviendrons longuement dans la troisième partie de ce chapitre, et précise à Klapisch la façon de les faire fonctionner sur l'ordinateur. Dans une lettre du 18 novembre 1971, Klapisch lui répond que l'opérateur du CIRCE, un certain Johannin, n'a pas été en mesure de faire lire les programmes à l'ordinateur. Elle transmet la conclusion suivante :

« Votre EXEC 8 est beaucoup trop évolué pour Orsay ».

Cet indice nous renseigne sur l'un des aspects du problème technique qui se joue ici. Tout d'abord, ces deux programmes, développés aux États-Unis par l'équipe de Herlihy sur un ordinateur UNIVAC 1108, dispose d'un système d'exploitation (EXEC 8) plus récent que celui de l'ordinateur dont dispose Klapisch-Zuber et Demonet - à ce moment-là, certainement l'un des IBM de la série 360 disponible au CIRCE. Cette incompatibilité repose donc en partie sur une différence entre les systèmes d'exploitation des ordinateurs français et américains.

Elle ajoute cependant dans cette même lettre, sans donner plus de précisions :

« Nous devons en revenir aux échanges classiques de cartes perforées ».

Cette seconde affirmation semble soulever un autre aspect du problème. Le changement de support des programmes, probablement envoyés par Herlihy sous une forme de stockage plus récente, la disquette, ne permet pas leurs lectures. Sans qu'il soit possible de trancher, l'hypothèse la plus probable est que l'incompatibilité vienne alors d'un problème de compilation du programme. Klapisch-Zuber suggère donc de revenir au mode d'échange antérieur : les cartes perforées. Cependant, Herlihy insiste et lui renvoie une nouvelle lettre le 20 décembre 1971, avec de nouvelles indications à destination de l'opérateur quant à la compilation et aux fonctionnements des programmes. Le 3 janvier 1972, dans une nouvelle lettre, Klapisch-Zuber lui annonce que le problème est résolu, l'un des programmes ayant été lu « *presque du premier coup* ».

Dans ce contexte, la diversité et la multiplicité des difficultés matérielles auxquelles sont confrontés les nombreux collaborateurs de cette enquête les obligent, comme nous allons le voir, à mettre en place de nombreuses règles de travail communes. Les acteurs ne partagent donc pas seulement la longue liste d'objets et de textes mentionnée plus haut, mais ils doivent surtout formaliser des modes de travail communs, pour le dépouillement, pour la saisie, pour le codage mais aussi pour l'analyse statistique des données ainsi produites. Pour mieux comprendre le poids des contraintes que ces difficultés matérielles entraînent sur la formalisation des méthodes historiques dans ce contexte, analysons les pratiques concrètes des acteurs, qui nous fourniront l'occasion d'examiner les méthodes de dépouillement, de saisie et de codage.

2. Opérations de constitution de l' « édition » : collecte des données, code et codage.

À la lumière des configurations matérielle et collective très lourdes décrites plus haut, quels sont les éléments qui nous permettent de saisir la *pratique des sources* que développent les historiens dans ce contexte ? Plusieurs choses entrent dans cette pratique. Tout d'abord, la décision de recourir à des méthodes informatiques et statistiques est corrélée à celle de se pencher sur des sources d'une certaine nature. Ici, le choix se porte sur une source unique, une série de documents avec des propriétés formelles communes,

et non, comme dans d'autres études historiques du même type, sur un ensemble de sources de natures différentes.

Ensuite, pour mener leur enquête sur la base de cette source, le collectif de travail mobilisé se trouve face à une première nécessité : ils doivent « éditer » le *catasto*, sous la forme d'un ensemble de plus de 120 000 cartes perforées, pour en permettre l'enregistrement sur bandes magnétiques puis l'exploitation statistique. Ces opérations de préparation des données répondent encore à l'exigence que se sont assignés Klapisch-Zuber et Herlihy, à savoir, la création d'un *instrument de travail* documentaire, utilisable par d'autres chercheurs. Pour saisir au mieux ces deux éléments de la pratique des sources que développent les acteurs, nous proposons d'étudier dans cette deuxième partie les opérations de production de cette « édition ».

À partir de 1967, l'objectif des acteurs est avant tout d'enregistrer les informations issues du *catasto* d'une manière qui permette d'en garantir l'exacte conformité à celles de la source. Il leur faut encore s'assurer que les informations enregistrées par les différentes sous-équipes, dispersées des deux côtés de l'Atlantique, soient strictement comparables. Dans ce contexte, les impératifs liés aux méthodes choisies par les acteurs se conjuguent avec le problème matériel de l'éparpillement des équipes et des documents. Ils doivent donc mettre en place des règles de travail communes et uniformes pour assurer la constitution d'une « édition » précise et exploitable.

Cette configuration de travail soulève plusieurs questions. Quelles sont les règles communes mises en place pour la collecte des données ? Sur quelles parties de la source s'appuient-ils pour extraire les informations ? Et pour les vérifier ? Quelles sont les règles communes mises en place pour la saisie des données sur les bordereaux de perforation ? Comment est élaboré le code (notion sur laquelle nous reviendrons) ? Enfin, quelles sont les procédures à suivre dans les cas où une information ne s'inscrit pas clairement dans une des catégories définies par le code ? Ces problématiques, qui se doivent d'être explicites entre les acteurs en raison des enjeux scientifiques qu'ils poursuivent, se manifestent dans la correspondance à chacune des étapes de la constitution de cette édition pour l'ordinateur.

2.1. La collecte des données.

2.1.1. Enjeux de la collecte des données.

Le premier enjeu du dépouillement d'une telle source est, comme pour tous les travaux historiques, de s'assurer de la nature des observations que fournissent les documents. Il faut, pour reprendre les bases de la méthodologie critique de Charles Seignobos, rendre le document pareil à une observation bien faite¹³⁷. Pour cela, Klapisch-Zuber applique les opérations et les méthodes traditionnelles de la discipline historique : critique externe et critique interne. Concernant les opérations de critique externe, on retrouve par exemple dans le fonds d'archives du CRH une pochette contenant des fiches portant sur les différentes lois fiscales du début du *Quattrocento* en Italie. Elles contiennent de nombreuses indications sur les raisons qui ont poussé l'administration florentine à mener ce type de recensement, sur l'organisation de ces enquêtes administratives, mais aussi des notes sur les travaux d'autres historiens de la Toscane comme Casini, Conti ou Roberto Sabatino Lopez (1910-1986).

La critique interne du document repose quant à elle sur la comparaison des différentes « couches » de la source. Il existe en effet dans le *catasto* trois séries de documents pour chaque déclaration fiscale faite par un chef de famille. Les *portate* constituent la série comportant les formes originales, brutes, des déclarations des citoyens florentins. Souvent « *autographes* », elles étaient « *apportées aux bureaux du fisc par les contribuables et décrivaient leurs biens et la composition de leur groupe domestique* »¹³⁸. Les *campioni* forment la deuxième couche de la source. Cette série est composée de « *registres où les scribes ont recopié ces déclarations et opéré les calculs fiscaux* ». Enfin, le *Sommario* reporte, « *par ordre alphabétique des prénoms, la liste des*

¹³⁷ Le passage exact auquel nous faisons référence ici est le suivant : « *Un document vaut exactement dans la mesure où il a pour origine une observation bien faite. Il semble donc que la science historique se retrouve, en dernière analyse, semblable à toutes les sciences d'observation ; il semble que la méthode historique repose sur le même principe que toute méthode scientifique, puisque le document, en dernière analyse, est une observation de faits.* », in Charles Seignobos, *La méthode historique appliquée aux sciences sociales* (Paris: Félix Alcan, 1901), 26.

¹³⁸ Christiane Klapisch-Zuber, Béatrice Marin, et Nicolas Veysset, « Le catasto florentin de 1427-1430 : présentation générale de l'enquête et du code », *L'Atelier du Centre de recherches historiques. Revue électronique du CRH*, 30 novembre 2016.

contribuables avec le montant de leur fortune imposable, le nombre de leurs bocche et celui des hommes adultes de la famille frappés d'une taxe personnelle (teste) »¹³⁹.

Par ailleurs, il faut ajouter qu'aucune de ces séries n'est complète, et qu'elles suivent aussi, pour chaque quartier ou province, une chronologie différente, puisque le travail d'établissement du document s'étale de 1427 à 1430. L'une des problématiques qui se pose aux médiévistes est alors de définir laquelle de ces trois séries de documents éditer. Il s'agit encore de répondre à la question suivante : que faire en cas de contradiction entre les informations données par ces différentes couches ? Sans rentrer pour le moment dans le détail de ces problématiques, il faut seulement préciser ici les grands principes mis en place par Klapisch-Zuber et Herlihy : le dépouillement s'appuiera avant tout sur les *campioni*, et les lacunes seront comblées « *en se basant sur les listes des sommaires et sur les descriptions contenues dans les portate* »¹⁴⁰.

Le deuxième enjeu du dépouillement consiste à concevoir le code de cette étude, qui doit permettre le déploiement de l'ensemble des projets prévus. Le terme code est ici employé pour renvoyer à l'établissement d'un ensemble fini de variables, choisies parmi les informations fournies par les *campioni*, et auxquelles les deux médiévistes doivent associer des modalités et un codage. Ces trois éléments trouvent leur expression matérielle dans les bordereaux de perforation, dont un exemplaire est reproduit à la figure 2. Pour établir le code, Klapisch-Zuber et Herlihy doivent appréhender la diversité des informations contenues dans la source, afin de définir les variables pertinentes. Il leur faut encore saisir les nombreuses situations que recourent chacune de ces variables, et définir pour chacune de ces modalités un codage à inscrire sur le bordereau¹⁴¹.

Quant à ce dernier point, sur lequel nous allons revenir au 2.2., le grand problème que pose le dépouillement d'une source comme le *catasto* est avant tout la masse insurmontable d'informations qu'elle contient. Avec plus de 60 000 déclarations, de parfois plus de cinq pages, il est tout simplement impossible de consulter toute la source, puis de regrouper pour chaque variable les modalités en catégories, comme on le ferait

¹³⁹ Klapisch-Zuber, Marin, et Veysset. Pour un aperçu plus concret de la matérialité de ces trois types de documents, voir les versions facsimilées ou retranscrites dans le livre Herlihy et Klapisch-Zuber, *Les Toscans et leurs familles*.

¹⁴⁰ Klapisch-Zuber, Marin, et Veysset, « Le *catasto* florentin de 1427-1430 ».

¹⁴¹ Par exemple, pour la variable métier, on a comme modalité : paysan propriétaire, codé 1 ; paysan louant des terres, codé 2 ; métayer, codé 3, etc... Pour une reproduction complète du code des métiers, voir l'annexe 3 de Klapisch-Zuber, Marin, et Veysset.

pour un ensemble de données plus petit¹⁴². Cette contrainte se conjugue par ailleurs avec un problème technique : l'impossibilité de réécrire les données sur les cartes perforées une fois celles-ci fabriquées. Le travail de codage doit donc se faire pendant le travail de dépouillement, sur la base d'un code qui évolue au fur et à mesure de la découverte de nouvelles modalités.

Ce problème a déjà été bien identifié par un acteur majeur de l'utilisation des ordinateurs au CRH, le directeur de la section informatique Marcel Couturier (1897-1973). Dans un article de 1966, il soulignait, sur la base de sa propre expérience¹⁴³, le coût de cette forme de préparation des données « *au courant de la collecte* »¹⁴⁴ :

« *Le résultat obtenu est clair : l'ensemble de la formalisation et du codage est revenu en moyenne cinq fois plus cher que la perforation* »¹⁴⁵.

Pour réduire les coûts, il proposait dans cet article une autre forme de préparation des données, plus économique, qui reposait, entre autres, sur un dépouillement des sources par lecture à haute voix, un enregistrement au magnétophone et sur un codage automatique des données. Ces deux éléments permettaient selon lui d'alléger « *la peine du chercheur sans accroître de façon correspondante les frais de l'exécution* »¹⁴⁶.

Une lettre du 18 mars 1967 atteste de la tentative de Le Roy Ladurie de convaincre Klapisch-Zuber et Herlihy d'utiliser la méthode de Couturier. Cependant, Herlihy s'y oppose. Les archives ne nous permettent pas d'affirmer pour quelles raisons, et nous en sommes donc ici réduit aux conjectures. Une raison, *a minima*, peut cependant être avancée. L'originalité de la méthode de Couturier supposerait la remise en cause du programme de travail mis en place depuis décembre 1966. De plus, le travail de dépouillement a déjà commencé aux États-Unis, bien avant qu'il ne débute en France. Ce

¹⁴² Pour avoir une idée générale de la pratique des méthodes quantitatives enseignées aujourd'hui dans certaines universités françaises, voir par exemple les étapes que préconisent deux historiennes : Claire Lemerrier et Claire Zalc, *Méthodes quantitatives pour l'historien* (Paris: La Découverte, 2010).

¹⁴³ Voir par exemple une recension portant sur son travail de thèse de 3^e cycle, soutenue en 1965 : Jean Jacquart, « M. Couturier, Recherches sur les structures sociales de Châteaudun : 1525-1789 », *Annales de Démographie Historique* 1970, n° 1 (1971): 410-13.

¹⁴⁴ Couturier, « Vers une nouvelle méthodologie mécanographique. La préparation des données », 769.

¹⁴⁵ Couturier, 770.

¹⁴⁶ Couturier, 778.

changement de méthode impliquerait donc non seulement de former les informaticiens américains à la méthode Couturier, mais aussi de reprendre le travail déjà effectué.

2.1.2. Quelques techniques mises en place.

En raison de sa connaissance de la source, Herlihy fait, au début de l'entreprise, figure de référent dans l'établissement des techniques de dépouillement et de codage. Dans une lettre du 11 février 1967, il suggère à Klapisch-Zuber de commencer par collecter les informations des contribuables des villes, pour lesquelles les déclarations sont plus longues, plus disparates et donc plus difficiles à coder :

« By working first with the city, you would get to know the character of the declarations, and the countryside would therefore present no difficulties ».

Mais au-delà des questions de priorité à donner à telle ou telle partie de la source, la principale difficulté pour les équipes est d'éviter, au cours de la saisie, les erreurs et les oublis liés à la masse de documents à analyser. Par exemple, les historiens ne doivent pas compter deux fois une famille, un quartier ou une paroisse. Ils doivent encore s'assurer de ne pas faire d'erreurs de lecture d'ordre paléographique ou de fautes de copie.

Pour être certains d'éviter ce type d'erreurs, Klapisch-Zuber et Herlihy mettent en place de nombreux moyens de contrôles et de vérifications. En voici un exemple, présenté par Herlihy dans cette même lettre du 11 février 1967 :

« I shall copy from the Sommario the names of the householders onto the coding forms, number them, include information (such as professions) which the Sommario may also reveal and send them to you. You could then, with the microfilm, fill in the remaining information on the basis of the full declarations. The order of the Sommario is the same as the Catasto (or nearly so). You would then have the opportunity to check my readings of names, and have the aid of knowing how I, or the copist of the Sommario, read particular names. »

L'intérêt de ce travail coopératif, par contraste avec d'autres, comme l'organisation au sein de laquelle Conti réalise son travail, apparaît ici clairement. L'échange des

bordereaux de perforation entre les équipes françaises et américaines permet dans ce cas une vérification collégiale et multilatérale des données, notamment pour la lecture des noms de familles. Les différentes « couches » de la source sont, elles aussi, utilisées pour éviter ces erreurs de lecture. L'ordre d'enregistrement du *Sommario* est tourné à l'avantage des historiens pour éviter tout oubli ou toute erreur d'inattention. La rationalité pratique des agents administratifs de Florence au début du *Quattrocento* devient donc un atout pour les deux médiévistes.

Cette stratégie est d'ailleurs poussée, en pratique, bien plus loin que ne le suggère cet extrait. On retrouve en effet dans les archives de nombreuses listes de quartiers, extraites de la source, qui permettent à Klapisch-Zuber, Triantafyllidou et Asdrachas de n'oublier aucune partie de l'État florentin, mais également de communiquer entre eux sur l'avancement du travail. Cet exemple atteste de la mise en place, pour les opérations de saisie des données, de techniques intellectuelles adaptées au maniement collectif d'une source d'une telle taille.

A cette injonction d'exactitude, traditionnelle pour les éditeurs scientifiques, s'ajoutent par ailleurs d'autres contraintes liées à la nature particulière de cette « édition » pour l'ordinateur. Dans la courte présentation de la méthodologie informatique publiée dans *Les Toscans et leurs familles*, Klapisch-Zuber exprime ces difficultés en termes :

« moins encore qu'un éditeur « traditionnel », il [l'éditeur préparant un document pour la machine] n'a le droit d'éditer les informations sélectionnées de manière incomplète ou inconséquente, en trahissant en cours de route ses propres critères de choix ou d'appréciation. Il ne peut en effet justifier par des notes et des variantes les retouches qu'il apporte à son interprétation originelle, les omissions volontaires les significations nouvelles attribuées à son code »¹⁴⁷.

La réalisation de cette « édition » suppose donc un effort de formalisation particulier. L'exigence de conformité entre les informations des *campioni* et celles enregistrées sur les cartes perforées est assurée par des procédés de vérifications à plusieurs mains, et s'appuie sur la rationalité propre à la source. Quant aux règles communes qui permettent d'assurer la constance des « critères de choix ou d'appréciation », elles reposent sur

¹⁴⁷ Herlihy et Klapisch-Zuber, *Les Toscans et leurs familles*, 102.

l'élaboration du bordereau de perforation, et sur les règles qui président au codage des informations extraites de la source.

2.2. Le code et le codage.

La conception du code pose deux grands problèmes à Klapisch-Zuber et Herlihy. Le premier concerne le choix des variables à prendre en compte, et la principale difficulté pour les acteurs est épistémologique. Il s'agit de justifier les choix intellectuels qui conduisent à passer de la source, riche d'une quasi-infinité de variables, à un ensemble fini de données, qui ne peut être qu'une *modélisation* de la source elle-même. Le code doit par ailleurs s'adapter au format des cartes IBM à 80 colonnes, qui permettent chacune l'enregistrement de 80 caractères. Klapisch-Zuber et Herlihy, qui cherchent à limiter le nombre de cartes perforées nécessaire, décident dès le début de l'entreprise de se limiter à deux cartes par déclaration. La première contiendra des données sur le chef de famille et le feu, et la seconde permettra de renseigner les informations sur dix membres du foyer. Dans le cas où le ménage est composé de plus d'individus, il faudra alors ajouter une carte, mais ce type de cas est très rare (moins de 1% des ménages)¹⁴⁸.

Dans cette optique, les acteurs élaborent des bordereaux de perforations spécifiques à l'enquête du *catasto*, qu'ils appellent aussi des « *feuilles de codages* »¹⁴⁹. Ils circuleront entre les équipes françaises, américaines, et les chercheurs italiens chargés de dépouiller certains volumes de la source¹⁵⁰.

¹⁴⁸ Voir le chapitre XVII du livre Herlihy et Klapisch-Zuber, *Les Toscans et leurs familles*.

¹⁴⁹ Herlihy et Klapisch-Zuber, 653.

¹⁵⁰ Il semble nécessaire de préciser ici que les bulletins de perforations retrouvés dans les archives du CRH sont rédigés en français, et qu'il existe donc certainement une version en anglais qui était utilisée par l'équipe américaine.

Figure 5. Liste des abréviations utilisées pour le bordereau de perforation.

Nous donnons ci-dessous en guise d'exemple une feuille de codage où figurent les trois déclarations publiées dans l'Appendice III, doc. 3-5.

Abréviations utilisées :

Sér. : série régionale

Réf. : code de la localité de résidence

Vol. : numéro du volume (le *campione* généralement) de la série d'archives *Catasto*

Pp. : folio (r° ou v°) où débute la déclaration

C. : caractéristique particulière de la déclaration

M. : maison d'habitation (mode de tenure)

A. : animaux (id.)

I. : origine ou destination de la famille

Oc. : métier

Inv. : valeur cadastrale des biens mobiliers (animaux non compris) (en florins)

Public : valeur cadastrale des parts de la dette publique (en florins)

Total : valeur cadastrale des biens possédés (*sostanza lorda*) (en florins)

Déduct. : valeur cadastrale des déductions autorisées (en florins)

Tax. : biens imposables déductions faites (en florins)

Cte : numéro d'ordre de la seconde carte perforée et nombre total de cartes démographiques.

Colonnes 9-15, 16-22, etc. : informations sur les individus composant le ménage. Chaque personne est décrite par

1^{re} colonne : le sexe

2-3^e colonne : l'âge

4^e colonne : l'état matrimonial

5-6^e colonne : la relation de parenté avec le chef

7^e colonne : éventuellement, une particularité personnelle.

Source : Herlihy et Klapisch-Zuber, 653.

L'exemple des professions est, à ce titre, toujours très éclairant¹⁵¹. Faut-il construire une modalité « soldat », qui engloberait des situations différentes telles que *soldato*, « *al soldo* », ou *provigionato*, ou créer une catégorie pour chacun des termes qui

¹⁵¹ Les historiens comme les sociologues s'interrogent depuis longtemps sur la pertinence des professions comme indices de l'existence de groupes sociaux. Au cœur de ces débats se retrouvent notamment les questions qui portent sur la construction des catégories socio-professionnelles. Pour un aperçu de ces débats, on pourra, par exemple, consulter Alain Desrosières et Laurent Thévenot, « Les mots et les chiffres : les nomenclatures socio-professionnelles », *Economie et Statistique* 110, n° 1 (1979): 49-65 ; ou encore, dans le contexte des études historiques, Christophe Charle et al., éd., *Histoire sociale, histoire globale?: actes du colloque des 27-28 janvier 1989* (Paris: Maison des Sciences de l'Homme, 1993).

apparaît dans la source ? Dans ce cas, le choix fait par Klapisch-Zuber et Herlihy est de regrouper les militaires sous le codage 14, en rassemblant tous les serviteurs militaires du gouvernement. Ils construisent ainsi 99 catégories de métiers qui pourraient toutes être discutées très longuement¹⁵². Ces opérations de catégorisation répondent à deux types d'impératifs différents. D'abord, un impératif technique : la place disponible sur les cartes perforées à 80 colonnes ne permet pas d'entrer les termes des déclarations sans multiplier les cartes nécessaires pour la saisie. En faisant le choix de catégoriser, les acteurs n'ont besoin que de deux caractères pour enregistrer une grande variété de profession. Toutefois, cette décision repose aussi sur un impératif épistémologique, puisque l'établissement de catégories va permettre la comparaison entre des groupes d'individus rassemblés au moment de la constitution des catégories (par exemple les soldats (14) et les employés de la commune (13)).

On comprend tout de suite que ces opérations intellectuelles sont en pratique bien plus complexes que ces exemples ne le laissent paraître. Que faire des individus qui déclarent plusieurs métiers ? Et quel codage appliquer dans le cas d'une profession jamais rencontrée auparavant ? Enfin, comment mettre en œuvre des règles communes quant au codage en l'absence d'une connaissance de toutes les modalités de toutes les variables ? Pour répondre à ce type de questions, nous nous intéresserons d'abord aux opérations qui conduisent à l'établissement du bordereau de perforation, *document de travail* central pour les acteurs chargés de dépouiller les volumes du *catasto*. Nous regarderons ensuite les règles communes mises en place par les acteurs pour le compléter.

2.2.1. Choix des variables et interdisciplinarité.

Le 20 novembre 1966, Herlihy envoie une lettre à Le Roy Ladurie pour lui présenter ce qui constitue la première version du projet d'enquête du *catasto*. Cette lettre, unique document qui nous permet de saisir la première forme du code imaginée par Herlihy, donne des indications sur son intérêt limité pour les questions démographiques.

¹⁵² Pour découvrir le code complet des professions, voir l'annexe 3 d'un article déjà cité : Klapisch-Zuber, Marin, et Veysset, « Le *catasto* florentin de 1427-1430 ». Pour une présentation complémentaire on pourra encore se reporter à l'appendice IV du livre David Herlihy et Christiane Klapisch-Zuber, *Les Toscans et leurs familles: Une étude du *catasto* florentin de 1427*. (Paris: Fondation nationale des sciences politiques, 1978). Ou encore au « Codebook of the 1427-29 *Catasto* Data File for Florence » mis en ligne par Robert Burr Litchfield sur le site de l'université Brown : https://cds.library.brown.edu/projects/catasto/newsearch/catasto_codebook.html

Les variables retenues portent en effet dans leur écrasante majorité sur des informations économiques : la valeur du patrimoine des feux et ses diverses composantes (dettes, propriétés foncières, animaux). Sur ces données se greffent des variables d'identité comme l'âge, le sexe et la localisation des feux sur le territoire florentin. Dans cette lettre, le médiéviste américain mentionne seulement deux variables d'ordre démographique : le nombre de bouches par feu et le statut marital des femmes.

En 1967, suite à des discussions entre Klapisch-Zuber et deux démographes que nous avons présentés plus haut, Louis Henry et Jacques Dupâquier, l'enquête va prendre une nouvelle tournure, résolument démographique. Dans une lettre adressée le 3 avril 1968 à de La Roncière, membre de la commission permanente de la RCP 181, Klapisch-Zuber affirme :

« Louis Henry a beaucoup insisté pour que nous tentions une étude de la composition des familles en utilisant les données sur les relations de parenté ».

Pour comprendre l'influence de Henry sur l'évolution des objets de l'étude, il faut cependant revenir rapidement sur deux points. Le premier concerne la position de Henry dans le paysage intellectuel français au milieu des années 1960, son expérience des enquêtes collectives informatisées en démographie, et ses relations avec la RCP 181. Le second concerne la nature des conseils qu'il donne à Klapisch-Zuber au cours d'une série de rendez-vous, qui consacreront la teneur interdisciplinaire de l'étude comme son devenir après publication.

Henry est souvent présenté comme le fondateur de la démographie historique en France¹⁵³. Malgré des avis divergents sur ce statut, les spécialistes s'accordent en tout cas sur le fait qu'il fut le premier à formaliser les méthodes de travail sur les registres paroissiaux, à travers, notamment, son *Manuel de dépouillement et d'exploitation de l'état civil ancien*, publié avec Michel Fleury en 1956. Henri Charbonneau présente, dans un article paru dans *Population* en 1995, l'importance du travail de Henry pour la démographie historique :

¹⁵³ Voir par exemple Sauvy, « Vues, travaux, cheminement de Louis Henry ».

« Pour la première fois dans l'histoire, il devenait possible que deux personnes dépouillent les vieux registres et reconstituent des familles de la même façon, en suivant exactement les mêmes règles techniques »¹⁵⁴.

Cet intérêt pour la formalisation des méthodes de dépouillement et de reconstitution des familles va par ailleurs de pair avec un intérêt pour l'informatisation des méthodes de la démographie. Henry s'y intéresse notamment, en raison de la charge de travail monotone et répétitive que peut représenter la reconstitution des familles¹⁵⁵. On peut percevoir l'impact de l'emploi de l'informatique en démographie sur l'histoire en relevant que les premiers systèmes experts développés dans cette seconde discipline sont justement des programmes de reconstitution automatique de généalogies à partir de registres paroissiaux¹⁵⁶.

Le démographe insiste par ailleurs sur un point clef de l'impact de l'informatique sur les techniques de travail : la nécessité, pour le programmeur, d'explicitier, pour la machine, toutes les opérations intellectuelles qui passent par son « esprit » de démographe. Dans l'article précédemment cité, Charbonneau exprime cette idée en ces termes :

« L'esprit, il [Henry] aimait à le dire, fonctionne très vite et multiplie les opérations simultanées, de sorte qu'on ne sait pas toujours ce que l'on fait exactement : à des intervalles relativement rapprochés, il avait, constatait-il un jour à son grand étonnement, pris des décisions opposées en face d'un même problème. Cela suffisait selon lui à justifier le recours à l'informatique. »¹⁵⁷

¹⁵⁴ Hubert Charbonneau, « Louis Henry et la démographie historique », *Population* 50, n° 6 (1995): 1667. Sur cette évolution des méthodes de la démographie historique, la place de Louis Henry et la méthode dite des fiches de familles, on regardera encore : Paul-André Rosental, « La nouveauté d'un genre ancien : Louis Henry et la fondation de la démographie historique », *Population* 58, n° 1 (2003): 103-36.

¹⁵⁵ Louis Henry évoque précisément cette application dans Louis Henry, « Simulation d'une reconstitution de familles par ordinateur », *Annales de Démographie Historique* 1972, n° 1 (1972): 303-9.

¹⁵⁶ Dans un article plus tardif, Marion Selz présente l'un de ces systèmes et en établit les origines : Marion Selz, « Procédures d'aide à la reconstitution des généalogies », *Histoire & Mesure* 4, n° 1 (1989): 3-19.

¹⁵⁷ Charbonneau, « Louis Henry et la démographie historique », 1668.

L'ordinateur induira, plus largement, ce type de processus de formalisation pour les sciences sociales, ce que François Furet (1927-1997) affirmait pour la discipline historique dans un article contemporain des travaux de Henry¹⁵⁸.

Ces éléments font de Henry une figure tutélaire pour le projet d'enquête du *catasto*. En 1966, il est non seulement directeur de l'Institut National d'Étude Démographique (INED), mais il est encore choisi pour faire partie de la commission permanente de la RCP 181, cadre administratif de l'enquête du *catasto*. Rappelons également que Dupâquier, l'un de ses proches collaborateurs, a le titre de maître-assistant au CRH. Pour Klapisch-Zuber, comme pour les directeurs d'études qui chapeautent le projet, Henry se présente pour ces raisons comme un expert, non seulement de la formalisation des méthodes de la démographie historique, mais encore de leur informatisation.

En 1967, alors que se façonne le dossier pour la demande de fonds au CNRS, au moins deux rencontres sont organisées entre Klapisch-Zuber, Henry et Dupâquier. Les conseils prodigués par les deux démographes vont alors bien plus loin que de simples suggestions. Dans une première note manuscrite, intitulée « *critique de Louis Henry* », Klapisch-Zuber note certaines indications de Henry quant au code et aux méthodes envisagées pour la collecte des données. Il lui conseille notamment de faire, dans un premier temps, pour définir les classes d'âges, une première classification sur 1000 personnes environ, avant de définir fermement les catégories. Ces indications portent encore sur un aspect plus technique, puisque Henry conseille à Klapisch-Zuber d'effectuer certains calculs manuellement avant le codage, l'ordinateur ne les faisant « *pas vite* ». Ce document, qui atteste du rôle de référent démographique et technique de Henry, ne nous donne cependant aucun détail sur la finalité de ces deux indications.

Une autre série de document nous révèle plus précisément la portée de ces consignes sur la modification du code initial. Intitulée « *feuilles d'idées d'exploitation résultant des discussions (avec Louis Henry, Dupâquier, etc.) en 1967* », on y découvre une liste de variables issue d'une discussion avec Henry. L'intérêt de ce document repose sur le fait que les variables sont mises en rapport avec les indicateurs recherchés par les démographes. Le premier ensemble mentionné concerne des variables déjà prises en compte dans le projet initial de Herlihy, comme l'âge, le sexe, le nombre de bouches par

¹⁵⁸ Furet, « Histoire quantitative et construction du fait historique ».

foyer et les variations régionales de ces informations. Les indicateurs qui y sont attachés montrent cependant des intérêts plus spécifiquement démographiques.

C'est, par exemple, le cas de l'enregistrement des nouveau-nés du *catasto*, qui doit permettre, mis en balance avec l'enregistrement du nombre de décès sur une année, d'étudier l'accroissement naturel¹⁵⁹. Dans le code final, on retrouve justement une variable destinée à indiquer ces deux informations. Pour la variable « commentaire » de la seconde carte, la personne chargée de la saisie peut en effet indiquer, pour chaque membre du feu, une valeur entre 0 et 9. Si celui-ci est un nouveau-né ajouté à la déclaration originale en 1428 ou 1429, elle indiquera le chiffre 2. Si le membre du feu a été rayé de la déclaration pour cause de décès, elle inscrira le chiffre 9.

Ce type de modification du code, visiblement suscité par les discussions de Klapisch-Zuber avec les démographes, se manifeste encore plus clairement à la lumière du second ensemble de variables indiqué dans le document de 1967. Il concerne plus directement les indicateurs portant sur les relations de parenté des membres du feu. Comme dans l'exemple précédent, la recherche de certains indicateurs demande, là aussi, une modification du code. Par exemple, l'enregistrement de l'état matrimonial, envisagé par Herlihy seulement pour les femmes, est étendu à tous les individus du feu. De plus, dans le but d'étudier la nuptialité, les noms rayés dans les déclarations pour cause de mariage seront indiqués par le chiffre 8 dans la variable « commentaire ».

Un autre changement de code apparaît encore si l'on compare la liste des variables établie en décembre 1966, le document dont il est question ici et le code final tel que présenté dans le livre de Klapisch-Zuber et Herlihy. Il s'agit cette fois de l'ajout d'une variable pour caractériser les relations de parenté de chaque membre du feu avec le chef de famille. Deux caractères sont prévus sur les bordereaux de perforation, et 99 types de relations de parenté sont définis¹⁶⁰. Ce dernier point concrétise la volonté de Henry

¹⁵⁹ Selon la définition donnée par l'Institut Nationale de la Statistique et des Études Économiques, « *le solde naturel est la différence entre le nombre de naissances et le nombre de décès enregistrés au cours d'une période. On parle d'accroissement naturel, ou d'excédent naturel, lorsque le nombre de naissances est supérieur à celui des décès* ». Voir : <https://www.insee.fr/fr/metadonnees/definition/c1110>. Pour un aperçu de l'intérêt de Henry pour ces questions, on pourra consulter Louis Henry et Paul Vincent, « Rythme maximum d'accroissement d'une population stable », *Population* 2, n° 4 (1947): 663-80.

¹⁶⁰ Le codage est consultable dans l'annexe 4 de l'article précédemment cité : Klapisch-Zuber, Marin, et Veysset, « Le catasto florentin de 1427-1430 ». Il semble par ailleurs que ce code ait été, au moins en partie, emprunté à d'autres enquêtes de démographie historique. Il ne nous a malheureusement pas été possible, pour le moment, de découvrir lesquels.

d'ajouter une dimension portant sur les relations de parentés dans l'étude du *cataso*, dont rendait compte Klapisch-Zuber dans sa lettre à de La Roncière d'avril 1968.

Ces différents éléments montrent le poids des démographes sur la refonte du code. Leurs indications conduisent Klapisch-Zuber et Herlihy à modifier substantiellement la teneur de l'étude, et va conditionner, nous le verrons dans la quatrième partie de ce chapitre, les réutilisations possibles de cette « édition » du *catasto*. L'ajout de ces variables dans le code contribuera à au moins un chapitre complet du livre, et à de nombreux autres éléments fondamentaux, qui feront de cette étude un travail qui compte en démographie historique¹⁶¹. Cependant, une fois le code et le bordereau de perforation effectivement établis, comment se passe en pratique ce codage à plusieurs mains ?

2.2.2. Construction de catégories et problèmes d'interprétation.

Comme nous l'avons vu, les acteurs de la collecte des données partagent un bordereau de perforation qui permet à tous les membres de ce collectif de suivre la même procédure. Klapisch-Zuber nous apprend par ailleurs, dans un entretien mené auprès d'elle en 2016, que des manuels de codage auraient circulé entre les membres de l'équipe, expliquant non seulement le code, mais encore les codages des différentes modalités. Malheureusement, ces manuels sont absents du fonds d'archives du CRH, et il ne subsiste aujourd'hui, pour comprendre le code, que les annexes d'un article déjà plusieurs fois mentionné et la correspondance conservée au CRH. L'existence de ce manuel de codage atteste néanmoins du fait que la question de la formalisation des règles d'interprétation est posée par les acteurs, dès le départ, comme une condition *sine qua non* de la réussite de l'entreprise.

Ces règles, nous l'avons dit, ne peuvent cependant pas être toutes anticipées. La rencontre avec de nouveaux cas fait ainsi automatiquement évoluer ces conventions. Herlihy présente en ces termes le processus de travail, dans une lettre du 19 janvier 1967 adressée à Klapisch-Zuber :

« the great bulk of the work is routine [...] but there are incidents of ambiguity in which it is necessary to follow consistent procedures »

¹⁶¹ Nous traitons de la réception du livre dans la quatrième partie de ce chapitre.

Pour mieux comprendre de quels types d'ambiguïté parle Herlihy, prenons un exemple. Si l'on considère la variable « âge », un individu de 24 ans sera codé 24. Les difficultés sont, sur ce point, minimales. Il faut seulement convenir d'un code qui correspond à une absence d'information dans la source. Dans une lettre du 8 avril 1967, Herlihy informe Klapisch-Zuber qu'il utilise pour ce dernier type de cas le nombre 79. Pour les variables catégorielles, comme les professions, par exemple, les difficultés redoublent. En effet, comme nous l'avons déjà vu, la taille de la source oblige le moment de la collecte des données et celui du codage à se superposer. Des professions apparaissent donc au fur et à mesure du dépouillement, occasionnant des modifications de codage au fil de la consultation des sources. Au final, le code comportera 99 modalités pour la variable « métier », dont environ 60% sont des catégories rassemblant plusieurs termes d'acteurs¹⁶².

Un autre vecteur d'évolution des conventions de codage vient de phénomènes que l'on peut qualifier d'accidentels. Klapisch-Zuber mentionne l'un d'entre eux dans une lettre adressée à Triantafyllidou le 15 octobre 1968 :

« Spyros [Asdrachas] a trouvé une fois un fils d'un figliastro (88) [beau-fils] et il a codé au crayon 49. C'est toi qui as choisi ce chiffre ou bien c'est lui qui a eu cette idée ? Il ne se souvient pas lui-même. De toute façon, si tu es d'accord, nous pouvons adopter ce chiffre comme fils de 88 puisque le 49 actuel est vraiment difficile à rencontrer. Dans ce cas, tu le dis aux autres ».

Dans ce cas précis, une décision personnelle de l'un des membres de l'équipe entraîne une modification des normes de codages, sans qu'il soit même nécessaire de la justifier. La commodité de la solution envisagée face à ce cas inédit (fils d'un beau-fils) transforme le codage des relations de parenté pour l'ensemble des équipes engagées. Dans les cas qui ne créent pas d'ambiguïté, comme celui-ci, ce type de modification ne pose pas de souci.

¹⁶² Voir le code des métiers, reproduit à l'annexe 3 de Klapisch-Zuber, Marin, et Veysset, « Le catasto florentin de 1427-1430 ».

Les problèmes arrivent au contraire, comme le notait Herlihy, lorsqu'apparaît une ambiguïté au sein même de la source. Deux exemples nous serviront ici d'illustration pour présenter les procédures mises en place dans ce second type de cas. Le premier exemple concerne une situation dans laquelle l'historien détecte une erreur dans les informations recueillies par la ville. Doit-il corriger cette erreur, ou la conserver pour rester fidèle à la source, « *au risque de perturber la démographie* », se demande Klapisch-Zuber dans une lettre à Herlihy du 9 juillet 1967. Un exemple de ce type nous est révélé dans un courrier du 12 mars 1967. Herlihy avait remarqué que certains âges enregistrés par les scribes de la ville de Florence n'étaient pas exacts, et formulait en conséquence la règle suivante :

« *We are interested only in the *catasto*, and we don't want to violate its integrity (i.e if both Portate and Campioni list a man as 60 years old, and you know that he is really 64, he should be entered as 60 ; even error can be revealing)* ».

Dans ce cas, le choix fait par les auteurs est de conserver les erreurs de la source, et d'en garder ainsi une image fidèle. La répétition de ce type d'erreurs dans le *catasto* permettra même aux historiens de conclure, comme nous le verrons plus bas, qu'il existait, à Florence, au XV^e siècle, un phénomène d'arrondissement des âges dépendant de variables sociales et culturelles. Un chapitre du livre sera même entièrement consacré à cette question, et particulièrement salué dans diverses recensions¹⁶³. Cet exemple montre bien une facette de la pratique des sources que les auteurs mettent en place au contact du *catasto*. Contrairement à leurs collègues démographes, à la recherche de la « réalité démographique », Klapisch et Herlihy, décident de respecter la source en tant qu'elle reflète les pratiques des acteurs qu'ils étudient.

Mais si la grande majorité des choix de saisie sont le fruit de considérations humaines, certains restent cependant le fait de limites techniques imposées par le format des cartes perforées. Dans une « Note on David Herlihy's Computer Files », publiée sur

¹⁶³ La quatrième partie de ce premier chapitre revient justement sur ces questions, à travers, notamment, la critique positive que le médiéviste Jean-Philippe Genet développe. Dans sa recension du livre, il salue cette analyse qui montre, selon lui, que l'utilisation des statistiques en histoire médiévale est non seulement possible, mais aussi fertile.

le site de l'université Brown¹⁶⁴, Robert Burr Litchfield (1936-) présente brièvement, pour les utilisateurs de l' « édition » informatique du *catasto*, quelques problèmes d'interprétation qu'ils pourraient rencontrer. Premier exemple : certains noms sont tronqués, pour des raisons qui tiennent autant à des contraintes venant de la source qu'aux limites techniques des cartes IBM à 80 colonnes. Plus anecdotique peut-être, pour Klapisch-Zuber et Herlihy comme pour les autres utilisateurs de cette « édition », le choix de limiter à cinq caractères la variable « biens imposables déductions faites » fait plafonner la fortune de l'un des contribuables à 99 999 florins, au lieu des 103 306 florins qu'indiquent la source¹⁶⁵.

Il ressort de ces éléments d'analyse que la collecte des données et l'élaboration du code combinent des opérations déjà bien connues par les historiens au milieu des années 1960 : identification des sources, critique externe et interne et application de la méthodologie statistique. Le choix d'utiliser des ordinateurs, pèse ici essentiellement sur l'organisation du travail et sur la nature des données collectées. Pendant cette première phase de l'enquête, l'utilisation des ordinateurs fait cependant évoluer cette méthodologie statistique de deux façons distinctes.

Premièrement, les procédés de collecte des données sont fortement affectés par l'utilisation des cartes perforées IBM à 80 colonnes. Ce support de stockage des données conditionne le nombre de variables auxquelles vont s'intéresser Klapisch-Zuber et Herlihy. Il suppose encore l'élaboration d'une forme originale de division du travail, adossée à des procédures de collecte, de codage et de vérification des données extraites de la source. Ces procédures reposent sur des documents de travail (listes des quartiers, bordereau de perforation, manuel de codage), qui circulent entre les acteurs dans le but de constituer un ensemble de données conforme aux informations du *catasto* et exploitable statistiquement. Ces différents éléments font des cartes perforées IBM à 80 colonnes un agent de l'évolution de la formalisation des méthodes dans ce contexte.

Deuxièmement, la décision d'utiliser les ordinateurs entraîne un ensemble de conséquences que nous qualifierons d'indirectes. La constitution de collectifs de travail d'une certaine forme, décrite dans la première partie de ce chapitre, ou le choix d'étudier

¹⁶⁴ Ce texte est accessible à l'adresse suivante : http://cds.library.brown.edu/projects/catasto/newsearch/note_on_herlihy_files.html

¹⁶⁵ Ce contribuable n'est autre que Palla di Nofri Strozzi, leader de l'opposition à la famille Medicis au début du *Quattrocento*.

une source aux dimensions monumentales, sont corrélés à la décision d'utiliser des ordinateurs pour mener l'enquête. Un autre élément ressort cependant de l'analyse de l'élaboration du code : les apports de la méthodologie démographique dans la conduite de l'enquête. Le rapport à la source se construit en effet, dans le cas que nous étudions, dans une tension entre deux types de rationalités disciplinaires.

D'une part, Herlihy et Klapisch-Zuber proposent une enquête sur le modèle de l'histoire sociale quantitative, dont la finalité repose sur l'analyse statistique des observations données par le *catasto*. À ce titre, le choix d'enregistrer les erreurs de la source rend compte d'une approche fondamentalement historique de l'étude. D'autre part, Henry et Dupâquier tentent d'imposer une approche de la réalité démographique de la Toscane du XV^e siècle qui implique un traitement différent des sources. La confrontation avec un modèle d'enquête quantitative informatisée préexistant conduit dans notre cas les deux médiévistes à affirmer la spécificité de leur discipline et à inventer une forme de collecte des données propre aux principes de la méthode historique.

De cette première phase de l'enquête ressort, après environ cinq années de travail, une base de données historiques stockée sur des bandes magnétiques, contenant près de 1 000 000 d'informations sur les foyers et de plus de 1 300 000 sur les individus¹⁶⁶, réparties en deux fichiers couplés par l'intermédiaire du nom du chef de famille. Cette édition du *catasto* ne permet cependant de saisir la réalité de la société florentine en 1427 autrement qu'au prisme de la source qu'est le *catasto*. Cette distinction est fondamentale pour comprendre la finesse de l'analyse qui sera développée par les auteurs dans leur livre. Elle est encore fondamentale pour comprendre comment les historiens s'approprient la méthodologie statistique et informatique dans ce contexte.

Avant de s'intéresser aux réutilisations de ce nouveau format d'édition par d'autres acteurs, regardons d'abord ce que Klapisch-Zuber, Herlihy et leurs équipes respectives en ont fait.

¹⁶⁶ Ces chiffres sont repris de Christiane Klapisch-Zuber, « Le *catasto* florentin et son environnement », *Les Cahiers du Centre de Recherches Historiques. Archives*, n° 14-15 (14 octobre 1995).

3. Raisonnements, programmes et documents de sortie.

L'étude du *catasto* s'inscrit, à bien des égards, dans deux chemins déjà tracés par les grandes enquêtes historiques et démographiques lancées en France depuis la fin des années 1950. Tout d'abord, nous l'avons dit, elle s'inscrit dans la continuité d'un programme historiographique que les acteurs qualifient volontiers d'histoire sociale, et qui a pour objet l'étude de toutes les strates sociales et économiques des sociétés du passé. Ensuite, elle met en œuvre des méthodes statistiques destinées à analyser une source issue du rapport entre l'administration et ses contribuables, dans une perspective monographique locale. Dans les années 1950 et 1960, la combinaison de ces deux éléments fonde un modèle d'histoire sociale quantitative, qui est alors l'une des formes paradigmatiques de la recherche en histoire moderne, au moins dans certaines grandes institutions parisiennes comme le CRH et la Sorbonne.

Parmi les figures incontournables de ce modèle, Camille-Ernest Labrousse (1895-1988) est certainement l'un des plus importants¹⁶⁷. Spécialiste de l'histoire moderne de la France, il est en 1966 non seulement professeur honoraire à la Sorbonne, mais aussi directeur d'étude à la VI^e section de l'EPHE. Il dirige alors officiellement et officieusement de nombreuses thèses qui prennent la forme de monographies régionales, ainsi que les premiers travaux de ce type utilisant des moyens mécanographiques ou informatiques¹⁶⁸. Pierre Chaunu aurait même affirmé au début des années soixante-dix : « *Toute l'école historique française est labrousienne* »¹⁶⁹. Toute non, puisque les médiévistes demeurent, jusqu'à l'enquête du *catasto*, en dehors du giron de ce modèle.

L'enquête du *catasto*, qui est en France la première grande enquête quantitative assistée par ordinateur en histoire médiévale, est profondément marquée par ce modèle

¹⁶⁷ Voir par exemple Jean-Yves Grenier et Bernard Lepetit, « L'expérience historique. À propos de C.-E. Labrousse », *Annales* 44, n° 6 (1989) : 1337-60. Pour une étude plus approfondie sur Labrousse voir Maria Novella Borghetti, *L'œuvre d'Ernest Labrousse : genèse d'un modèle d'histoire économique*, Recherches d'histoire et de sciences sociales 106 (Paris: Éd. de l'École des Hautes Études en Sciences Sociales, 2005).

¹⁶⁸ A titre d'exemple, on peut ainsi citer le travail de Pierre Vilar sur la Catalogne à l'époque moderne (1962), celui d'Emmanuel Le Roy Ladurie sur les paysans du Languedoc (1966), ou encore, plus tardivement, et ayant pour ambition le recours aux méthodes informatiques, la thèse de Jean-Claude Perrot sur Caen au XVIII^e siècle. Labrousse et Braudel dirigeront, dans la deuxième moitié des années, 1970 la publication d'une synthèse en plusieurs tomes construite sur la base de ces monographies : Fernand Braudel et Ernest Labrousse, éd., *Histoire économique et sociale de la France* (Paris, France: Quadrige : Presses universitaires de France, 1977).

¹⁶⁹ Grenier et Lepetit, « L'expérience historique. À propos de C.-E. Labrousse », 1138.

d'études historiques. Trois principaux éléments viennent toutefois en assurer l'originalité : la collaboration avec une équipe américaine, le fait d'étudier une source médiévale et le recours aux ordinateurs. L'objet de cette troisième partie est justement de caractériser la particularité des raisonnements qui se développent dans l'étude du *catasto* par rapport au dernier de ces trois éléments. Il faudra cependant prendre en compte les deux premiers pour rendre justice à ce dernier.

Quelles sont donc, si elles existent, les particularités de cette étude par rapport au modèle d'études historiques présenté ci-dessus ? Les capacités de calcul des ordinateurs, mais aussi l'organisation du travail qu'ils supposent, jouent-elles sur les raisonnements mis en œuvre par les historiens dans ce contexte ? Et si oui, comment saisir ces différences, ces écarts, avec les travaux des historiens quantitativistes en histoire moderne ?

Pour traiter ces questions, nous nous appuyons ici sur une documentation plurielle : 1) la correspondance, déjà mentionnée, et plus particulièrement les échanges qui portent sur l'utilisation de deux programmes (MARSTA et HHLD) ; 2) les documents de travail retrouvés dans les archives du CRH ; 3) les figures reproduites dans la publication de l'étude (graphiques, tableaux et cartes) et 4) des recensions portant sur cette publication et 5) des articles publiés par les auteurs entre 1966 et 1978. Ces matériaux permettront de saisir, en contexte, les raisonnements des acteurs, autant que la fonction des ordinateurs dans ces raisonnements.

3.1. Raison statistique : quelles innovations et comment les saisir ?

L'étude du *catasto* aboutit, en 1978, à la publication d'un livre en français de 702 pages, découpé en cinq parties : *Les Toscans et leurs familles. Une étude du catasto florentin de 1427*¹⁷⁰. Les deux premières parties sont consacrées à la présentation critique du document et du contexte dans lequel apparaît cette loi fiscale. Les raisonnements qui y sont développés par les auteurs relèvent de méthodologies sur lesquelles nous ne nous

¹⁷⁰ Herlihy fait paraître une version abrégée du livre, en anglais sous le même titre en 1985 : David Herlihy et Christiane Klapisch-Zuber, *Tuscans and Their Families: A Study of the Florentine Catasto of 1427*, Yale Series in Economic History (New Haven London: Yale Univ. Press, 1985). Une traduction italienne de l'étude en français est par ailleurs publiée en 1988 : David Herlihy, Christiane Klapisch-Zuber, et Mario Bensi, *I toscani e le loro famiglie : uno studio sul catasto fiorentino del 1427* (Bologna, Italie: Mulino, 1988).

attarderons pas, puisqu'elles sont indépendantes de l'utilisation des ordinateurs. Les trois parties restantes, consacrées respectivement à la population et aux fortunes, aux comportements démographiques et aux images de la famille, constituent le cœur du travail. Les raisonnements développés dans ces parties nous donnent la matière pour mieux comprendre les particularités de cette étude au regard de travaux antérieurs, réalisés sans ordinateurs.

3.1.1. Puissance de calcul et multiplication des opérations statistiques.

Tout d'abord, il faut se demander quels types d'opérations permettent les programmes informatiques mobilisés par les auteurs. En l'absence d'autres documents, la correspondance constitue la source privilégiée pour éclairer ce que les acteurs retirent du fonctionnement de ces programmes. Une lettre de Herlihy à Klapisch-Zuber du 3 novembre 1971 permet de saisir certaines des opérations de l'un d'entre eux, intitulé HHLD. On y apprend que le programme résume (« *summarize* »), donne des pourcentages (« *gives percentage* »), montre (« *it also shows* ») ou encore classe (« *it ranks* »). Une autre lettre nous apprend qu'un autre programme, nommé MARSTA, propose aussi des visualisations des données (« *it draws an elegant graph* », « *it also include scattergrams* ») et calcule (« *calculate* ») des coefficients de corrélations pour l'ensemble de la population.

Sur la base de ces différents types d'opérations mentionnées par les acteurs, nous chercherons à caractériser les innovations qu'apporte l'utilisation des ordinateurs dans ce contexte. La méthode générale consistera à isoler un raisonnement, à le situer par rapport à d'autres raisonnements développés en histoire médiévale ou dans d'autres sciences sociales, et à comprendre ainsi comment l'utilisation des ordinateurs vient altérer, ou modifier, l'enchaînement des opérations intellectuelles dont ce raisonnement procède.

Il se présente cependant un problème fondamental pour mener à bien cette démarche : aucun raisonnement ne peut être caractérisé seulement par ses liens avec l'utilisation des ordinateurs. Prenons un premier exemple. En parcourant le livre de Klapisch-Zuber et Herlihy, on remarque que de nombreuses figures des *Toscans et leurs familles* sont des tableaux de contingence, résultats d'opérations de tris croisés. Ce type d'opérations, qui consiste à trier les individus sur la base de deux variables, compte parmi

les plus courantes qui soient en histoire quantitative dans les années 1960. Leur utilisation en histoire ne saurait donc être à elle seule caractériser par l'utilisation des ordinateurs, puisque la plupart des historiens mettent en œuvre ce type d'opération statistique sans pour autant avoir recours à des moyens informatiques depuis, *a minima*, les années 1930.

Les procédures statistiques les plus complexes, comme les analyses multivariées, préexistent, elles aussi à l'apparition des ordinateurs et peuvent être menées sans leur aide. Dès lors, qu'est-ce que l'utilisation de ce nouvel instrument apporte à ce type de raisonnement dans ce contexte ? Ou, en d'autres termes, est-il possible, de trouver dans la publication de l'enquête du *catasto* des raisonnements pour lesquels l'utilisation de l'ordinateur a manifestement joué un rôle prépondérant sur les résultats historiographiques ?

La réponse n'est pas aisée. Pour traiter ces questions, il faut apprécier les possibilités qu'offre l'utilisation des ordinateurs, les limites qu'ils imposent et les effets de ces possibilités limitées sur les raisonnements. Nous formulerons ici une hypothèse, qu'il s'agira de confronter aux raisonnements développés par les acteurs dans la publication de leur étude : l'emploi des ordinateurs va de pair avec une augmentation du nombre d'opérations statistiques et du volume de données traitées. Pour étayer cette hypothèse, il faut insister sur trois apports de l'ordinateur mentionnés par les auteurs dans leur livre de 1978.

Pour en comprendre la nature, il faut commencer par insister sur le fait qu'ils concernent essentiellement la possibilité pour les deux médiévistes de traiter l'ensemble des 60 000 déclarations du *catasto*, possibilité elle-même corrélée au caractère collectif de l'enquête. Klapisch-Zuber et Herlihy l'exprime en ces termes dans le troisième chapitre du livre :

« Les additions toutes bêtes portant sur une pareille masse de données, le croisement et la recherche des rapports entre les diverses caractéristiques échappent vite à l'effort isolé. Les limites physiques du chercheur qui devrait dépouiller des centaines de volumes ne sont pas seules en cause ; mais la probabilité d'erreur, que laisse prévoir la manipulation de tous ces chiffres, et la

*difficulté de certains calculs deviennent autant d'obstacles infranchissables sans l'aide d'un outil plus puissant, l'ordinateur. »*¹⁷¹

Cette citation nous révèle, dans les termes des acteurs, deux apports de l'utilisation des ordinateurs dans ce contexte : la réduction du risque d'erreur dans les opérations statistiques et la possibilité de mettre en œuvre certains calculs trop difficiles.

Un autre apport doit encore être mis en lumière pour mieux comprendre l'impact de l'utilisation des ordinateurs sur les raisonnements de l'enquête du *catasto*. Il concerne le fait que la multiplication des opérations d'un même type se trouve grandement facilitée par le recours à cet instrument de calcul. S'il n'existe malheureusement pas de données, à notre connaissance, permettant de comparer le temps que prend la mise en œuvre d'une série de tris croisés avec et sans ordinateur, il est évident pour tout praticien des statistiques que l'ordinateur permet de multiplier, à un rythme plus soutenu, ce type d'opérations, une fois la base de données et le programme développés¹⁷².

3.1.2. Conséquences de la multiplication des opérations statistiques.

À défaut de données, un premier indice des conséquences de ce troisième apport peut cependant être mis en lumière. Dans les deuxième et troisième parties du livre, Klapisch-Zuber et Herlihy, développent par exemple une approche systématiquement statistique qui tranche avec d'autres études réalisées à partir de la même source. Dans son étude sur la ville de Pistoia, publiée en 1967¹⁷³, Herlihy s'appuie lui aussi sur une méthodologie statistique pour étudier une partie du *catasto* de 1427, mais sans avoir recours à l'automatisation des calculs. Dans ce livre de 297 pages, on trouve 35 tableaux et 5 graphiques. Dans *Les Toscans et leurs familles*, livre de 702 pages, on ne retrouve pas moins de 194 tableaux, diagrammes et cartes. À défaut d'un indicateur plus intelligent, on constate que la densité de figures (tableaux, graphiques et cartes) est plus importante dans l'étude assistée par ordinateurs.

¹⁷¹ Herlihy et Klapisch-Zuber, *Les Toscans et leurs familles*, 101.

¹⁷² On pourra notamment regarder, pour un aperçu contemporain dans un autre contexte : Pierre Dagnelie, « L'ordinateur et l'enseignement de la statistique », *Revue de Statistique Appliquée* 19, n° 1 (1971): 5-15.

¹⁷³ Herlihy, *Medieval and Renaissance Pistoia*.

Une autre approche permet encore de saisir cette différence plus finement. Dans son chapitre 7, dédié à l'étude de l'économie urbaine, Herlihy publie un tableau intitulé « Merchant, Artisans and Shops at Pistoia, 1427 ». Il a pour objet d'éclairer « *the relative importance of these commercial and industrial activities at Pistoia* » en s'appuyant sur « *the number of people involved in the various trades, and the number of shops mentioned* »¹⁷⁴. Les neuf catégories qu'il utilise sont empruntées à la liste des guildes établie en 1330 à Pistoia.

Figure 6. Tableau des professions publié dans *Medieval and Renaissance Pistoia* (1967).

TABLE 26 Merchants, Artisans and Shops at Pistoia, 1427

	Persons	Shops
(1) Notaries	24	
(2) Dealers in spices, gold, silks, girdles and money changers	24	12
(3) Cloth merchants, cloth workers, and porters	43	16
(4) Butchers, leather workers and furriers	25	9
(5) Iron workers	51	10
(6) Stone masons and wood workers	24	1
(7) Shoemakers, painters and barbers	64	10
(8) Vintners, innkeepers and food salesmen	13	4
Unspecified laborers	22	

The table shows the number of persons involved in the various trades, and the number of shops mentioned, in the Catasto of 1427.

Source : Herlihy, *Medieval and Renaissance Pistoia : The Social History of an Italian Town, 1200-1430*, 175.

Dans le commentaire du tableau, Herlihy éclaire les modalités de la prépondérance de certaines industries sur d'autres. Pour la production du fer par exemple (5), il s'appuie sur des exemples d'investissement de riches familles florentines, découverts dans des sources primaires du XV^e siècle comme l'*Estimario della Citta* de Pistoia de 1415.

Dans les *Toscans et leurs familles*, Klapisch-Zuber et Herlihy abordent le même sujet dans leurs chapitre X sur « Les activités économiques ». Cette fois, ce sont trois tableaux et six cartes qui viennent alimenter l'analyse des auteurs. L'un de ces tableaux reprend la structure du précédent, mais en indiquant les dix métiers les plus représentés

¹⁷⁴ Herlihy et Klapisch-Zuber, *Les Toscans et leurs familles*, 174.

des sept plus grandes villes de Toscane. Le saut quantitatif entre les deux études, du point de vue de la masse de données prise en compte, ainsi que du nombre d'analyses statistiques mises en œuvre, est ici particulièrement frappant.

Figure 7. Tableau des professions publié dans *Les Toscans et leurs familles* (1978).

Les métiers les plus représentés à Florence et dans les grandes villes de Toscane

	<i>Florence</i>	<i>Pise</i>	<i>Pistoia</i>	<i>Arezzo</i>	<i>Prato</i>	<i>Volterra</i>	<i>Cortona</i>
1.	Notaire 307 (8,3)	Cordonnier 76 (10,4)	Cordonnier 50 (11,3)	Cordonnier 54 (13,1)	Paysan p. 43 (12,2)	Paysan t. 116 (34,7)	Paysan p. 45 (13,8)
2.	Administr. 261 (7,0)	Notaire 62 (8,5)	Paysan p. 48 (10,9)	Lainier 39 (9,5)	Lainier 36 (10,2)	Paysan p. 82 (24,6)	Lainier 26 (8,0)
3.	Tisserand 261 (7,0)	Paysan p. 44 (6,0)	Notaire 47 (10,7)	Notaire 29 (7,0)	Métayer 34 (9,7)	Lainier 27 (8,1)	Cordonnier 22 (6,8)
4.	Cordonnier 244 (6,6)	Tanneur 41 (5,6)	Forgeron 32 (7,3)	Forgeron 21 (5,1)	Cordonnier 33 (9,4)	Notaire 25 (7,5)	Notaire 17 (5,2)
5.	Lainier 222 (6,0)	Paysan t. 39 (5,3)	Lainier 21 (4,8)	Pourpoint. 17 (4,1)	Notaire 19 (5,4)	Métayer 21 (6,3)	Epicier 14 (4,3)
6.	Peigneur 202 (5,4)	Lainier 33 (4,5)	Epicier 17 (3,9)	Manœuvre 16 (3,9)	Forgeron 18 (5,1)	Administr. 10 (3,0)	Boucher 11 (3,4)
7.	Cardeur 188 (5,1)	Epicier 31 (4,2)	Barbier 13 (3,0)	Epicier 12 (2,9)	Meunier 15 (4,3)	Mercier 6 (1,8)	Forgeron 11 (3,4)
8.	Charpentier 169 (4,3)	Barbier 31 (4,2)	Charpentier 13 (3,0)	Charpentier 11 (2,7)	Epicier 13 (3,7)	Chaufourn. 6 (1,8)	Religieux 10 (3,1)
9.	Toilier 149 (4,0)	Courtier 25 (3,4)	Manœuvre 11 (2,5)	Charcutier 11 (2,7)	Charpentier 13 (2,7)	Charpentier 5 (1,5)	Charcutier 8 (2,5)
10.	Teinturier 117 (3,1)	Orfèvre 23 (3,1)	Tonnelier 10 (2,3)	Boucher 11 (2,7)	Paysan t. 10 (2,8)	Meunier 5 (1,5)	Bâtier 6 (1,9)

Le premier chiffre dans chaque colonne indique le nombre des membres de la profession chefs de ménage ; le nombre entre parenthèses donne le taux pour 1 000 habitants.

Administration : *vigilaio, piffero, meziere, corriere*
 Barbier : *barbiere*
 Bâtier : *bastiere, brigliaio, sellaio, bastaio*
 Boucher : *beccaio, pollaiolo, carnaiolo, lardaiolo*
 Cardeur : *scardassiere*
 Charcutier : *pizzicagnolo, cacciaino, treccone*
 Charpentier : *legnaiolo, falegname, tavolaccio*
 Chauffournier : *fornaciaio, stufaiolo, occhialaio*
 Cordonnier : *calzolaio, ciabattino*
 Courtier : *sensale*
 Epicier : *speziale*
 Forgeron : *fabbro*
 Lainier : *lanaiolo, ritagliatore*
 Manœuvre : *lavoratore*
 Mercier : *merciaio*
 Métayer : *mezzadro*
 Meunier : *mugnaio*
 Notaire : *notaio, giudice*
 Orfèvre : *orafo, gioielliere, orpellaio*
 Paysan p. : *paysan propriétaire*
 Paysan t. : *paysan tenancier*
 Peigneur : *carditore, pettinatore*
 Pourpointier : *farsettaio*
 Tanneur : *cuoiaio, cerbolattaio*
 Teinturier : *tintore*
 Tisserand : *tessitore di lana*
 Toilier : *linaiolo, rigattiere*
 Tonnelier : *bottaio, barattolaio*

Source : Herlihy et Klapisch-Zuber, *Les Toscans et leurs familles*, 35.

Par ailleurs, trois autres points de divergences apparaissent à la comparaison entre les deux tableaux (figure 6 et figure 7).

Le premier concerne les catégories utilisées. Contrairement au tableau de 1967, Klapisch-Zuber et Herlihy n'utilisent non pas les catégories des acteurs mais des catégories qu'ils ont eux-mêmes constituées. Pour donner au lecteur les clefs de cette opération intellectuelle, ils précisent dans la légende du tableau les termes de la source inclus dans chaque catégorie. Par contraste, les catégories du premier tableau réalisé par Herlihy sont beaucoup plus floues, puisqu'il ne précise nulle part dans son livre quels sont les termes des professions qui correspondent à ces neuf catégories. Pour la ville de Pistoia, les dénombrements qu'il effectue donnent d'ailleurs lieu à des résultats différents de ceux effectués au cours de l'enquête qu'il mène avec Klapisch-Zuber, comme pour la profession « notaire », pour laquelle le premier tableau fait état de 24 individus, contre 47 pour le second.

Cette remarque nous permet de saisir le second point de divergence entre les deux tableaux : la nature des données qui le constituent. Contrairement au tableau publié par Herlihy dans son livre sur Pistoia (figure 6), le tableau réalisé dans le cadre de l'enquête sur l'ensemble des documents du *catasto* (figure 7) présente les résultats de dénombrements pour sept villes de Toscane, contre une seule pour celui d'Herlihy. La quantité de données traitées est donc tout infiniment plus grande dans le cadre de l'étude réalisée à l'aide d'un ordinateur. De plus, par contraste avec la première étude de Herlihy, le résultat du dénombrement produits dans le cadre de l'enquête du *catasto* en précise explicitement la méthode : les médiévistes prennent ici en compte les professions des chefs de familles, élément qui ce qui n'est pas clairement indiqué dans la figure 6. Comme précédemment, la définition de la nature de ces dénombrements est donc beaucoup plus claire dans l'étude publié en 1978 que dans l'étude menée par Herlihy en 1967.

Sans juger une fois pour toutes les erreurs potentiellement faites par Herlihy dans ses premiers dénombrements, il apparaît que le fait d'explicitier les modalités de la construction des catégories, et de préciser la nature des données employées, assurent les acteurs d'une forme d'exactitude beaucoup plus solide. Comme nous l'avons vu dans la seconde partie de ce chapitre, ces deux éléments sont fortement corrélés aux méthodes de préparation des données rendues nécessaires par l'utilisation des ordinateurs mais, si l'on ajoute à cela le fait que les dénombrements réalisés par ordinateur sont infiniment plus sûrs que ceux effectués à la main, il est permis d'affirmer que l'un des apports majeurs de l'informatique dans ce contexte est la fiabilité qu'il assure aux données.

Une troisième différence entre ces deux types d'études peut encore être dégagée à partir de cette comparaison. Elle concerne l'utilisation des données par les acteurs dans leurs raisonnements. En effet, au contraire du premier exemple, Klapisch-Zuber et Herlihy n'utilisent pas d'autres sources pour donner du sens à leurs résultats statistiques. Leur analyse repose exclusivement sur le tableau lui-même. Il dégage ainsi, à partir des différences entre les sept villes choisies, une image de la « *spécialisation régionale* »¹⁷⁵ des activités économiques. Cette comparaison est rendue possible non seulement par le fait de traiter la source exhaustivement, mais aussi par le calcul d'un indice statistique qui n'apparaît pas dans le premier tableau : le taux de chaque profession pour mil habitants. Une fois de plus, si ce taux aurait pu être calculé sans ordinateur, il semble possible d'affirmer que l'ordinateur facilite grandement la réalisation d'une telle série d'opérations pour les soixante-dix professions présentes dans le tableau. En acceptant cette dernière affirmation comme vraie, on peut alors certifier que le recours aux ordinateurs joue aussi sur les raisonnements développés par les acteurs.

La comparaison entre ces deux tableaux met en lumière le caractère systématiquement statistique de l'étude mentionnée plus haut. Il nous apparaît que la multiplication des opérations, combinée à la faible probabilité de faire des erreurs de comptages ou de calculs, permet le développement d'une analyse plus complète, mais aussi plus exacte de la source. Ces deux aspects seront mis en valeur dans les recensions qui portent sur le livre, comme nous le verrons dans la quatrième partie de notre chapitre. Mais derrière ces éléments basés sur des publications, quelle est la réalité du rythme du travail ? La possibilité de multiplier les calculs et les opérations est-elle si évidente que cela, dans un contexte, nous l'avons vu, dans lequel les contraintes matérielles sont extrêmement prégnantes ?

Un exemple illustre parfaitement l'envers du décor, c'est-à-dire, la réalité matérielle du contexte de cette recherche. Dans une lettre du 3 novembre 1971 adressée par Herlihy à Klapisch-Zuber, le médiéviste affirme :

¹⁷⁵ Herlihy et Klapisch-Zuber, 296.

« I have the suspicion that wealth becomes more of a factor in fertility as women age (i.e. poor women cease having babies sooner than richer women), and this part of MARSTA was meant to investigate that possibility ».

À cette lettre est adjointe une nouvelle partie du programme MARSTA, dont Herlihy parle dans cette citation. À travers cet indice, Herlihy laisse transparaître ici la réalité matérielle qui sous-tend les raisonnements mis en œuvre. À rebours de l'hypothèse de facilité de reproduction des opérations du même type présentée plus haut, certaines questions des acteurs entraînent en pratique la nécessité de concevoir de nouvelles parties de programmes, et donc, un nouveau temps de programmation. Cette facilité est donc à mesurer à l'aune de la lourdeur du processus de fabrication d'un programme informatique, mais également des modalités d'échanges des programmes, que nous avons découverts dans la première partie de ce chapitre. Ces deux facteurs nuancent donc cette première conclusion.

Cet exemple, basé sur le cas des opérations de calculs les plus courantes en histoire quantitative, laisse cependant de côté un autre aspect de l'apport des ordinateurs : le bond « qualitatif » qu'ils permettent d'effectuer dans l'analyse de certains phénomènes. Pour mieux saisir ces autres apports et leurs conséquences historiographiques, il faut s'intéresser à des raisonnements statistiques plus complexes, très peu répandus dans le champ de l'histoire médiévale avant la fin des années 1970, comme l'utilisation des analyses multi variées ou l'élaboration de matrices de corrélations.

3.2. Une nouvelle pratique de la critique des sources.

De nombreuses constructions statistiques plus complexes que des dénombrements entrent en jeu dans l'analyse du *catasto* pendant les huit années que durent l'exploitation de l'« édition » et la rédaction du livre. Certaines d'entre elles, comme l'analyse factorielle des correspondances, sont cependant absentes du livre, et nous les laisserons de ce fait de côté, malgré leurs usages avérés pendant la phase de recherche, et en particulier dans plusieurs articles publiés au fil de l'enquête par Klapisch-Zuber et

Demonet¹⁷⁶. Pour mieux saisir les apports des programmes informatiques sur les raisonnements de ce type, nous ne nous intéresserons ici qu'à un seul chapitre du livre, consacré aux mariages.

Ce choix est motivé par le fait que ses conclusions ont été abondamment commentées par les démographes comme par les historiens, et que cet exemple pourra être repris dans la partie suivante pour éclairer les conséquences historiographiques de cette étude. Nous laisserons ainsi de côté d'autres types de raisonnements, qui sont pourtant d'un grand intérêt pour comprendre les usages des ordinateurs en contexte, comme la construction d'indices permettant l'analyse des anomalies repérées dans la source, l'utilisation d'un test graphique pour étudier les phénomènes de seuil dans des opérations de catégorisation, ou encore la recherche d'une définition statistique des mots qui désignent des classes socio-économiques.

3.2.1. Trois méthodes pour l'évaluation des âges moyens au mariage.

Le chapitre XIV du livre, qui nous intéresse ici, pose la question suivante : le mariage florentin se conforme-t-il au modèle ouest-européen du mariage caractérisé par le démographe John Hajnal (1924-2008) pour les XVII^e et XVIII^e siècles ? Pour répondre à cette question, les auteurs reprennent deux critères retenus par Hajnal dans un article fondateur de 1965 : l'âge au premier mariage des hommes et des femmes et la proportion de personnes qui ne se marient jamais¹⁷⁷. Dans le cas de l'étude du *catasto*, trois méthodes permettraient théoriquement, à partir de la source, d'apprécier l'âge moyen au premier mariage. En raison des faiblesses de chacune d'entre elles, ce sera finalement la combinaison des trois qui permettra aux deux médiévistes de déterminer l'âge moyen au premier mariage dans la société toscane en 1430.

La première méthode repose sur le fait que, dans les déclarations des feux d'origine des femmes mariées, les chefs de familles biffent leurs noms si le mariage a eu lieu entre 1427 et 1430, année des dernières corrections apportées par les scribes

¹⁷⁶ Christiane Klapisch-Zuber et Michel Demonet, « A uno pane e uno vino : la famille rurale toscane au début du XVe siècle », *Annales* 27, n° 4 (1972): 873-901 ; Christiane Klapisch, « A Correspondence Analysis of a XVth Century Census: The Florentine *Catasto* of 1427. », *Journal of European Economic History* 4, n° 2 (1975): 415-28.

¹⁷⁷ John Hajnal, « European Marriage Patterns in Perspective. », in *Population in History*, éd. par D. V. Glass et D. E. C. Eversley, 1st Edition (Routledge, 1965), 44.

florentins. Mais cet indice ne saurait suffire à caractériser les âges au premier mariage pour deux raisons principales : l'absence d'information sur les hommes et l'impossibilité de saisir s'il s'agit d'un premier ou d'un second mariage.

La deuxième méthode s'appuie sur les déclarations faites par les chefs de familles dans les feux d'arrivée des femmes mariées. Ces documents mentionnent, quant à eux, l'entrée dans le feu des femmes nouvellement mariées, et, pour une grande majorité d'entre elles, les âges de leurs époux. Cependant, ces deux ensembles de données comportent des limites. Pour des raisons tenant aux règles d'imposition appliquées par le gouvernement florentin, elles concernent dans leur grande majorité des familles plutôt aisées, possédant plus de 400 florins.

Pour contourner ces limites, une troisième méthode, indirecte cette fois, est utilisée. Elle est empruntée à Hajnal, qui la présente dans un article de 1953¹⁷⁸. Elle consiste à estimer l'âge moyen au premier mariage à partir des proportions de célibataires, hommes et femmes, trouvées dans les groupes d'âges successifs d'un recensement. Pour éviter les ambiguïtés, les auteurs doivent prendre en compte les pratiques d'arrondissement des âges repérées dans un de leur chapitre précédent, et faire l'hypothèse que les hommes d'états matrimoniaux indéterminés sont célibataires. L'immense avantage de cette méthode repose sur le fait qu'elle permet d'éclairer les pratiques matrimoniales des classes pauvres et paysannes, dont les caractéristiques ne peuvent être étudiées à l'aide des deux autres approches.

3.2.2. Une construction de la critique des sources en dialogue avec les sciences sociales.

Pour présenter et rendre comparables les résultats de ces trois méthodes, Klapisch-Zuber et Herlihy proposent un tableau des « Âges observés et des âges estimés aux mariages »¹⁷⁹, qui montre que les valeurs données par les trois indicateurs concordent.

¹⁷⁸ John Hajnal, « Age at Marriage and Proportions Marrying », *Population Studies* 7, n° 2 (1953): 111-36.

¹⁷⁹ Herlihy et Klapisch-Zuber, *Les Toscans et leurs familles*, 399.

Figure 8. Tableau reprenant les résultats des trois méthodes d'évaluation de l'âge au premier mariage.

Âges observés et âges estimés au mariage*

	Villes			Campagnes		Total
	Florence	grandes	petites	Contado	District	
<i>1. Mariées biffées de la déclaration paternelle</i>						
Nombre.....	67	29	11	26	15	150
Age moyen.....	17,56	21,98	22,77	17,00	16,60	18,20
<i>2. Mariées ajoutées à la déclaration du mari</i>						
Mariées						
Nombre.....	55	16	26	245	26	369
Age moyen.....	20,83	18,25	19,04	19,26	19,42	19,17
Epoux						
Nombre.....	76	17	27	238	27	375
Age moyen.....	34,41	29,15	25,80	23,42	25,37	25,03
Deux âges connus...	51	14	26	228	26	345
Différence moyenne	13,63	10,29	6,88	6,63	7,04	7,89
<i>3. Age estimé à partir des % de célibataires</i>						
Femmes.....	17,96	17,92	17,93		18,36	
Maris.....	29,95	27,40	26,39		25,63	
Différence moyenne	11,99	9,48	8,46		7,27	

* Seules les personnes d'âge connu ont été retenues.

Source : Herlihy et Klapisch-Zuber, *Les Toscans et leurs familles*, 399.

Après une analyse complémentaire portant uniquement sur le célibat, et sur laquelle nous ne nous appesantirons pas, les auteurs concluent que le modèle matrimonial toscan se caractérise par :

« un âge encore tendre au premier mariage des femmes, un nombre presque négligeable de vieilles filles en dehors des couvents ; des hommes d'âge mûr à leur première union, une proportion plus forte de célibataires prolongés, puis définitifs ; et, conséquence de tout cela, une différence d'âge très accentuée entre mari et femme. Surtout dans les villes, le comportement masculin correspond assez bien au mariage tardif et aux tendances au célibat

du modèle européen défini par J. Hajnal ; mais les femmes s'en écartaient considérablement »¹⁸⁰.

L'analyse conduit donc au rejet de l'hypothèse formulée par Hajnal en 1965, qui le conduira, et nous le verrons, à rectifier sa théorie dans un article incontournable pour la démographie historique publié en 1982¹⁸¹.

Ce raisonnement appelle plusieurs remarques intéressantes pour ce qui est des questions qui nous préoccupent ici. La première concerne la pratique de l'observation historique développée sur la base de la source. La nature des informations disponibles dans les différentes déclarations des chefs de familles conduit les historiens à combiner plusieurs méthodes d'analyses pour parfaire la critique de ces informations. Ils développent ainsi une forme de critique des données de la source qui s'appuie sur la possibilité de multiplier les opérations statistiques. En effet, on peut supposer qu'en l'absence d'instrument de calcul telle que l'ordinateur, les médiévistes n'auraient eu recours qu'à l'une de ces trois méthodes.

Par ailleurs, la conclusion du raisonnement permet l'inscription de l'étude dans un dialogue avec les sciences sociales, que l'on retrouve tout au long du livre. Les auteurs empruntent en effet de nombreux indicateurs et de nombreuses méthodes à d'autres sciences sociales, comme l'indice de Gini à l'économie, ou l'analyse spatiale à la géographie. Mais quel rôle joue l'ordinateur dans ces emprunts ? La réponse n'est, là encore, pas facile à donner.

Il semble tout d'abord que le recours aux ordinateurs permet d'assurer une forme de précision et de rigueur dans l'analyse des données, en raison de la multiplication des mesures que cela rend possible. Les résultats de ces analyses sont dès lors comparables avec ceux d'autres sciences sociales, qui utilisent elles-mêmes des ordinateurs pour construire ce type d'indices (comme la démographie par exemple). Outre l'ordinateur, l'importation de ce type de procédures statistiques en histoire suppose par ailleurs la présence d'un statisticien de métier aux côtés des historiens. Ces derniers, qui sont issus pour la plupart de formations littéraires dans les années 1960, sont le plus souvent incapables de mettre en œuvre ce type de procédures. Certains spécialistes de la

¹⁸⁰ Herlihy et Klapisch-Zuber, 404.

¹⁸¹ John Hajnal, « Two Kinds of Preindustrial Household Formation System », *Population and Development Review* 8, n° 3 (1982): 449-94.

quantification y voient d'ailleurs la principale limite au développement des méthodes quantitatives en histoire¹⁸².

Dans ce contexte, le recours aux ordinateurs marque l'originalité de l'enquête du *catasto* dans le champ des études médiévales, non seulement par la multiplication et l'articulation de différentes méthodes statistiques pour la critique du document, mais aussi par l'emprunt d'outils statistiques à des disciplines elles-mêmes déjà informatisées. L'enquête produit ainsi, à la fin des années 1970, des données pour et à partir des sciences sociales. Pourtant, si l'approche démographique est fondatrice pour ce chapitre, l'analyse des auteurs ne s'arrête pas là. Les conclusions démographiques servent en effet de base pour une réflexion plus historique, qui passe notamment par l'articulation des résultats issues de l'étude du *catasto* avec d'autres sources.

3.3. Utilisation des coefficients de corrélations et conséquences historiographiques.

3.3.1. Étude de l'intensité des relations entre trois variables.

Dans la seconde partie du même chapitre, les auteurs utilisent leurs conclusions relatives au mariage pour aborder une autre problématique : celle des relations entre pratiques matrimoniales, sexes, lieux de résidences et fortunes. Ce type de question est, rappelons-le, au cœur du projet initial de l'enquête puisque les acteurs se proposaient dès 1966 d'étudier les corrélations entre variables démographiques et économiques. Pour explorer les relations entre ces variables de natures différentes, ils décident d'attribuer une valeur numérique aux modalités des deux variables catégorielles (sexe et résidence¹⁸³) et de calculer le coefficient des corrélations entre chaque paire de variables. La finalité de ces opérations est présentée ainsi par les auteurs dans leur chapitre :

¹⁸² C'est l'analyse que proposent certains acteurs dans plusieurs entretiens que nous avons pu mener entre 2016 et 2018 auprès d'historiens comme Antoine Prost, Françoise Waquet ou Alain Guerreau. Pour une vision plus argumentée de cet état de fait, dans le contexte américain, on pourra notamment consulter l'excellent article de Robert William Fogel, « The Limits of Quantitative Methods in History », *The American Historical Review* 80, n° 2 (1975): 329-50.

¹⁸³ Les valeurs ainsi définies sont les suivantes : le sexe masculin reçoit la valeur 1, le féminin la valeur 2. Florence reçoit la valeur 4, les dix grandes villes la valeur 3, les bourgades la valeur 2 et les campagnes la valeur 1.

« L'analyse des corrélations permet de préciser et de mesurer, au moins partiellement, la puissance de l'influence du milieu ou de la résidence et celle de la fortune sur le mariage »¹⁸⁴.

Pour comparer ces coefficients de corrélations, ils proposent une matrice de corrélations, qui représente les résultats pour les différentes corrélations qui unissent le sexe, la résidence, la fortune et la proportion de mariés, la proportion de veufs et la proportion de célibataires.

Figure 9. Matrice de corrélations.

Matrice de corrélations - Sexe, résidence, fortune et état matrimonial (1427)

	Sexe	Résidence	Fortune	Proportion de mariés *	Proportion de veufs *	Proportion de célibataires *
Sexe	1,00					
Résidence	0,00	1,00				
Fortune	0,00	0,00	1,00			
Proportion de mariés*	- 0,03	- 0,74	- 0,46	1,00		
Proportion de veufs*	0,94	0,19	- 0,06	- 0,19	1,00	
Proportion de célibataires*	- 0,90	0,18	0,25	- 0,30	- 0,88	1,00

* Sexes réunis

Source : Herlihy et Klapisch-Zuber, *Les Toscans et leurs familles*, 410.

Le tableau présente en lignes et en colonnes les six variables qui intéressent Klapisch-Zuber et Herlihy. Les valeurs indiquées dans chaque case, comprises entre -1 et 1, caractérisent l'intensité de la relation linéaire entre les paires de variables. Plus cette valeur est proche de 0, moins la corrélation entre les deux variables est forte. Les auteurs donnent par exemple la lecture suivante :

« Ces coefficients montrent donc que l'association entre sexe masculin et célibat est un peu moins étroite que celle entre sexe féminin et veuvage. Un coefficient

¹⁸⁴ Herlihy et Klapisch-Zuber, *Les Toscans et leurs familles*, 407.

négligé élevé (-0,88) exprime bien la variation inverse des proportions de veufs et de célibataires »¹⁸⁵.

Les autres coefficients les plus remarquables concernent les relations entre le sexe et la proportion de veufs (0,94), qui atteste que la grande majorité des veufs sont des femmes, ou entre sexe et proportion de célibataires (-0,90), qui montre à l'inverse que presque tous les célibataires sont des hommes.

Deux autres de ces coefficients sont incontournables pour répondre aux questions que se posent les deux médiévistes. Il s'agit de ceux calculés pour établir la dépendance entre résidence et proportion de mariés (-0,74), et entre fortune et proportion de mariés (-0,46). Ces deux valeurs, qui attestent d'une corrélation plus ou moins forte entre ces deux paires de variables, conduisent les auteurs à détailler les mécanismes à l'œuvre en s'appuyant notamment sur le calcul de nouveaux coefficients de corrélation, entre l'âge au premier mariage, la résidence et la fortune. Il découvre qu'en fonction des milieux, les pratiques matrimoniales sont différentes :

« Autrement dit, la fortune encourageait le paysan à se marier, alors qu'elle incitait les habitants des villes à retarder les unions. »¹⁸⁶.

Pour clore cette partie du chapitre, et conclure sur l'influence de ces deux variables sur le pourcentage de mariés en Toscane, Klapisch-Zuber et Herlihy proposent une analyse de corrélation multiple qui confirme statistiquement l'importance de la dépendance entre ces deux derniers facteurs et le pourcentage de mariés.

Cette dernière opération est particulièrement intéressante pour comprendre l'apport des ordinateurs dans ce raisonnement. Elle constitue certainement l'une des opérations qualifiées de difficiles par Klapisch-Zuber et Herlihy dans le passage cité au début du 3.2.. En effet, encore plus que pour le calcul de coefficients de corrélation qui se présente comme des indices d'observations des relations entre les variables, l'analyse de corrélation multiple est un véritable test statistique. Le résultat de ce test joue donc la

¹⁸⁵ Herlihy et Klapisch-Zuber, 410.

¹⁸⁶ Herlihy et Klapisch-Zuber, 411.

fonction d'une preuve dans le raisonnement de Klapisch-Zuber et Herlihy, puisqu'il affirme la significativité du lien entre ces variables.

3.3.2. Exploitation des résultats de la matrice de corrélations.

Une fois ces résultats établis, les auteurs se demandent quelles sont les implications sociales de ces caractéristiques matrimoniales. Ils avancent une explication de deux traits de la société florentine que nombre de sources mentionnent. Tout d'abord - et cet aspect intéressera, comme nous le verrons, de nombreux historiens - de nombreux hommes se retrouvent, en ville, privés « *d'exutoire sexuel* »¹⁸⁷ avant leurs 30 ans, ce qui explique, selon les auteurs, pourquoi Florence avait pu être reconnue par les contemporains comme une ville réputée pour la prostitution et les pratiques homosexuelles¹⁸⁸.

Cette analyse conduit par ailleurs Herlihy à y voir l'une des racines démographiques de la violence dans les villes toscanes au *Quattrocento*, connues pour leurs luttes de factions et les nombreuses guerres privées auxquelles participaient les jeunes hommes¹⁸⁹. Une autre conséquence de ces mariages tardifs est à rechercher du côté des familles riches. La forte mortalité qui touche l'ensemble de la population toscane au XV^e siècle entraîne une raréfaction des hommes en âge de se marier, et, donc, une augmentation du montant des dots. En conclusion, les auteurs croisent leurs résultats avec d'autres sources, comme des livres de familles, des testaments ou le *Monte delle doti*, qu'étudie un élève de Herlihy, Anthony Mohlo (1939-).

Dans quelle mesure les nouvelles possibilités offertes par l'ordinateur ont-elles joué un rôle dans de telles conclusions ? Deux réponses peuvent ici être formulées. La première concerne, comme dans l'exemple précédent, la possibilité qu'offre cet instrument de calcul de multiplier les tests de corrélation pour parfaire les raisonnements

¹⁸⁷ Herlihy et Klapisch-Zuber, 414.

¹⁸⁸ Voir par exemple David Herlihy, « Vieillir au Quattrocento », *Annales* 24, n° 6 (1969): 1338-52.

¹⁸⁹ Les auteurs affirment que ces violences permettaient « *de dépenser leur trop plein d'énergie* », in Herlihy et Klapisch-Zuber, *Les Toscans et leurs familles*, 414. Pour une analyse plus poussée et plus claire de ce phénomène, il faut cependant consulter David Herlihy, « Some Psychological and Social Roots of Violence in the Tuscan Cities », in *Violence and Civil Disorder in Italian Cities, 1200-1500*, éd. par Lauro Martines (University of California Press, 1972), 129-55. Cette problématique a par ailleurs été soulevé dès le début des années 1960, sur la base de sources d'une autre période et d'un autre lieu, dans Georges Duby, « Dans la France du Nord-Ouest au XII^e siècle : les « jeunes » dans la société aristocratique », *Annales* 19, n° 5 (1964): 835-46.

sur les relations entre les différentes variables. L'utilisation d'une analyse de corrélation multiple renforce l'idée que l'ordinateur joue un rôle primordial sur l'établissement de la nature de ces relations, au titre de preuve de la significativité des liaisons entre les variables. L'analyse des facteurs qui influencent l'âge au mariage s'en trouve fortement raffinée, et éclaire les sources traditionnellement mobilisées par d'autres historiens pour traiter du même sujet. Dans ce contexte, les ordinateurs s'avèrent être autant un agent de l'automatisation des méthodes de l'histoire quantitative, permettant l'analyse de données plus nombreuses, qu'un instrument permettant, par la multiplication des opérations et le recours à des constructions statistiques complexes, une amélioration qualitative de l'analyse comme de la synthèse.

La seconde réponse qui peut être donnée repose sur un type de figures jusqu'ici passé sous silence, et pourtant fondamental pour comprendre une autre dimension de l'apport qualitatif qui dérive de l'utilisation des méthodes informatiques : les visualisations produites à l'aide des programmes. Pour mieux comprendre l'apport de ces visualisations, nous nous intéresserons à l'une des principales originalités du livre : la présence de nombreuses cartes, produites par le laboratoire de cartographie automatique de l'EPHE. Les modalités de productions des cartes, comme leurs usages et leurs contextes de publication nous renseigneront sur leurs fonctions dans les raisonnements développés dans *Les Toscans et leurs familles*.

3.4. Cartographie automatique et raisonnements historiques.

Quels sont les apports de l'informatique dans le traitement et la représentation de l'information en sciences sociales ? Et comment les nouvelles visualisations rendues possibles par l'utilisation des ordinateurs permettent-elles de rendre compte d'un apport dans la recherche en histoire, dans le contexte qui nous intéresse ? Pour répondre à ces questions, il faut revenir sur l'une des figures de cette enquête : le cartographe en chef de l'étude, Jacques Bertin, fondateur et directeur du laboratoire de cartographie de la VI^e section de l'EPHE depuis 1957. Dans l'introduction de son célèbre traité de 1964, intitulé *Sémiologie graphique*, il affirme que la graphique, entendue comme l'équivalent des

mathématiques pour la perception visuelle¹⁹⁰, a subi un changement fondamental sous la pression des technologies informatiques :

*« Devenue manipulable par superpositions, juxtapositions, transformations, permutations, autorisant groupements et classements, l'image graphique est passée de l'image morte, de l'illustration, à l'image vivante, à l'instrument de recherche accessible à tous. »*¹⁹¹

Aux yeux de Bertin, l'utilisation des ordinateurs a fait muter la fonction des graphiques, qui n'opèrent plus exclusivement comme un vecteur de communication, mais aussi comme un instrument heuristique permettant le traitement des informations¹⁹². Pour saisir les implications de ces possibilités ouvertes par l'informatique dans le contexte de l'étude du *catasto*, nous nous intéresserons aux fonctions dévolues aux 67 cartes présentées dans *Les Toscans et leurs familles*.

3.4.1. La cartographie automatique « n'a rien à voir avec la géographie ».

Deux caractéristiques principales fondent la valeur des cartes pour notre propos. La première concerne leurs modes de production. Les cartes produites dans le cadre de l'étude résultent pour la très grande majorité de traitements statistiques des données fournies par l'« édition » informatique de la source. C'est à ce type de cartes que nous nous intéresserons particulièrement. Elles sont imprimées automatiquement à l'aide d'une « machine à imprimer des points »¹⁹³, et donne le type de résultats suivants :

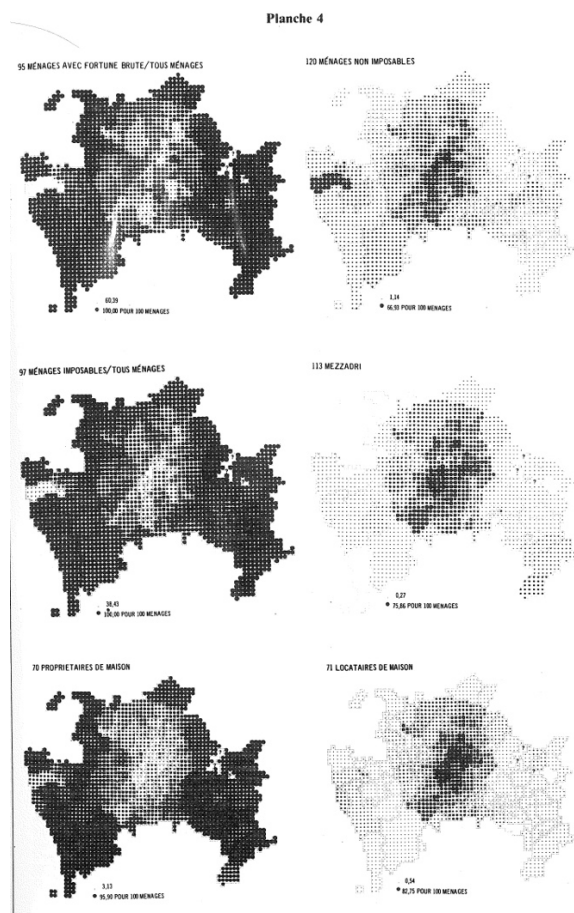
¹⁹⁰ L'édition utilisée ici est la suivante : Jacques Bertin et al., *Sémiologie graphique : les diagrammes, les réseaux, les cartes* (Paris, France: Éd. EHESS, 2013).

¹⁹¹ Jacques Bertin et al., 8.

¹⁹² Un autre exemple d'analyse de ce type d'outil heuristique est étudié dans le chapitre 6 de cette thèse.

¹⁹³ Cette formulation est reprise d'une lettre de Klapisch-Zuber adressée à Herlihy le 13 décembre 1971.

Figure 10. Exemple d'une planche de cartes, produites de façon automatique, dans le cadre de l'enquête du *catasto*.



Source : Herlihy et Klapisch-Zuber, *Les Toscans et leurs familles*, 243.

Du fait de ces modalités de production, ces cartes ont le même statut que le résultat d'une opération statistique comme un dénombrement ou un calcul de coefficient de relation. Ces cartes figurent les résultats du traitement des données, et se présente donc comme un outil heuristique pour les acteurs. En ce sens, ce sont de formidables indices pour saisir les raisonnements mis en œuvre dans le cadre de l'enquête. Herlihy présente cette pratique cartographique particulière en ces termes dans une lettre du 12 août 1971 :

« mapping has nothing to do with geography, it is computer language [...] a process which fits together all parts of a program ».

L'autre caractéristique essentielle de ces objets pour comprendre leur valeur dans les raisonnements des acteurs concerne l'efficacité que Bertin attribue à un certain type

de représentation graphique informative, par rapport à d'autres formes graphiques, comme les tableaux par exemple. Dans son traité, Bertin affirme :

« il faudrait au moins 20 000 instants successifs de perception pour comparer deux tableaux de chiffres de 100 lignes sur 100 colonnes »¹⁹⁴,

au contraire :

« Que les chiffres soient transcrits graphiquement et la comparaison est aisée, et peut même être instantanée [...] la perception visuelle dispose de trois variables sensibles : la variation des taches et les deux dimensions du plan »¹⁹⁵

et il conclut donc :

« dans un instant de perception, les systèmes linéaires ne nous communiquent qu'un seul son ou signe, tandis que les systèmes spatiaux [...] nous communiquent dans le même instant les relations entre les trois variables. »¹⁹⁶.

Les possibilités de traitement de l'information offertes par ces cartes, tout comme les analyses qu'elles rendent possibles pour les acteurs, s'en retrouvent donc fortement multipliées.

L'observateur qui voudrait étudier ce type de phénomènes dans notre contexte se retrouve cependant face à un obstacle, qui traverse de nombreuses recherches explorant la matérialité des raisonnements. Tout d'abord, la documentation conservée au CRH ne permet pas de travailler sur d'autres documents graphiques que ceux publiés dans le livre. Comment savoir, dès lors, si ces figures sont utilisées pour leur bonne représentation de l'information, c'est-à-dire, dans une visée argumentative, ou dans une visée heuristique, c'est-à-dire, pour la conduite des raisonnements ? Face à l'impossibilité pratique de répondre, nous considérerons ici que ces cartes assument toujours une double fonction : heuristique et argumentative. En nous intéressant à leur inscription dans l'argumentation,

¹⁹⁴ Jacques Bertin et al., *Sémiologie graphique : les diagrammes, les réseaux, les cartes*, 7.

¹⁹⁵ Jacques Bertin et al.

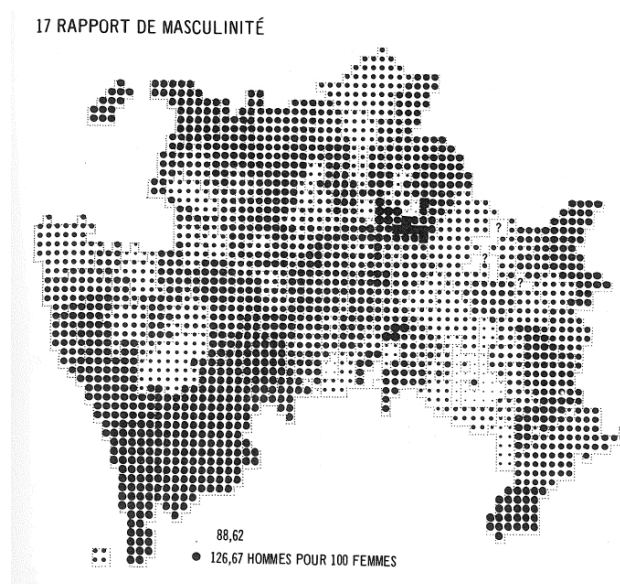
¹⁹⁶ Jacques Bertin et al.

nous essaierons d'identifier leur fonction dans le raisonnement, et de saisir, par-là, les apports de ces documents pour les auteurs.

3.4.2. Apport de la cartographie automatique dans l'étude de la Toscane médiévale.

Le premier raisonnement qui nous intéressera ici concerne l'analyse des variantes régionales du rapport de masculinité (proportion d'hommes pour cent femmes). Klapisch-Zuber et Herlihy remarquent en effet, après d'autres¹⁹⁷, que la Toscane florentine comprend bien plus d'hommes que de femmes. Après avoir examiné si ce déficit venait d'un sous-enregistrement des femmes, ou d'un ensemble de facteurs réduisant réellement leur effectif, les auteurs s'intéressent aux relations entre le rapport de masculinité et d'autres variables dont le lieu de résidence. Pour l'analyse de la relation entre cette paire de variables, les auteurs ont recours à une carte, publiée au tout début de la partie concernant cette question.

Figure 11. Carte représentant la distribution géographique du rapport de masculinité.



Source : Herlihy et Klapisch-Zuber, *Les Toscans et leurs familles*, 347.

¹⁹⁷ Cette remarque avait été formulé dès le début du XIV^e siècle par certains contemporains, comme Giovanni Villani, qui avait remarqué que le nombre de nouveau-nés masculins dépassait celui de sexe féminin, in Giovanni Villani, *Cronica di Giovanni Villani: a miglior lezione ridotta coll'ajuto de'testi a penna, con note filologiche di I. Moutier e con appendici storico-geografiche, compilate da Franc. Gherardi Dragomanni...* (Firenze, Italie: S. Coen, 1844), 184.

À propos de cette carte, ils affirment :

« La cartographie des variations rurales du rapport indique que les campagnes de Toscane ne sont pas homogènes. La carte 17 [figure 11] présente en effet une alternance de larges bandes diagonales, orientées du nord-est au sud-ouest, disposition en écharpe qui se retrouve sur de nombreuses autres représentations de données statistiques du catasto ».

Nous pouvons saisir ici deux principaux indices de la manière dont les cartes sont utilisées dans les raisonnements. Le premier concerne la fonction de la cartographie automatique : à la manière d'un tableau de contingence, ici effectué entre un indice (le rapport de masculinité), et une variable (le lieu de résidence), les auteurs se servent de la carte pour dégager la structure d'une liaison entre deux éléments. Mais contrairement à une représentation sous forme de tableau, qui aurait certainement permis de dégager cette même conclusion, la carte permet de découvrir la structure géographique de ces variations. La fonction de ce type de carte, prôné par Bertin dans son traité, apparaît ici en contexte.

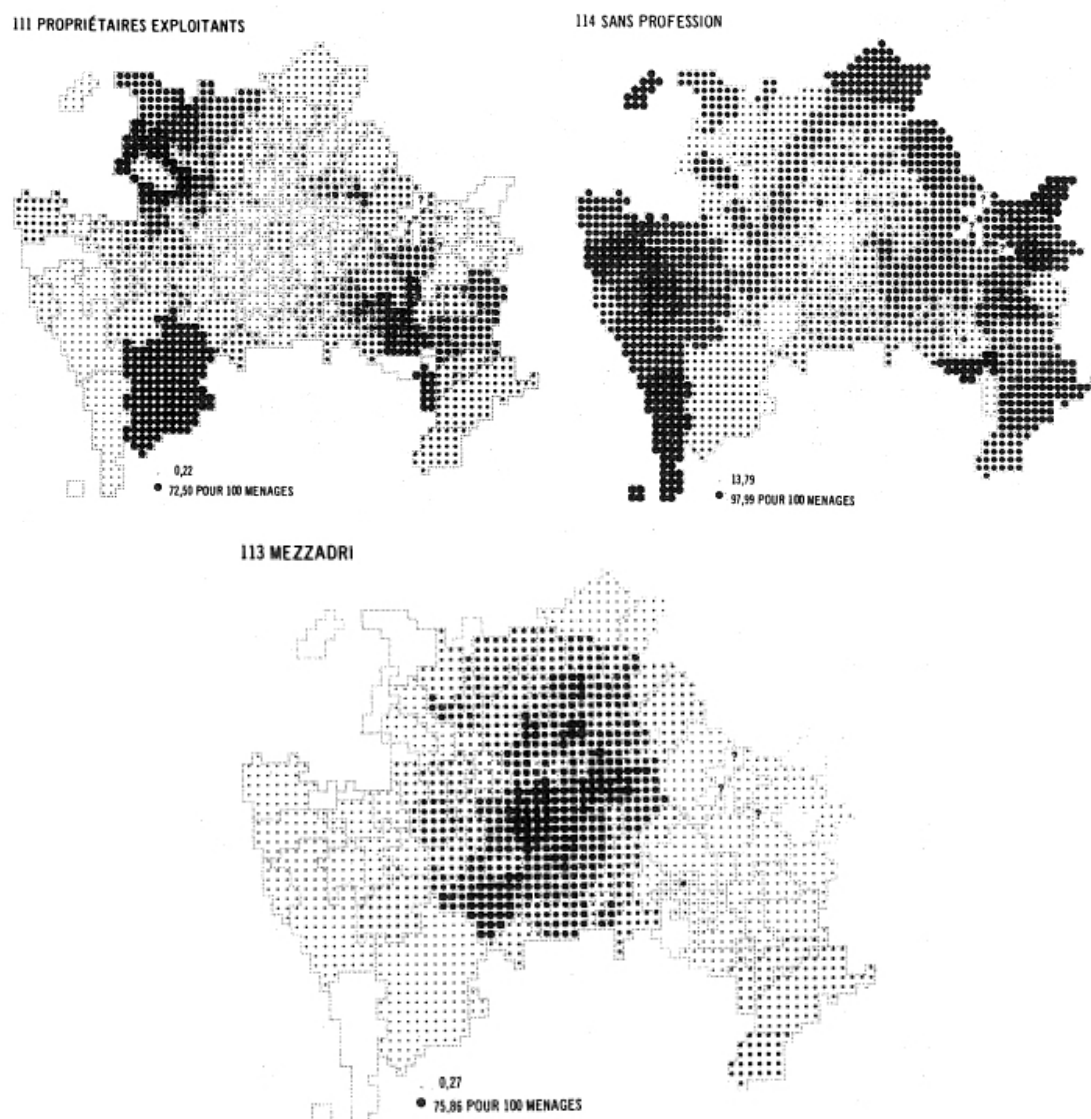
Des analyses de ce type, menées à l'aide de cartes du même genre se retrouvent tout au long du livre. Elles permettent d'identifier les distributions géographiques des données pour différentes variables portant sur les populations rurales de Toscane. Pour mieux comprendre comment la lecture combinée de plusieurs cartes permet de faire émerger de nouvelles questions historiques, nous nous attarderons sur un autre raisonnement, tiré du chapitre portant sur les activités économiques.

Au début de ce chapitre, les auteurs font le choix d'explorer en détail les activités de métayage (*mezzadria*), qu'ils considèrent comme un très bon cas d'étude pour saisir les rapports entre les caractères démographiques et les caractères économiques dans les campagnes. Pour aborder cette question, les auteurs s'intéressent notamment à la distribution géographique de trois types de catégories socio-économiques du monde rural : celle des *mezzadri*, celle des propriétaires exploitants et celle des « sans

profession », qui regroupent, du fait du codage mis en œuvre par les auteurs, certains propriétaires vivant de leurs terres¹⁹⁸.

Trois cartes, correspondant à chacune de ces situations, sont réalisées, et publiées dans le livre sur la même planche¹⁹⁹.

Figure 12. Trois cartes de la distribution géographique de professions.



Source : Herlihy et Klapisch-Zuber, *Les Toscans et leurs familles*, 285.

¹⁹⁸ Les raisons de ce codage tiennent aux modalités de déclarations de cette catégorie de paysans qui, en raison du montant de leurs revenus issus de la terre, n'ont pas à déclarer précisément leurs activités. Faute de meilleure solution, les auteurs les ont donc désignés comme sans profession. Voir Herlihy et Klapisch-Zuber, *Les Toscans et leurs familles*, 272.

¹⁹⁹ Cette planche est du type de celle présentée à la figure 10. Pour rendre la lecture plus simple, nous ne reproduisons ci-dessous que les trois cartes en question.

Les auteurs indiquent aux lecteurs comment les utiliser en ces termes :

*« les cartes 111 et 114 doivent être lues ensemble. Car s'il est malaisé d'identifier avec précision les mezzadri, il est encore plus difficile d'isoler les propriétaires exploitants. »*²⁰⁰

Ils précisent encore, à la page suivante :

*« Les deux cartes superposées restituent en revanche l'image inversée de la carte des mezzadri »*²⁰¹

L'utilisation heuristique de ces trois cartes par les médiévistes apparaît, à la lumière de ces citations, de deux façons distinctes. Tout d'abord, elles sont utilisées pour l'exploration de phénomènes de seuils entre des situations socio-économiques différentes, mais proches (*mezzadri*, propriétaires exploitants, sans profession). Ensuite, elles le sont comme outil de comparaison, à travers la superposition des cartes, opération suggérée, comme nous l'avons vu, par Jacques Bertin en tant que nouvelle pratique heuristique à l'ère de la cartographie automatique. Ces deux procédés permettent aux auteurs de faire émerger plus précisément la distribution géographique de ces trois situations difficiles à démêler autrement.

L'autre question qui anime les auteurs quant à la vie de ces métayers tient à la fortune attachée à leurs situations. En effet, l'historiographie des XIX^e et XX^e siècle avait vanté les mérites de l'harmonie sociale qui naissait du contrat de *mezzadria* à la fin du Moyen-Âge²⁰². Les auteurs entendent examiner cette question, là encore, à l'aide de cartes. Pour cela, ils proposent une autre comparaison entre trois cartes :

²⁰⁰ Herlihy et Klapisch-Zuber, 271.

²⁰¹ Herlihy et Klapisch-Zuber, 272.

²⁰² Herlihy et Klapisch-Zuber, 273.

*« Les cartes 95 et 97, comparées à la carte 113 représentant la proportion de métayers, mettent en évidence la relation entre leur présence, et la diminution des contribuables ayant du bien, et imposables. »*²⁰³

Cette fois, ce n'est plus pour affiner l'observation des catégories que ces cartes sont mobilisées, mais bien pour établir une « *corrélation régionale* »²⁰⁴ entre fortune et métayage, et ouvrir ainsi des questions historiques sur les conditions de vie des *mezzadri* au début du XV^e siècle. Après une analyse infructueuse portant sur les moyens alloués par les propriétaires aux exploitants (bœufs, bâtiments d'habitation et d'exploitation), les auteurs concluent finalement cette partie par une liste de questions, qui aboutit au constat des limites de leur matériau d'analyse :

*« Autant de questions que l'examen des données globales du catasto ne permet pas de trancher »*²⁰⁵.

Finalement, la cartographie automatique est devenue pour les auteurs, au fil de l'étude, l'un des outils heuristiques et argumentatifs majeurs. Un exemple nous renseigne cependant sur les difficultés matérielles que la mise en œuvre de ces méthodes a pu susciter. Dans une lettre adressée à Herlihy le 13 décembre 1971, Klapisch affirme que la cartographie automatique conduit à des « *retards exaspérants* », dus à des « *pannes quotidiennes du terminal de l'IBM* ». Elle ajoute, plus généralement :

« J'aurais vraiment servi de cobaye, comme on me l'a dit du reste, mais la situation de cobaye vaudra peut-être un peu plus d'originalité à l'entreprise ».

Cette originalité a notamment permis, à la lumière de l'analyse que l'on vient de mener, d'étudier et de représenter les distributions géographiques des activités économiques, mais aussi de la densité de population, des fortunes, des ménages imposables, des âges ou encore de la taille des ménages. L'exploitation de ces visualisations est donc à

²⁰³ Herlihy et Klapisch-Zuber, 272.

²⁰⁴ Herlihy et Klapisch-Zuber, 272.

²⁰⁵ Herlihy et Klapisch-Zuber, 277

considérer comme l'un des apports fondamentaux de l'utilisation des ordinateurs pour les recherches historiques dans ce contexte.

Comme nous l'avons vu, le choix d'utiliser des méthodes informatiques pour l'analyse des données conduit les auteurs à multiplier les raisonnements statistiques, jusqu'à développer une approche systématiquement statistique, destiné à tester des hypothèses, qui s'apparentent à une démarche expérimentale. Certains programmes ont encore permis de mettre en œuvre des opérations de statistiques multidimensionnelles, comme le calcul de coefficient de corrélations multiples ou les analyses factorielles des correspondances. D'autres sont aussi utilisés pour visualiser les données, sous la forme de diagrammes ou de cartes, et pour donner ainsi une autre image des sources.

Ces différents apports des programmes peuvent par ailleurs être combinés, et conduisent ainsi à la mise en œuvre de raisonnements complexes qui transforment la façon de critiquer les données issues de la source, ou permettent d'établir les relations entre les caractères des populations et d'en tester la significativité. Pour saisir pleinement les conséquences historiographiques de ces évolutions, nous nous attacherons dans la quatrième partie à la réception de l'étude, comme à la réutilisation de l'« édition » produite dans ce cadre.

4. Réutilisations de l'« édition » et des conclusions de l'étude.

L'analyse des conditions de fabrication de l'« édition » pour l'ordinateur et des raisonnements mis en œuvre au cours de l'étude a permis d'examiner les pratiques, les moyens matériels et les organisations collectives développés par les acteurs pour atteindre leurs objectifs. Ils visaient, comment nous l'avons vu, à produire une étude répondant à des problématiques démographiques et historiques autant qu'à fournir un outil de travail aux historiens voulant étudier le *catasto* florentin. Pour analyser l'impact du recours aux ordinateurs dans ce contexte, il nous faut également nous intéresser à la réception par les pairs des deux objets qui résultent de cette entreprise : l'« édition » pour l'ordinateur du *catasto* et le livre publié en 1978, *Les Toscans et leurs familles*. Cette analyse nous permettra de saisir comment l'utilisation des ordinateurs a pu jouer un rôle dans l'évolution de l'historiographie de l'Italie médiévale à la fin des années 1970.

Pour comprendre le devenir de ces deux objets scientifiques, nous suivrons, à partir des archives conservées au CRH et d'une série d'entretiens menés auprès de Klapisch-Zuber et de Demonet, les reprises de l' « édition » informatique du *catasto* par d'autres acteurs, ainsi que les références à l'ouvrage dans d'autres travaux, aussi bien en démographie, qu'en histoire et en histoire de l'art. Comme nous allons le voir, la trajectoire dessinée au terme de cette analyse présente les conséquences historiographiques de l'étude sous un jour à la fois intra-disciplinaire et extra-disciplinaire.

4.1. Réutilisations de l' « édition ».

La réutilisation de l' « édition » est une problématique centrale si l'on veut comprendre la portée du projet mené par Klapisch-Zuber et Herlihy. Premièrement parce que la volonté d'élaborer un instrument de travail pour l'ensemble des chercheurs intéressés par le *catasto* était un objectif clairement formulé par Herlihy dès 1966. Mais aussi parce que les commentateurs voient dans cette « édition » informatique un instrument de travail utile à de nombreux chercheurs. Philippe Wolff, qui préface l'édition française de l'étude, met d'ailleurs en lumière l'importance des promesses que porte avec elle cette entreprise :

*« Des nuées d'érudits de tout poil – dont je serai, bien sûr – vont s'abattre sur la mine si habilement exploitée, si prometteuse encore. »*²⁰⁶.

Wolff cite les historiens, les démographes, mais aussi les ethnologues parmi les érudits potentiellement intéressés par ce document. Au cours d'une série d'entretiens réalisée entre 2016 et 2017, Klapisch-Zuber nous assure par ailleurs que le fichier a été massivement utilisé, au moins ponctuellement, par des chercheurs qui s'intéressaient à certains personnages florentins de la période.

²⁰⁶ Herlihy et Klapisch-Zuber, *Les Toscans et leurs familles*, 5.

4.1.1. Réutilisations ponctuelles.

Des documents dans les archives du CRH permettent de préciser ces déclarations générales et attestent de cet intérêt pour le fichier informatique, parfois même bien avant la publication du livre. Dans une lettre du 13 décembre 1971, Klapisch-Zuber informe par exemple Herlihy de l'intérêt d'un archiviste de Florence, responsable de la conservation d'une partie du *catasto*, pour un objet dérivé de cette « édition » :

« [Pampaloni] s'est montré très intéressé par l'acquisition d'une liste des chefs de familles, avec les références d'archives, qui pourrait servir d'index aux chercheurs travaillant sur le catasto, ou voulant l'utiliser pour une recherche particulière ».

Selon Klapisch-Zuber, cet index imprimé est encore aujourd'hui utilisé par les visiteurs de l'*Archivio di Stato* de Florence. Cet élément nous renseigne sur les nombreuses possibilités que l'édition des données pour l'ordinateur offre, par exemple pour les archivistes du *catasto*. La possibilité d'établir des index de noms ou de matière sur la base de l'édition constitue, dans ce contexte, l'un des apports documentaires fondamentaux de ce nouveau type d'objet scientifique.

Les archives permettent de saisir un autre exemple de reprise de l'édition. Il s'agit d'une demande formulée par Stuart Laws, épidémiologiste à l'université de Newcastle. Ce chercheur s'intéresse alors à la fréquence et la distribution des maladies dans les populations humaines dans la longue durée. Les données du *catasto* lui serviraient donc d'étalon pour la fin du Moyen-Âge. Une lettre, qu'il adresse à Klapisch-Zuber en 1994, atteste de la réception d'un échantillon du fichier, portant sur 9 780 familles. Laws demande par ailleurs dans cette même lettre si le reste du fichier peut lui être envoyé. Malheureusement, les archives ne nous permettent pas de savoir si Klapisch-Zuber a bien envoyé le reste du fichier. Il est par ailleurs difficile de dire si la recherche de Laws a abouti, d'autant que son travail n'a, à notre connaissance, débouché sur aucune publication. Il semble cependant ressortir de la correspondance entre Laws et Klapisch que le choix des informations enregistrées, comme le codage défini par Herlihy et Klapisch-Zuber n'aurait pas permis à Laws d'effectuer les opérations qui l'intéressaient.

Malgré l'absence de détails, cet exemple nous renseigne sur deux points particulièrement intéressants. Le premier concerne l'intérêt d'un chercheur extérieur aux disciplines des sciences humaines pour le jeu de données. Les promesses qu'incarne cette étude dépassent donc non seulement les frontières de l'histoire et de la démographie, mais aussi la frontière des sciences humaines et sociales. Ce cas fait par ailleurs apparaître les limites posées à la reprise de l'édition produite par Klapisch-Zuber et Herlihy, puisque l'épidémiologiste se trouve dans l'incapacité d'exploiter le fichier pour ses recherches. En serait-il de même pour la réutilisation des données par un autre historien ?

En l'absence de cas de ce type dans les archives, il nous est malheureusement impossible de répondre en s'appuyant sur des faits. Il nous est cependant possible de donner un éclairage sur les limites de cette « édition » informatique. Le code établi par Klapisch-Zuber et Herlihy dérive d'intérêts disciplinaires et d'injonctions matérielles qui, de fait, réduisent la collection de questions qui peuvent être posées sur la base de leur édition. Parmi les données écartées, on peut par exemple citer les données cadastrales, déjà étudiées par Conti, ou les données onomastiques, qui auraient demandé l'établissement de plus de cartes perforées.

Ce caractère incomplet ne disqualifie cependant pas cette édition aux yeux de leurs auteurs. Klapisch-Zuber n'a par exemple jamais cessé, jusqu'en 2016, de rechercher la pérennisation de cet objet éditorial, certainement convaincue de son intérêt pour la communauté des chercheurs. Qu'en reste-t-il aujourd'hui ? Et qui utilise encore cette base de données ? Au contraire de la grande majorité des travaux de ce type réalisés dans les années 1970²⁰⁷, le fichier du *catasto* existe toujours, sous la forme d'un fichier Excel. En effet, grâce à l'aide de Nicolas Veysset notamment, mais aussi de Béatrice Marin, tous deux archivistes au CRH, le fichier a pu migrer, année après année, du format bande magnétique initial vers un format SAS en 1990, puis vers le format Excel entre 2013 et 2016, passant au travers des différentes révolutions *hardware* et *software* qui se sont succédées.

²⁰⁷ Dans la plupart des archives laboratoires que nous avons pu visiter au cours de cette thèse, les éditions pour l'ordinateur élaborées dans les années 1970 et 1980 n'ont pas pu être pérennisées. Les nombreuses révolutions technologiques *hardware* et *software* nécessitaient en effet de nombreuses migrations des formats de stockage. Deux laboratoires font cependant figures d'exception : le CRH, qui a su faire pérenniser quatre de ces projets, et l'Institut de Recherche et d'Histoire des Textes (IRHT). Les raisons de ce succès sont abordées au chapitre suivant, à partir de la comparaison avec une entreprise de pérennisation ratée.

Si l'accès à l'édition est aujourd'hui facilité par cette migration vers des formats dits commerciaux, qu'en était-il pendant les années 1980 ? En l'absence d'archives, il reste possible de regarder ailleurs pour éclairer ces zones d'ombres. Dans un article de 1988, Manfred Thaller, spécialiste du traitement de l'information historique, présente les problèmes que rencontre l'ensemble des historiens qui cherchent à réutiliser ce type d'« édition »²⁰⁸. Il identifie trois obstacles à l'utilisation plus large des bases de données en histoire : 1) le doute des utilisateurs face aux difficultés d'en comprendre les logiques de constitution ; 2) l'extrême diversité des programmes de gestion de telles bases, qui rend la consultation des données extrêmement complexes et 3) les difficultés d'accès à ces objets liés aux modalités d'échanges prévus par leurs concepteurs, que les bases de données soit accessibles seulement sur les lieux de leurs créations, ou via un réseau informatique. Ces maigres indices que nous donne Thaller en 1988 permettent d'envisager les conditions de réutilisation de l'édition du *catasto* dix ans après son élaboration.

4.1.2. Réutilisations programmatiques.

Un autre type de réutilisations apparaît encore dans les archives du CRH. Il s'agit de la reprise du code par l'un des élèves de Herlihy, Anthony Mohlo (1939-), qui a pour ambition de construire, sur la base de l'exemple du *catasto* de 1427, une édition informatique d'un autre *catasto* florentin, celui de 1480. La finalité de son étude est de comparer les données de ces deux sources. La filiation avec le projet de Herlihy est ici totale, non seulement au regard de la position institutionnelle et hiérarchique de son auteur, mais aussi de l'inscription du projet dans la droite ligne de celui qu'avait formulé Herlihy en 1966 : créer une bibliothèque d'éditions pour l'ordinateur de sources médiévales. Un modèle de ce type de bibliothèque existait d'ailleurs déjà aux États-Unis depuis 1962 pour l'histoire politique contemporaine : l'Inter-University Consortium for Political and Social Research, établi par le politologue Warren E. Miller (1924-1999) à l'université du Michigan, qui avait justement pour objet « *to keep on file quantitative data of all kinds* »²⁰⁹.

²⁰⁸ Manfred Thaller, « Data bases v. critical editions », *Historical social research/Historische sozialforschung* 13, n° 3 (1988): 129-39.

²⁰⁹ Shorter, *Historian and the Computer*, 16.

Dans l'optique de mettre en œuvre un pôle de ce type en histoire médiévale, Herlihy adresse une lettre à Klapisch-Zuber le 28 avril 1967, dans laquelle il lui demande s'il existe des travaux de ce type en France. Klapisch-Zuber demande alors autour d'elle et répond le 8 mai à Herlihy qu'un centre à Nancy, dirigé par Jean Glénisson (1921-2010), mène, en partenariat avec l'Institut de Recherche et d'Histoire des Textes (IRHT), des enquêtes basées notamment sur l'analyse de contenu et l'analyse littéraire des textes du Moyen-Âge. La réutilisation du code par Mohlo et cette lettre de Herlihy nous montrent qu'au-delà de l'enquête sur le *catasto* se constitue à la fin des années 1960 une forme d'ambition disciplinaire autour de l'utilisation des moyens informatiques en histoire, au moins aux États-Unis. Malheureusement, en l'absence d'archives permettant l'étude, nous regarderons un autre type de réception des *Toscans et leurs familles*, plus traditionnelle, par la communauté des médiévistes, puis par certains démographes.

4.2. Réception du livre du côté des historiens : un étalon pour l'histoire de l'Italie médiévale.

L'analyse des réutilisations des données et de la méthodologie ne saurait faire oublier que le principal succès de l'entreprise de Herlihy et Klapisch-Zuber semble être la partie recherche et conclusions de l'étude, qui se manifeste à travers la publication du livre *Les Toscans et leurs familles* en 1978. Une rapide recherche sur les citations qui renvoient à l'ouvrage nous mène à identifier près de 500 références, dans de nombreux domaines des sciences humaines et sociales : démographie, histoire sociale, histoire économique ou sciences politiques. Il faut encore ajouter à celles-ci les citations des versions américaines et italiennes de l'étude, que nous n'avons pas prises en compte dans ce rapide dénombrement. Elles permettraient sans doute de montrer encore plus évidemment l'étendue du succès de ce travail. Mis à part ces considérations sur le succès du livre, on peut se pencher sur les recensions du livre pour saisir plus précisément la nature concrète de la réception.

4.2.1. Critiques élogieuses.

Du côté des historiens tout d'abord, l'étude est saluée presque unanimement. Jean-Philippe Genet (1944-) l'encense notamment, dans une recension publiée dans *Francia* en 1978 :

« grâce à David Herlihy et à Christiane Klapisch, grâce aussi à la discipline imposée par l'ordinateur qu'ils ont su dompter à la masse fourmillante des données, le médiéviste dispose d'un fantastique « étalon », à l'aune duquel il est possible de mesurer l'histoire démographique et sociale de l'Europe du XIV^e au XVI^e siècle »²¹⁰.

Aux États-Unis, le *Journal for Interdisciplinary History* lui consacre même trois recensions différentes dans un numéro paru à l'hiver 1981. Dans l'une d'entre elles, l'historien américain Edward Muir (1946-) va lui aussi dans le sens de Genet, en affirmant :

« The author's compilation of data is impressive and likely to become the cornerstone of future investigation of Renaissance Italian social history »²¹¹.

Ces deux auteurs, parmi d'autres²¹², vantent l'intérêt de l'étude en tant qu'elle constitue une base pour la comparaison avec des analyses quantitatives futures. Ce statut d'étalon de mesure historique et démographique repose sur deux principaux éléments.

Tout d'abord, l'enquête donne des résultats quantitatifs pionniers pour l'étude de la famille et de la démographie à la fin du Moyen-Âge, et apparaît ainsi comme un élément indispensable pour de futures comparaisons avec d'autres régions ou d'autres

²¹⁰ Jean-Philippe Genet, « David Herlihy et Christiane Klapisch-Zuber, Les Toscans et leurs familles. Une étude du catasto florentin de 1427, 1978 », *Francia* 7 (1979): 711.

²¹¹ Edward Muir, « New Light on Old Numbers: The Political and Cultural Implications of Les Toscans et Leurs Familles », *The Journal of Interdisciplinary History* 11, n° 3 (1981): 477.

²¹² Il faut ajouter à ces textes deux autres recensions publiées dans le *Journal of Interdisciplinary History* de 1981 : Susan Mosher Stuard, « Family Confronts the Renaissance Household: A Review of Les Toscans et Leurs Familles », *The Journal of Interdisciplinary History* 11, n° 3 (1981): 495-501 ; Myron P. Gutmann, « Demographic and Economic Issues in Les Toscans et Leurs Familles », *The Journal of Interdisciplinary History* 11, n° 3 (1981): 487-93. Un dossier du même type, constitué de quatre articles portant directement sur le livre de Klapisch-Zuber et Herlihy, paraît la même année, dans le numéro 12 de la revue italienne *Società e storia*.

périodes historiques. Ensuite, la précision et la justesse des observations développées par Klapisch-Zuber et Herlihy à l'aide des méthodes statistiques et informatiques assurent à ces résultats une fiabilité applaudie par les pairs. Dans une note critique publiée en 1980 dans les *Annales*, Philippe Braunstein (1933-), historien médiéviste et maître-assistant au CRH, présente ces deux éléments en ces termes :

*« La perfection des détails, saisis dans un vaste panorama, tient à la fois à la qualité des sources qui se prêtaient aux comptages, et à la volonté de deux historiens qui ont choisi de traiter par l'informatique les données d'un recensement des biens et des personnes. »*²¹³

Et il continue :

*« La Toscane du premier tiers du XV^e siècle risque d'être, et pour longtemps, un lieu de référence obligé pour les historiens de la société et de la famille, qu'ils soient médiévistes ou modernistes, et des sociologues soucieux d'une archéologie des comportements. »*²¹⁴

À ce titre, un exemple montre bien l'intérêt que peut porter un médiéviste convaincu par l'intérêt des méthodes assistées par ordinateur, comme Genet, à un tel travail. Il trouve notamment, dans le choix fait par Klapisch-Zuber et Herlihy d'étudier les causes culturelles de la déformation des âges, un argument pour affirmer que les sources médiévales sont parfaitement exploitables et pertinentes pour une analyse statistique, et ce, en dépit de leurs incohérences. Dans sa recension de 1979, il défend ainsi sa cause :

²¹³ Philippe Braunstein, « La Toscane dans le premier tiers du XV^e siècle : informatique et paysage social (note critique) », *Annales* 35, n° 1 (1980): 42.

²¹⁴ Philippe Braunstein.

« *L'historien superficiel aurait une fois de plus conclu à l'impossibilité d'une statistique à partir des sources médiévales, mais par une observation méticuleuse nous pénétrons l'attitude de l'homme médiéval face aux nombres* »²¹⁵.

La pratique de la source mise en oeuvre par Klapisch-Zuber et Herlihy, qui ont, rappelons-le, décidé d'enregistrer fidèlement les âges donnés par les contribuables dans leurs déclarations, alors même qu'ils savaient, pour certains d'entre eux, que ces âges étaient faux, est applaudie par Genet. Elle montre, selon lui, la valeur d'une approche statistique pour l'étude des sources médiévales, qui est pourtant mise en doute par certains médiévistes²¹⁶.

Pour Muir, la valeur de l'étude dérive d'un autre aspect méthodologique, rendu possible, nous l'avons vu, par l'utilisation des ordinateurs : le choix de traiter l'ensemble des déclarations du *catasto* et non seulement une partie. Dans sa recension de 1981, il salue cette décision des auteurs :

« *The book will find favor among historians to supplant the method of sampling only a few distinguished lineages, used in previous studies of the Florentine family and extends its net far beyond the urban elite, abandoning the communal orientation of so much of Italian historiography for a regional one.* »²¹⁷

Là aussi, c'est le rapport à la source développé par les auteurs, à travers la réalisation d'une édition des données portant sur l'ensemble des individus du *catasto*, qui est valorisé. Au contraire d'études précédentes, qui ne se concentraient que sur certains individus, ou d'autres travaux, qui avaient choisi de n'étudier qu'un échantillon des déclarations, le travail de Klapisch-Zuber et Herlihy permet de donner, pour la première fois, une vision complète du territoire florentin. Ce second aspect de la pratique des sources encensé ici par Muir renforce le statut d'« *étalon* » de l'enquête pour l'histoire de l'Italie médiévale.

²¹⁵ Genet, « David Herlihy et Christiane Klapisch-Zuber, Les Toscans et leurs familles. Une étude du *catasto* florentin de 1427, 1978 », 709.

²¹⁶ Nous renvoyons ici à la deuxième partie de cette thèse, et, notamment, au chapitre 3, dans lequel nous conduisons une analyse des débats qui animent les premiers médiévistes intéressés par les usages des ordinateurs.

²¹⁷ Muir, « New Light on Old Numbers », 478.

4.2.2. Limites méthodologiques de l'étude.

Pourtant, en dépit de cette reconnaissance du travail dans son ensemble, les auteurs de ces recensions formulent aussi des critiques plus négatives, qui tiennent notamment aux méthodes d'analyses des corrélations entre les variables démographiques et économiques. Genet regrette par exemple que :

« l'ordinateur ait été souvent employé à peaufiner la description plutôt qu'à rechercher des corrélations ou chercher à faire apparaître les structures complexes au moyen d'analyses factorielles (par contre utilisées au moins par Ch. Klapisch dans des articles isolés) »²¹⁸.

L'une des limites du modèle d'histoire quantitative des années 1960-1970, souvent pointée par certains médiévistes comme Genet ici, ou comme Alain Guerreau (1948-) ailleurs²¹⁹, apparaît comme flagrante pour certains lecteurs de l'étude. L'intérêt qui se développe dans les années 1970, chez les médiévistes notamment, pour les apports « qualitatifs » de l'usage des ordinateurs, semble en effet absent du livre. Demonet, le statisticien en charge de l'enquête du *catasto* du côté français, est pourtant un véritable spécialiste de l'analyse factorielle de correspondances, méthode qu'il a notamment mise en œuvre dans le cadre d'autres enquêtes historiques.

Dès lors, comment comprendre l'absence d'approches comparables dans *Les Toscans et leurs familles* ? Herlihy s'est-il opposé à l'utilisation de ces nouvelles méthodes ? Si oui, pour quelles raisons ? Ou, encore, les auteurs ont-ils fait le choix de ne pas exposer cette méthodologie dans leur argumentation, tout en la prenant en compte dans leurs raisonnements ? Dans un entretien que nous avons réalisé en 2020, Klapisch-Zuber affirme que Herlihy aurait en effet été opposé à l'utilisation des analyses factorielles dans le livre. Cependant, l'absence d'archives étayant ces discussions au CRH

²¹⁸ Genet, « David Herlihy et Christiane Klapisch-Zuber, *Les Toscans et leurs familles. Une étude du catasto florentin de 1427, 1978* », 711.

²¹⁹ Les critiques formulées par Guerreau sur cet usage descriptif des méthodes informatiques sont notamment développées dans le chapitre 4 de cette thèse.

nous oblige à la prudence, et donc, à reporter ces questions à plus tard afin de leur donner une réponse plus précise et plus documentée.

Un autre type de critique portant spécifiquement sur la méthode est aussi formulé par Muir. En 1981, il pointe l'échec des auteurs à réconcilier les approches quantitatives et les autres, plus traditionnelles, basées sur l'énorme corpus de sources littéraires à la disposition de l'historien de Florence pour la période médiévale. Muir développe notamment cette critique à partir de l'exemple d'un débat historiographique portant sur les profits que pouvaient tirer de la guerre les différentes classes sociales de Florence²²⁰. Chez les contemporains, Rinaldo degli Albizzi (1370-1442), homme politique florentin, affirmait en effet que la guerre profitait surtout au *popolo minuto*, c'est-à-dire, aux petites gens, qui en profitaient pour vendre leurs provisions aux mercenaires et ne supportaient pas la charge de l'impôt dû à la guerre. Au contraire, Machiavel (1469-1527) affirmait que la guerre enrichissait les grandes familles de Florence. Face à cette dissonance dans les sources de l'époque, Muir attendait de l'étude du *catasto* une réponse qui n'apparaît pas dans le livre de Klapisch-Zuber et Herlihy, puisque les conclusions portant sur la distribution de la richesse permettent, selon lui, de soutenir les deux hypothèses.

Muir conclut alors en ces termes :

« *On this issue, cliometry does not appear to have advanced us far beyond choosing side* »²²¹.

Pour Muir, si l'étude est, peut-être, plus scientifiquement précise (« *scientifically accurate* »)²²² que ses prédécesseurs, elle ne donne pas d'arguments qui permettraient de répondre à la problématique de l'enrichissement en période de guerre, du fait de son extrême spécialisation méthodologique. L'historien de la Toscane reste donc, en dépit de l'étude quantitative proposée, confronté à deux discours opposés entre lesquels il ne peut trancher. Ces deux citations de Muir, et son choix d'utiliser le terme « *cliometry* », éclaire

²²⁰ Cette question, soulevée auparavant par Nicolas Machiavel dans son *Histoire de Florence*, est d'actualité aux États-Unis parmi les historiens de l'Italie médiévale, plus particulièrement depuis 1965 et la publication d'une nouvelle traduction de Machiavel par un professeur américain : Niccolò Machiavelli, *Machiavelli: The Chief Works and Others, Vol. I* (Duke University Press, 1989).

²²¹ Edward Muir, « New Light on Old Numbers: The Political and Cultural Implications of Les Toscans et Leurs Familles », 482.

²²² Muir, 483.

l'un des aspects fondamentaux du contexte historiographique dans lequel s'inscrit de la réception de l'enquête du *catasto* outre-Atlantique.

En effet, les années 1970 marquent l'essoufflement du programme historiographique de l'histoire sociale quantitative comme de la *cliometry*, tous deux dominants dans le champ académique, respectivement en France et aux États-Unis, dans les années 1950 et 1960. Entre 1973 et 1976, plusieurs publications majeures viennent cependant remettre en cause l'idée de la possibilité d'une pratique de l'histoire calquée sur les valeurs épistémiques des sciences sociales. C'est notamment le cas des travaux de l'historien de la critique littéraire Hayden White (1928-2018), qui publie en 1973 le fameux livre *Metahistory : The Historical Imagination in Nineteenth-Century Europe*, qui conduira à de nombreuses réactions dans les milieux académiques anglophones autour des approches narrativistes de l'écriture de l'histoire²²³.

On peut, dans le même ordre d'idée, citer les travaux qui se développent en Italie autour de ce que l'on a pu qualifier, *a posteriori*, de *microstoria*, dont le livre de Carlo Ginzburg (1939-), *Il formaggio e i vermi. Il cosmo di un mugnaio del '500*, publié en 1976, est un élément marquant. À rebours de l'exhaustivité et de la scientificité qu'assurerait l'utilisation des méthodes quantitatives pour l'étude des classes populaires, Ginzburg y défend au contraire une méthodologie basée sur l'analyse d'un cas particulier, qui doit permettre, selon lui, de dépasser les limites des autres approches, et en particulier de celles des historiens quantitativistes français²²⁴. Ce dernier exemple nous éclaire sur le contexte historiographique de la critique formulée par Muir en 1981, en montrant que se sont développées, depuis le début de l'enquête du *catasto* initiée en 1966, de nouvelles approches méthodologiques qui sont à l'opposé de celles mises en œuvre par Klapisch-Zuber et Herlihy.

4.2.3. Réutilisations des conclusions.

Un autre aspect de la réception de l'étude concerne évidemment l'utilisation des conclusions de l'étude par les historiens. Les références au travail de Klapisch-Zuber et

²²³ Voir, par exemple Wulf Kansteiner, « Hayden White's Critique of the Writing of History », *History and Theory* 32, n° 3 (1993): 273-95.

²²⁴ Voir en particulier la préface du livre, dans laquelle Ginzburg détaille les spécificités de son approche par contraste avec les autres tendances historiographiques de l'époque : Carlo Ginzburg, *Il Formaggio e i vermi: il cosmo di un mugnaio del '500* (Torino, Italie: G. Einaudi, 1976).

Herlihy présentent un panorama de la diversité de ces utilisations au sein d'autres raisonnements historiques. Il reste cependant possible de dégager des tendances, comme l'inscription des données numériques du cas florentin dans une chronologie plus longue²²⁵, la comparaison avec d'autres aires géographiques²²⁶, l'utilisation d'une conclusion au service d'un raisonnement plus large²²⁷, ou comme éléments de contextualisation pour une étude sur une période postérieure²²⁸.

Des champs de recherche peuvent encore être identifiés comme particulièrement liés au travail sur le *catasto* : l'histoire économique, à travers l'étude historique des dettes gouvernementales, l'histoire des épidémies, et notamment à travers l'analyse des effets politiques et démographiques des périodes de pestes, l'histoire du genre et l'histoire de la sexualité. Quant à cette dernière catégorie, on peut par exemple mentionner le livre de l'historien italien Guido Ruggiero (1944-) : *The Boundaries of Eros, Sex Crime in Renaissance Venice*²²⁹, mais aussi le travail de Richard Trexler (1932-2007)²³⁰ ou de Michael J. Rocke²³¹. Ce dernier utilise, dans son livre de 1996, *Forbidden friendships*, certaines conclusions de Klapisch-Zuber et Herlihy sur la structure sociale du mariage en Toscane et ses conséquences sur la présence massive d'hommes dans les villes²³². Dans ce travail, l'utilisation des conclusions du *catasto* permet à l'auteur d'obtenir certaines données démographiques absentes des archives sur lesquelles il s'appuie : les *Ufficiali di notte*.

Il ressort de ces considérations que les historiens se servent bien des résultats du livre de Klapisch-Zuber et Herlihy comme d'un point de comparaison pour explorer des questions différentes, mais néanmoins connexes, à celles qu'abordaient les auteurs.

²²⁵ Paolo Malanima, « The Long Decline of a Leading Economy: GDP in Central and Northern Italy, 1300–1913 », *European Review of Economic History* 15, n° 2 (août 2011): 169-219 ; Giovanni Federico et Paolo Malanima, « Progress, Decline, Growth: Product and Productivity in Italian Agriculture, 1000-2000 », *The Economic History Review* 57, n° 3 (2004): 437-64.

²²⁶ Steven A. Epstein, *Genoa and the Genoese, 958-1528* (Univ of North Carolina Press, 1996); K. Geert Rouwenhorst et William N. Goetzmann, *The Origins of Value: The Financial Innovations That Created Modern Capital Markets* (Oxford University Press, 2005); Brent D. Shaw, « The Age of Roman Girls at Marriage: Some Reconsiderations », *The Journal of Roman Studies* 77 (1987): 30-46.

²²⁷ Richard Lachmann, *States and Power* (Polity, 2010).

²²⁸ R. Burr Litchfield, *Emergence of a bureaucracy: the Florentine patricians (1530-1790)* (Princeton (N.J.), États-Unis d'Amérique: Princeton University press, 1986).

²²⁹ Guido Ruggiero, *The boundaries of eros: sex crime and sexuality in Renaissance Venice* (New York, États-Unis d'Amérique, Royaume-Uni de Grande-Bretagne et d'Irlande du Nord, 1985).

²³⁰ Richard Trexler, « La prostitution florentine au XV^e siècle », *Annales* 36, n° 6 (1981): 983-1015.

²³¹ Michael Rocke, *Forbidden friendships: homosexuality and male culture in Renaissance Florence* (New York, États-Unis d'Amérique, 1996).

²³² Rocke.

Certains résultats apparaissent cependant plus réutilisés que d'autres, comme le constat d'une présence abondante d'hommes dans les villes. Les questions soulevées par les auteurs des *Toscans et leurs familles* n'ont toutefois pas, à notre connaissance, été reprises telles quelles par d'autres historiens, en raison de facteurs qui tiennent autant à la spécialisation méthodologique de l'étude qu'à l'évolution des intérêts historiographiques des historiens des années 1990 (histoire culturelle, histoire globale, *microstoria*). Il n'en va cependant pas de même en démographie, comme nous allons le voir.

4.3. Citations et réutilisations en démographie.

Sans revenir sur certains points précis de l'étude, sur lesquels les critiques, positives ou négatives, abondent (intérêt pour l'histoire de la famille, pour l'apparition d'une morale de la richesse, effets de la peste de 1347 sur la démographie), réaffirmons que les références au livre furent pléthore et qu'elles débordent largement du cadre disciplinaire de l'histoire. Parmi les travaux qui le citent, on compte d'ailleurs une majorité d'ouvrages et d'articles de démographie, qui portent particulièrement sur les structures familiales en Europe. Pour saisir la nature de cette réception, nous nous appuyerons ici sur deux recensions publiées dans des revues de démographie par des spécialistes de la démographie historique.

4.3.1. Une « référence obligée de toute recherche sur les sociétés médiévales ».

La première est de la plume de l'un des collaborateurs du projet, Dupâquier, et paraît dans la revue de démographie *Population* en 1979. Dans ce contexte, les évaluations sont plus dithyrambiques :

*« un grand livre, qui me semble le plus important de tout ce qu'a produit la démographie historique sur le Moyen-âge [...] une œuvre monumentale »*²³³.

²³³ Jacques Dupâquier, « Herlihy D., Klapisch-Zuber C. — Les Toscans et leurs familles. Une étude du Catasto Florentin de 1427 », *Population* 34, n° 1 (1979): 216.

Plus loin, Dupâquier se félicite notamment de voir l'une de ses hypothèses, formulée en 1972²³⁴, confirmée dans le cas étudié par Klapisch-Zuber et Herlihy : le nombre de feux sur un territoire donné est lié aux ressources économiques disponibles. L'analyse rejoint d'autres conclusions chères aux démographes, comme l'efficacité de l'utilisation de la distribution log-normale pour la définition des types de localités (grandes villes, villes moyennes, villages), développée dans une étude sur la province belge de Hainaut pour une période comparable.

La seconde recension qui nous intéresse ici est le fait du médiéviste français Henri Dubois (1923-2012), spécialiste de la démographie médiévale, qui se joint aux compliments formulés par Dupâquier, en affirmant dans les *Annales de démographie historique*, en 1978, que le livre est désormais une « référence obligée de toute recherche sur les sociétés médiévales »²³⁵.

Toutefois, cet intérêt des démographes dépasse le cadre des simples recensions. En effet, en parcourant les travaux qui citent l'ouvrage, on découvre qu'un grand nombre d'entre elles sont le fruit de reprises des tableaux, des graphiques, ou même seulement des valeurs numériques publiés par Klapisch-Zuber et Herlihy. Dans le contexte de la démographie, les *Toscans et leurs familles* se présente ainsi également comme un étalon, qui va jusqu'à être utilisé comme exemple pour la confirmation ou l'infirmité de théories démographiques. La plus célèbre des réutilisations du *castato* est, certainement, à ce titre, la reprise des données pour comparaison des résultats avec d'autres travaux du même type par John Hajnal dans un article de 1982²³⁶.

4.3.2. Une réutilisation marquante des résultats démographiques.

Dans son introduction, Hajnal précise l'objet de son article : il s'agit d'étudier le système de formation des ménages du Nord-Ouest de l'Europe²³⁷ au XVII^e et XVIII^e siècles, en le comparant avec d'autres systèmes. Pour poursuivre cette comparaison,

²³⁴ Jacques Dupâquier, « De l'animal à l'homme : le mécanisme autorégulateur des populations traditionnelles », *Revue de l'Institut de Sociologie*, n° 2 (1972): 177-211.

²³⁵ Henri Dubois, « D. Herlihy et Ch. Klapisch-Zuber, Les Toscans et leurs familles, une étude du castato florentin de 1427, 1978 », *Annales de Démographie Historique* 1979, n° 1 (1979): 466.

²³⁶ John Hajnal, « Two Kinds of Preindustrial Household Formation System ».

²³⁷ Hajnal inclut, dans cette aire géographique, les pays scandinaves (Finlande exclue), les îles britanniques, les pays germanophones, le nord de la France et les Pays-Bas. Hajnal, 449.

Hajnal formule un certain nombre de règles que suivrait le modèle du mariage nord-ouest européen. L'une d'entre elles concerne l'âge auquel les hommes deviennent chefs de familles. Malheureusement, comme l'indique Hajnal dans son article, il est très compliqué de trouver des données pour étudier ce phénomène dans le cas des *joint households system*²³⁸. Pour la période préindustrielle et pour le Nord de l'Europe, Hajnal choisit de s'appuyer sur le travail de Jørgen Elklit (1942-) sur la démographie du Danemark. Pour comparer ces données avec celles d'autres régions, Hajnal affirme :

« *The only comparable set of data for a joint household system known to me comes from a population not yet discussed: that of fifteenth century Tuscany (Italy), whose rich records (compiled in connection with taxation) have been analyzed in magnificent detail by Herlihy and Klapisch-Zuber* »²³⁹.

La comparaison des modèles danois et toscans passe alors par une reprise des données de chacune des études, sous la forme de tableau et de graphes.

Deux éléments particulièrement intéressants pour notre sujet peuvent être dégagés de cet exemple. Tout d'abord, Hajnal intègre les données du *catasto* dans un corpus plus large, composé des résultats d'autres travaux de recherche en démographie. En ce sens, et comme l'affirme Hajnal, le travail de Klapisch-Zuber et Herlihy vient combler un vide dans la bibliographie sur les « *joint household systems* ».

Le second élément concerne les modalités de réutilisations du travail de Klapisch-Zuber et Herlihy. Hajnal précise en effet dans son article qu'il emprunte ces données à « *Klapisch et Demonet (1972)* », c'est-à-dire, certainement à l'article « *A uno pano e uno vino. La famille rurale toscane au début du XVème siècle* » paru dans les *Annales* en 1972²⁴⁰. On remarque pourtant que le graphe et le tableau qu'il présente sous cette référence ne sont pas publiés tels quels dans l'article des médiévistes. De plus, à notre connaissance, Hajnal n'a pas eu de contact direct avec Klapisch-Zuber ou Demonet, et n'a donc pas pu profiter de l'« édition » pour l'ordinateur qu'ils avaient constituée. La

²³⁸ Pour une définition de ces *joint household systems*, voir l'introduction de l'article de Hajnal, et plus particulièrement les pages 450 et 451.

²³⁹ John Hajnal, « Two Kinds of Preindustrial Household Formation System », 464.

²⁴⁰ Christiane Klapisch-Zuber et Michel Demonet, « A uno pane e uno vino : la famille rurale toscane au début du XV^e siècle ».

question qui se pose est donc de savoir si Hajnal a pu extraire les données de l'article, les retravailler, et les utiliser pour mettre en œuvre une comparaison avec celles produites par Elklit sur le Danemark. Dans ce cas, une autre utilisation des résultats apparaîtrait et consisterait à extraire des données de la version publiée sur papier pour les réutiliser dans d'autres opérations statistiques.

Ces deux éléments montrent que la réception des résultats a été plus sensiblement une réussite que celle de l'« édition ». Il ne faudrait cependant pas ici perdre de vue le biais qui pourrait affecter notre étude du fait que nous n'avons pu nous appuyer que sur les archives françaises. L'idéal, pour compléter cette recherche, serait donc de consulter les archives de Herlihy, conservées aux archives de l'université Brown à Providence (Rhode Island). En dépit de ce biais documentaire, il ressort que l'intérêt des chercheurs en démographie pour les valeurs numériques produites dans le cadre des analyses statistiques semble plus important que celui qu'y trouvent les historiens. En effet, la force de cette étude du *catasto* repose plus sur la solidité épistémologique des données produites, qui permettent la remise en cause ou la confirmation d'hypothèses, notamment démographiques, que sur l'originalité des conclusions historiques formulées par les auteurs.

Conclusion.

Au terme de cette longue analyse, quelles évolutions liées aux usages des ordinateurs est-il possible de mettre en lumière ? Et qu'est-ce que cette enquête du *catasto* nous apprend de l'utilisation des ordinateurs en histoire médiévale dans les années 1960 et 1970 ? Pour répondre à ces deux questions, il faut reprendre les conclusions de chacune des quatre parties de l'analyse : organisations collectives, opérations de constitution du corpus, raisonnement statistiques et réception de l'étude par les pairs.

Tout d'abord, nous avons vu que l'enquête du *catasto* s'inscrit, du côté français, dans un modèle d'organisation collective qui existe déjà à la VI^e section de l'EPHE en 1966. Klapisch-Zuber et Herlihy peuvent ainsi compter sur une constellation de ressources humaines et matérielles qui préexistent à leur étude : 1) des « petites mains », vacataires ou thésards, employés par le CRH (Asdrachas et Triantafyllidou), 2) des

spécialistes d'autres disciplines que l'histoire comme la démographie (Dupâquier), la mécanographie (Couturier), les statistiques et l'informatique (Démonet) ou la cartographie (Bertin) ; un réseau qui s'étend en dehors de l'École et qui permet 3) des relations avec le CNRS pour la recherche de financement et l'accès aux centres de calculs et 4) des contacts avec une entreprise privée de mécanographie pour la perforation des cartes.

Nous avons ensuite montré que l'enquête du *catasto* s'intègre dans un modèle de division du travail à l'échelle disciplinaire, qui existe par exemple en histoire moderne ou en démographie historique, et qui s'est développé sous l'impulsion d'historiens comme Labrousse ou Braudel. Ce modèle s'organise de la façon suivante : le développement d'enquêtes monographiques régionales, basées sur des sources « sérielles », doit permettre la mise en œuvre de comparaisons entre différentes périodes et/ou régions, sur la base de mesures établies à l'aide de méthodes statistiques. L'enquête du *catasto* se présente, sous cette lumière, comme l'un des éléments de la mise en œuvre de ce modèle pour l'étude des sociétés et de la démographie du Moyen-Âge.

Ces considérations permettent même d'affirmer que le travail de Klapisch-Zuber et Herlihy s'inscrit, du côté français, dans la continuité d'une forme encore plus ancienne d'enquête collective, mise en œuvre dès le début des années 1930 par Marc Bloch (1886-1944) et Lucien Febvre (1878-1956), les deux fondateurs de la revue des *Annales d'histoire économique et sociale* en 1929. L'historien André Burguière (1938-), auteur d'une histoire intellectuelle des *Annales*²⁴¹, en donne les trois caractères principaux dans un article daté de 2005 : un « questionnaire interdisciplinaire », une « perspective comparatiste » et l'« exploration d'un nouveau type de source »²⁴².

À la lumière de cette analyse, il ressort que, dans ce cas, le recours aux ordinateurs ne modifie pas fondamentalement les modalités de la constitution d'un collectif de travail, mais s'inscrit dans la continuité d'un modèle préexistant. Par contraste, nous avons cependant mis en lumière que les modalités de la mise en œuvre du travail collectif évoluent, et ce, notamment, en raison des méthodes de constitution du corpus et de préparation des données.

²⁴¹ André Burguière, *L'école des annales : une histoire intellectuelle*.

²⁴² André Burguière, « Archéologie du Centre de Recherches Historiques. Les enquêtes collectives de Marc Bloch et de Lucien Febvre et leur postérité », *Les Cahiers du Centre de Recherches Historiques. Archives*, n° 36 (30 juin 2005).

Sur ce point, le premier impact fondamental de l'utilisation des moyens informatiques est corrélé à la dimension du corpus traité. En effet, pour la première fois, les auteurs sont en mesure d'étudier exhaustivement les 60 000 déclarations des contribuables de l'administration florentine. Pour ce faire, Klapisch-Zuber et Herlihy sont contraints de s'appuyer sur un type de document propre aux technologies informatiques de l'époque, les cartes perforées IBM à 80 colonnes, qui supposent, du fait de la structure de l'information qu'elles impliquent, la formalisation de règles de travail communes extrêmement rigides.

De nouvelles procédures de collecte des données sont développées, sur la base de l'élaboration d'une grille d'analyse qui prend matériellement la forme du bordereau de perforation. Sur la base de la rationalité pratique des administrateurs de l'État florentin, les historiens mettent aussi en place des procédures de vérifications et de correction des données à plusieurs mains. La *culture de travail*, qui s'invente ici à l'échelle locale pour la collecte et l'édition des données, dérive manifestement de l'option prise d'utiliser des ordinateurs pour traiter cette gigantesque source.

La formalisation de la collecte des données qui s'opère dans ce cadre constitue le fond de ce que Furet avait qualifié de passage « *de l'implicite à l'explicite* », dans un article désormais célèbre, publié en 1971²⁴³. Il précisait dans ce texte cette expression en ces termes :

*« l'utilisation de l'ordinateur par l'historien n'est pas seulement un immense progrès pratique, par le gain de temps qu'il permet [...]; c'est aussi une contrainte théorique très utile, dans la mesure où la formalisation d'une série documentaire destinée à être programmée oblige par avance l'historien à renoncer à sa naïveté épistémologique, à construire son objet de recherche, à réfléchir à ses hypothèses et à passer de l'implicite à l'explicite. »*²⁴⁴

Cette « *contrainte théorique* » que suscite l'ordinateur s'observe très nettement dans notre étude de cas. Tout d'abord, d'un point de vue pratique, avec la nécessité, du fait du

²⁴³ François Furet, « Histoire quantitative et construction du fait historique », 67.

²⁴⁴ Furet, 67.

travail collectif, de rendre explicite les opérations mentales autrefois muettes, afin que chacun des membres de l'équipe puisse dépouiller la source et remplir les bordereaux d'une façon identique. Ensuite, d'un point de vue épistémologique, du fait de la nécessité d'élaborer un code qui contribue à définir, en amont de l'analyse, des hypothèses de travail, et, donc, à réduire les objets qu'il sera possible d'étudier.

Avec cette première étude de cas, et pour les opérations de collecte des données et d'élaboration du matériau d'analyse, on remarque que l'utilisation des ordinateurs constitue une véritable révolution dans les formes du travail de ces historiens. Cette révolution tient non seulement, nous l'avons dit, à l'originalité de l'environnement technique dans lequel ils évoluent, par rapport à d'autres équipes d'historiens aux environnements de travail plus traditionnels.

De plus, nous avons montré que l'ordinateur produit de la méthode, en raison de la formalisation parfaitement claire de l'enchaînement des opérations qu'il exige. Comme le faisait remarquer Henry dans une citation mentionnée dans la deuxième partie de ce chapitre, l'informatique présente un intérêt du seul fait de la nécessité qu'elle impose d'explicitier comment sont prises les décisions à chaque étape de la recherche. Au niveau du traitement des données, la programmation des opérations d'analyse systématise encore plus clairement cette évolution majeure.

Par ailleurs, ce phénomène apparaît, pour l'observateur que nous sommes, comme une formidable aubaine, puisqu'elle permet 1) de découvrir dans le même mouvement une transformation épistémologique et 2) d'analyser des documents sur la base desquels les historiens détaillent leurs raisonnements de manière explicite. Il nous appartient, pour la suite de notre recherche, de saisir en quoi les raisonnements ainsi développés constituent des évolutions par rapport aux méthodes de l'histoire quantitative développées sans ordinateur, tout en prenant en compte le fait que ces dernières étaient le plus souvent implicites, et donc, connaissables à travers des types de documents différents de ceux que nous avons déployés ici.

Un second élément de transformation majeure que nous avons mis en lumière à travers cette étude est qu'il résulte de cette nouvelle manière de formaliser la collecte des données historiques une nouvelle forme d'« édition » des sources historiques. Cet *instrument de travail*, que les auteurs nomment différemment en fonction des périodes et

des langues qu'ils parlent – « *computer edition* », « *édition simplifiée* », « *édition abrégée* », « *data bank* » – apparaît dans notre contexte comme une nouvelle forme de publication scientifique, et constitue, à ce titre, l'un des apports incontournables de l'utilisation des ordinateurs dans les années 1970.

Cet exemple, dont nous avons pu voir qu'il en existait d'autres dans d'autres disciplines à la même période (sciences politiques, histoire moderne, démographie), marque une rupture forte dans le monde académique, qui se doit, selon nous, d'être étudiée en détail, à la fois pour être intégrée à l'histoire plus générale de l'édition scientifique, mais également pour permettre d'en saisir les implications sur l'évolution de la discipline historique, dans son organisation comme dans ses méthodes. Ce sujet nous semble d'autant plus important que la production d'instruments de travail numérique (bases de données, éditions électroniques, logiciels) est aujourd'hui l'un des éléments clef dans de nombreux projets bénéficiant de larges financements.

Cet axe de recherche a notamment été initié par le spécialiste de l'histoire de l'archéologie Sébastien Plutniak, dans un article à paraître²⁴⁵. Il y présente l'ambition, portée par l'archéologue et spécialiste de l'automatique documentaire Jean-Claude Gardin dès la fin des années 1950 en France, de

« *développer et promouvoir une forme éditoriale mieux adaptée aux besoins de consultation - et non de lecture - des chercheurs* »²⁴⁶.

Plutniak met en avant trois éléments qui caractérisent ce projet : 1) la volonté de « *faire valoir le fichier mécanographique en tant que format éditorial et réalisation scientifique à part entière* »²⁴⁷ ; 2) les choix technologiques qui s'imposent pour permettre la production, le stockage et la consultation de ces nouvelles formes de ressources et 3) le coût de ces opérations, qui redéfinissent les logiques de l'édition scientifique.

Ces trois éléments mis en avant par Plutniak dans son article se retrouve dans notre étude de cas. Pour le cas de l'édition informatique du *catasto*, nous avons montré

²⁴⁵ Sébastien Plutniak, « Le fichier comme format de publication en sciences sociales : les expériences mécanographiques du CADA à Marseille », in *Marseille en capitales*, éd. Christelle Rabier et Marie-Luce Rauzy (Marseille: Éditions Gaussien, à venir).

²⁴⁶ Plutniak, 3.

²⁴⁷ Plutniak, 3.

que les pairs saluent ce travail pour sa précision et le niveau de détail qu'il offre aux spécialistes. Nous avons encore mis en lumière, dans le même mouvement, qu'en raison des limites techniques et du choix des données enregistrées, la base de données n'a été essentiellement utilisée que de manière ponctuelle par les pairs.

Quant au troisième point, nous avons vu que les éditeurs d'une version pour l'ordinateur du *catasto* envisagent, dès 1968, la constitution d'une bibliothèque de bases de données, semblable, pour l'histoire médiévale, au projet de Gardin mentionné plus haut, ou au travail mené à l'Inter-University Consortium for Political and Social Research, dont nous avons présenté le rôle dans la quatrième partie de ce chapitre. L'ambition qui apparaît ici chez ces médiévistes a vocation à transformer en profondeur l'organisation de la recherche au sein de la discipline, par la production de données sous une forme qui en permette un nouveau type d'utilisation.

Le rêve scientifique qui se dessine ici, derrière la révolution que constitue l'utilisation des ordinateurs, nous permet d'affirmer que ce projet d'édition pour l'ordinateur du *catasto* s'inscrit dans une dynamique plus générale en sciences humaines et sociales : la volonté de transformer la nature des instruments de travail de la recherche. Qu'en est-il pour d'autres publications du même type (édition de texte pour l'ordinateur, programme informatique) ? Quels sont les critères qui président à la définition des logiques éditoriales de ces nouvelles publications ? Et comment ces nouvelles formes de publications scientifiques sont-elles valorisées par leurs auteurs ? Ces questions seront abordées dans le troisième chapitre de cette thèse, à travers l'étude des échanges qui prennent forme au cours d'un colloque qui s'était tenu à Rome en 1975.

Un autre axe central de l'analyse menée dans ce premier chapitre était de comprendre quels avaient pu être les effets de l'usage des ordinateurs sur la pratique des statistiques en histoire, et donc, plus généralement, de ce qu'on appelle communément l'histoire quantitative. Sur ce point, trois principales conclusions ont pu être dégagées.

Premièrement, nous l'avons dit, on constate un changement d'échelle dans la masse d'information à laquelle il est possible d'appliquer des opérations statistiques. L'ordinateur permet sur ce point un bond quantitatif évident. Il faut y ajouter un autre changement d'échelle, qui ne tient plus aux dimensions des corpus, mais à la multiplication significative du nombre d'opérations statistiques par rapport à une étude

antérieure menée sans ordinateur. Cette conclusion se retrouve également pour l'analyse d'une même problématique historique (par exemple la distribution des professions, ou l'âge aux mariages).

Ces deux phénomènes, que nous constatons ici sur pièces, ne sont pas surprenants. Ils tiennent au gain de temps que permet le recours aux ordinateurs dans les opérations de calcul et de classement, qui sont au cœur des procédures statistiques, et qui constituaient, dès le début de l'entreprise de Klapisch-Zuber et Herlihy, l'une des raisons du recours à ces instruments électroniques. Notons toutefois que, dans le cadre de cette étude, les ordinateurs permettent de gagner du temps dans les opérations d'analyses, tout en en faisant certainement perdre au moment de l'édition des données, puisque cette phase dure près de huit ans.

C'est cependant l'analyse des effets de ce double bond quantitatif sur la critique des sources qui constituent, selon nous, le résultat le plus intéressant. Il apparaît en effet, à travers les deux exemples de raisonnement étudiés dans ce chapitre (distribution des métiers et étude de l'âge au premier mariage), que la multiplication des analyses permet d'affiner l'observation historique dans deux directions différentes. Tout d'abord, le recours aux ordinateurs évite aux historiens des erreurs de dénombrement ou de calcul qu'une telle quantité d'information rend presque inévitables. Ensuite, la combinaison d'analyse portant sur des variables différentes permet l'établissement d'observations plus précises. Les historiens peuvent en effet multiplier les analyses et aboutir à une observation plus fiable par la synthèse entre différentes approches.

La deuxième conclusion qu'il est possible de formuler sur ce point s'inscrit dans la continuité de cette dernière considération, et concerne ce que Klapisch-Zuber qualifiait, dans l'introduction de la version française du livre, de « *calculs compliqués* ». Nous l'avons vu, le recours aux ordinateurs permet aux acteurs de mettre en place une série d'opérations statistiques plus complexes que les opérations traditionnelles de l'histoire quantitative. Parmi celles-ci, on retrouve notamment les analyses de corrélations, et, plus marginalement, une analyse de corrélations multiples, qui permettent, pour nos acteurs, d'établir la significativité des relations entre les variables économiques et démographiques (voir **3.3**).

Cette dernière opération, qui s'inscrit plus généralement dans ce qu'on appelle les analyses multivariées, puisqu'analysant dans le même temps plus de deux variables, est

caractéristique de l'intérêt que présentent les ordinateurs pour l'application de méthodologies statistiques à partir des années 1960. En effet, avant l'apparition des ordinateurs électroniques de la deuxième génération, les analyses multivariées ne sont, le plus souvent, possibles que de manière théorique, ou sur des jeux de données de dimension sans comparaison avec la base du *catasto*.

Ce tournant apparaît d'autant plus important que les premiers utilisateurs des analyses multivariées ont, dès le début des années 1970, mis en lumière le bouleversement profond qu'impliquait l'utilisation de ce type de procédures statistiques. Dans un article de 1974, intitulé « Multivariate analysis: Not so Much a Technique as a *Weltanschauung* » le zoologiste Robert Edmund Blackith (1923-2000) affirmait :

*« Some techniques open up new approaches to biological problems; microscopy did not simply enable us to see better what we had already seen dimly, it allowed us to discover micro-organisms with all that discovery implies. Similarly, multivariate analysis does not simply enable us to discriminate ever more effectively as we incorporate more characters into the discriminant, it opens up ways of handling masses of data meaningfully, data which would have defeated us by sheer bulk »*²⁴⁸.

Cette transformation de la vision du monde (*Weltanschauung*) des scientifiques qui utilisent ce type d'analyse n'est cependant pas tout à fait manifeste avec le cas de l'enquête que nous présentons ici, certainement en raison du fait que les auteurs n'y recourent que de façon extrêmement ponctuelle. Le choix que nous avons fait d'exclure de notre analyse les applications des analyses factorielles des correspondances aux données du *catasto*, du fait de l'opposition de Herlihy à ces nouvelles méthodes, rejoint cette dernière considération. Nous reviendrons cependant longuement sur l'apport de cette procédure dans le travail d'autres médiévistes, dans les chapitres 2 et 6 de cette thèse.

Un troisième élément doit encore être mis en lumière pour présenter les évolutions des méthodes de l'histoire quantitative qui se manifestent dans l'enquête du *catasto* en

²⁴⁸ Robert Edmund Blackith, « Multivariate Analysis: Not so Much a Technique as a *Weltanschauung* », *Biometrics* 2, n° 30 (1974): 387.

relation avec l'utilisation des ordinateurs. Il s'agit de l'utilisation de méthodes de cartographie automatique développée par Bertin, et, donc, d'un nouveau type de visualisation rendu possible par le recours aux moyens électroniques de calcul. L'analyse de ces nouvelles formes de visualisations nous a permis de mettre le doigt plus concrètement sur le fait que le changement produit par l'ordinateur ne manifeste pas seulement un double changement d'échelle, mais conduit, également, à des changements d'ordre plus « qualitatif ». Dans notre cas, la raison graphique qui se développe sur la base de ces visualisations enrichit les analyses géographiques des auteurs.

Pourtant, en dépit de la nouveauté de ces deux approches (analyses multivariées et cartographie automatique), leurs utilisations par Klapisch-Zuber et Herlihy rejoignent les conclusions formulées plus haut : elles ont pour objet d'affiner la critique interne des sources, et ne provoquent pas de bouleversement dans les méthodes quantitatives mises en œuvre. De fait, cette nouvelle puissance de calcul ne bouleverse donc pas, dans le contexte de l'enquête du *catasto*, le programme historiographique dans lequel cette étude s'inscrit. Le recours aux ordinateurs ne fait qu'augmenter ce que l'historien Pierre Chaunu (1923-2009) nommait « *le pouvoir d'analyse* » des historiens²⁴⁹, au même titre que l'introduction des statistiques l'avait fait pour l'histoire économique au début du siècle.

Derrière cette utilisation des ordinateurs par ces acteurs, c'est une certaine représentation de la fonction des ordinateurs pour la recherche historique qui se dégage ici. Pour ces médiévistes, l'ordinateur est un outil qui doit permettre de faire plus de calculs, sur plus de données, tout en s'assurant de faire le moins d'erreurs de calculs possibles au cours de ce processus. L'objectif de cette utilisation est de rendre possible une critique des sources plus précise, qui passe par l'élaboration de méthodes de critique des sources plus complexes.

De ce fait, les critiques adressées par les pairs à l'enquête du *catasto* sont presque unanimes. Ce travail constitue un formidable étalon pour étudier les aspects démographiques, sociaux et économiques de la fin du Moyen-Âge en Italie et, plus largement, en Europe. Ce statut peut être résumé par l'expression utilisée par Muir pour qualifier cette enquête : les résultats établis par Klapisch-Zuber et Herlihy sont plus précis

²⁴⁹ Voir par exemple Pierre Chaunu, « Pour une histoire sérielle du Brésil au XVIII^e siècle : Réflexions préalables », *Revue d'histoire économique et sociale* 49, n° 4 (1971): 466.

scientifiquement (« *scientifically accurate* ») – selon une certaine représentation de ce qu'est la science, ce sur quoi nous reviendrons dans la conclusion générale -, ce qui enchante, nous l'avons vu, les démographes. Certains historiens sont quant à eux plus critiques, notamment sur le fait que l'enquête est trop descriptive, ou sur le fait qu'elle ne permet pas de résoudre certaines questions pourtant centrales de l'historiographie de la Toscane pré-moderne.

Ces critiques s'adosent, selon nous, à la représentation de la fonction de l'ordinateur qui marque la recherche de Klapisch-Zuber et Herlihy. En effet, cette dernière s'inscrit dans la continuité des valeurs épistémologiques défendues par le programme historiographique de l'histoire quantitative, qui avait notamment pour objet d'intégrer les méthodes des sciences sociales à la pratique des historiens. À ce titre, l'utilisation des méthodes des statistiques devaient assurer un statut plus scientifique à l'histoire, qui passait par une plus grande rigueur méthodologique et une exigence d'exactitude dans la critique des sources. En 1978, la situation a cependant évolué, et des historiens plus jeunes, nés dans les années 1940, comme Genet ou Muir notamment, critiquent autant les méthodes que les objectifs de cette forme d'histoire quantitative.

Derrière la critique qu'ils font des résultats de l'enquête du *catasto*, c'est en réalité une critique à un certain usage des ordinateurs qui est formulée, qui se rattache directement aux critiques du programme historiographique de l'histoire sociale quantitative défendue, notamment, à l'EHESS. Nous verrons, par comparaison avec le travail d'un autre collectif dans le chapitre suivant, mais plus généralement au fil de cette thèse, que d'autres utilisations des ordinateurs sont possibles, et que ceux-ci s'articulent à la fois avec les moyens financiers et humains de la recherche, avec l'évolution des technologies informatiques et avec les ambitions historiographiques des chercheurs.

Chapitre 2 : Lexicométrie et histoire médiévale : le programme ALINE de l'Équipe de Recherche Associée 713 (1971-1984).

Le premier chapitre de cette thèse nous a permis de dégager certains éléments propres à une enquête historique ayant recours à des ordinateurs en histoire médiévale entre 1966 et 1978. Cependant, l'étude d'un seul cas, bien qu'emblématique dans ce contexte, ne saurait suffire pour apprécier les multiples configurations dans lesquelles ces instruments de calcul ont été mis en œuvre en histoire médiévale. En effet, pour le seul cadre de la France, on compte au début des années 1970 près d'une dizaine de collectifs engagés dans des projets de ce type²⁵⁰. Chacun de ces collectifs développe des formes d'organisations singulières, propose des méthodes de collecte de données plurielles, sur la base de sources de natures différentes, et poursuit des objectifs de recherche variés.

Pour aller plus avant dans cette analyse, nous avons choisi comme objet de ce deuxième chapitre l'un des projets de l'un de ces collectifs. Il s'agit de l'élaboration d'un programme de traitement automatique des textes, menée au sein d'une équipe de recherche en histoire médiévale du CNRS, travaillant à l'université Paris 1 Panthéon-Sorbonne : l'Équipe de Recherche Associée 713 (ERA 713). Nous allons le voir, ce choix est motivé par les possibilités de comparaison qu'il offre par rapport au cas de l'enquête du *catasto* étudié dans le chapitre précédent. Cette comparaison portera notamment 1) sur le contexte institutionnel dans lequel se développe ce projet ; 2) sur l'équipement qu'il mobilise ; 3) sur les influences extra-disciplinaires qu'il implique ; 4) sur la nature de l'instrument de travail développé par ce collectif et 5) sur les raisonnements développés par les acteurs.

À l'échelle mondiale, au début des années 70, l'ERA 713 n'est que l'une des nombreuses équipes de recherche qui travaillent avec des programmes de traitement automatique des textes. En 1967, la revue américaine *Computer and the Humanities*

²⁵⁰ Une liste non exhaustive de ces projets est donnée dans l'introduction de cette thèse. Pour une liste semblable, mais certainement plus complète, voir Lucie Fossier et al., « Vingt ans d'informatique en histoire médiévale », *Actes des congrès de la Société des historiens médiévistes de l'enseignement supérieur public* 20, n° 1 (1989): 501-25.

recense ainsi 159 projets de ce type dans diverses disciplines (études littéraires, linguistique, études des classiques, philologie, histoire, sciences politiques, etc.)²⁵¹. Ils sont conduits un peu partout en Europe et en Amérique du Nord (en Italie, en Belgique, en Allemagne, en France, au Royaume-Uni, aux États-Unis, au Canada ou encore aux Pays-Bas), et visent à traiter automatiquement des genres de textes très différents, comme de la poésie, de la littérature en prose, du théâtre, des textes administratifs ou de la philosophie, et ce dans de nombreuses langues, vivantes ou mortes, telles que l'italien, le français, l'allemand, le grec ancien, le sumérien ou encore le latin. Les méthodes et les objectifs de ces recherches sont, eux aussi, très diversifiés, puisqu'on relève des études stylistiques, l'établissement de concordances ou d'index, de la traduction automatique, de l'analyse de grammaire ou encore des approches phonétiques.

Si l'on se concentre sur le contexte français, la recension de la revue mentionne quatre équipes de recherche²⁵², mais il existe, en réalité, bien plus de chercheurs engagés dans cette voie. Quatre collectifs sont particulièrement importants pour le cas que nous avons choisi d'étudier, et ce pour différentes raisons, sur lesquelles nous reviendrons au cours de ce chapitre. Au centre de recherche de lexicologie politique de l'École Normale Supérieure (ENS) de Saint-Cloud²⁵³, plusieurs petites équipes travaillent depuis 1966 à la constitution d'index sur la base de corpus de natures diverses²⁵⁴. Parmi eux, un groupe, composé entre autres des linguistes Annie Geoffroy, Robert Arnault, Liliane Cavaciuti, Françoise Theuriot et de l'historienne Régine Robin (1939-2021), se concentre sur l'étude du vocabulaire de la Révolution française, et publie des index et des tableaux statistiques, sans pour autant avoir recours à des moyens informatiques²⁵⁵.

²⁵¹ « Directory of Scholars Active », *Computers and the Humanities* 2, n° 2 (1967): 71-93.

²⁵² Les quatre collectifs mentionnés ici développent tous des travaux sur la période médiévale. On y trouve le Répertoire général sur cartes perforées des manuscrits antérieurs au XVI^e siècle en latin, hébreu, grec, français, provençal et espagnol, d'Édith Bayer, Françoise Ferrand et Jacqueline Steunou de l'Institut de Recherche et d'Histoire des Textes (IRHT) ; le programme de Traitement automatique des textes médiévaux non littéraires de Lucie Fossier et Jean Schneider, encore à l'IRHT ; le projet de Traitement automatique des textes d'ancien français du XIII^e au XIV^e siècle d'Hélène Naïs au Centre de Recherche Automatique des Langues de Nancy (CRAL) ; et le travail intitulé « Emploi de la machine électronique dans les études médiévales pour tout type de textes de philosophie », mené par Jacques Froger en collaboration avec l'École Pratique des Hautes Études et l'IRHT.

²⁵³ On peut se reporter, par exemple, à l'entretien de Maurice Tournier publié dans *Mots* en 2010 : Maurice Tournier, « Mots et politique, avant et autour de 1980 Entretien », *Mots. Les langages du politique*, n° 94 (30 novembre 2010): 211-33.

²⁵⁴ Maurice Tournier, « Le centre de recherche de lexicologie politique de l'E.N.S. de Saint-Cloud », *Langue française* 2, n° 1 (1969): 82-86.

²⁵⁵ Voir, par exemple le chapitre V de Régine Robin, *Histoire et linguistique* (Paris, France: A. Colin, 1973).

Une autre équipe, qui mène pour sa part une étude historique, collabore au même moment avec ce centre. Elle utilise, quant à elle, les moyens informatiques mis à disposition par l'ENS de Saint-Cloud pour étudier le vocabulaire des proclamations électorales à la fin du XIX^e siècle. Ce projet, brièvement mentionné dans le chapitre précédent, est dirigé par l'historien Antoine Prost (1933-), et mobilise une petite équipe composée de quelques étudiants avancés. Pour la mise en œuvre des opérations statistiques et informatiques, Prost bénéficie de l'aide de Michel Demonet, acteur majeur, rappelons-le, de l'enquête assistée par ordinateur du *catasto* florentin menée par Christiane Klapisch-Zuber (1936-) et David Herlihy (1930-1991). L'équipe de Prost et Demonet exploite alors le Barodet, recueil de proclamations électorales conservé aux archives de l'Assemblée nationale, dans le but d'établir des inventaires de mots et mener des analyses statistiques sur le lexique politique de la fin du XIX^e siècle²⁵⁶.

L'autre collectif qui s'avèrera important pour notre étude n'est pas basé en France, mais en Belgique, et s'intéresse plus particulièrement au vocabulaire de la période médiévale. Il a été constitué dans le cadre d'une collaboration entre les équipes du Laboratoire d'Analyse Statistique des Langues Anciennes de Liège (LASLA)²⁵⁷ et celles du Centre de Traitement Électronique des Documents situé à Louvain (CETEDOC)²⁵⁸. Depuis 1962, les deux fondateurs de ce second centre de recherche, le philologue Paul Tombeur (1936-), formé aux méthodes informatiques au sein du LASLA, et le médiéviste Léopold Genicot (1914-1995), avaient notamment développé « *un projet de mise en ordinateur des textes édités écrits en latin avant 1200* » sur le territoire de l'actuelle Belgique²⁵⁹, dans un but plus philologique que lexicologique.

²⁵⁶ Les résultats de cette enquête sont publiés dans Antoine Prost, *Vocabulaire des proclamations électorales de 1881, 1885 et 1889* (Paris, France: Presses universitaires de France, 1974).

²⁵⁷ Voir, par exemple Louis Delatte et Étienne Évrard, « Un laboratoire d'analyse statistique des langues anciennes à l'Université de Liège », *L'Antiquité Classique* 30, n° 2 (1961): 429-44. Voir par ailleurs les publications dans la *Revue de l'Organisation Internationale pour l'Étude des Langues Anciennes par Ordinateur (R.O.I.E.L.A.O)*, et plus particulièrement les articles suivants : Étienne Évrard, « Deux programmes d'ordinateur pour l'étude quantitative du vocabulaire », *Revue de l'Organisation Internationale pour l'Étude des Langues Anciennes par Ordinateur*, n° 2 (1967): 81-96; Arthur Bodson et Suzanne Govaerts, « Codification d'un texte latin sur cartes mécanographiques 80 colonnes », *Revue de l'Organisation Internationale pour l'Étude des Langues Anciennes par Ordinateur*, n° 1 (1966): 1-49.

²⁵⁸ Pour une présentation du CETEDOC par son directeur voir : Paul Tombeur, « Le Centre de Traitement Électronique des Documents, au service des disciplines ecclésiastiques », *Revue Théologique de Louvain* 1, n° 4 (1970): 433-43. Pour avoir une idée des résultats de ce type de travaux, voir par exemple Paul Tombeur, *Raoul de Saint-Trond: Gesta abbatum Trudonensium* (Hildesheim, Allemagne: G. Olms, 1969).

²⁵⁹ Pour une brève présentation de ce projet par l'un de ses auteurs, voir Léopold Genicot, « Index Scriptorum Operumque Latino-Belgicorum Medii Aevi. Nouveau répertoire des œuvres médiolatines belges publié sous la direction de L. Genicot et P. Tombeur. Première partie : VIIIe-Xe siècles par A. Stainier

Enfin, le quatrième collectif qu'il faut mentionner se trouve au Centre de Recherche et d'Application Linguistique (CRAL) de l'université de Nancy, où une équipe, notamment composée de médiévistes, enregistre, puis traite à l'aide des ordinateurs les chartes originales antérieures à 1220 conservées dans les archives départementales françaises. Cette initiative permet aux médiévistes, selon les mots de Michel Parisse (1936-2020), directeur de la section des textes diplomatiques médiévaux au CRAL, de

« définir pour chaque mot la chronologie de son emploi [...], le type d'actes où il se trouve le plus employé [...], également la géographie de son emploi. »²⁶⁰

Ces quatre exemples dessinent un programme d'automatisation de l'analyse du vocabulaire, conduit par des historiens, qui relève d'une dynamique disciplinaire plus profonde. Cette dynamique n'est pas directement liée à l'apparition des ordinateurs, mais va trouver un nouvel essor avec l'apparition de ces nouveaux moyens électroniques de calcul. Dès le début des années 1930, quelques historiens s'intéressent à l'étude du vocabulaire, notamment dans le contexte de ce que les observateurs ont qualifié, dans les années 1960, d'histoire des mentalités²⁶¹. Lucien Febvre (1878-1956), célèbre cofondateur en 1929 de la revue des *Annales d'histoire économique et sociale*, s'intéresse par exemple à la linguistique historique et à la linguistique pré-structurale, dans le but d'étudier la trajectoire de certains mots dans les œuvres de personnages marquants du XVI^e siècle (François Rabelais et Martin Luther). L'objet de son étude est alors de saisir ce qu'il appelle « *l'outillage mental* » de ces acteurs²⁶². Pour l'étude du vocabulaire médiéval, Marc Bloch (1886-1944) avait lui aussi proposé, dès 1928, une enquête sur les mots de l'histoire, ceux des sources, mais aussi ceux des historiens²⁶³.

(Académie royale de Belgique. Comité national du dictionnaire du latin médiéval), Bruxelles, 1973 », *Bulletins de l'Académie Royale de Belgique* 60, n° 1 (1974): 29-30.

²⁶⁰ La citation, pour laquelle la source n'a pas pu être retrouvée, est reprise de l'annexe « Lexicographie, typologie informatique » du livre suivant : Jean-Pierre Poly et Éric Boumazel, *La mutation féodale : X^e-XII^e siècles* (Paris, France: Presses universitaires de France, impr. 1980, 1980), XXIII.

²⁶¹ Voir Georges Duby, « Histoire des mentalités », *L'histoire et ses méthodes*, 1961, 937-65.

²⁶² Voir, par exemple Lucien Febvre et al., *Civilisation: le mot et l'idée* (Paris, France: F. Alcan, 1930); Lucien Febvre, *Le problème de l'incroyance au XVI^e siècle: la religion de Rabelais* (Paris, France: Éditions Albin Michel, 1942).

²⁶³ Voir l'article de Michael Werner, « « Les Mots de l'histoire » et la Begriffsgeschichte / sémantique historique », *Revue de l'IFHA. Revue de l'Institut français d'histoire en Allemagne*, n° 4 (30 septembre 2012): 187-94.

Toutefois, ces premières approches sont remises en question par des linguistes, qui critiquent particulièrement les critères de sélection des mots étudiés. Le sémioticien français Algirdas Julien Greimas (1917-1992), spécialiste de la lexicologie, instruit par exemple cette critique dans une recension de 1958, à l'attention d'un ouvrage consacré à l'étude du vocabulaire de la médecine au XVII^e siècle de son collègue Bernard Quemada (1926-2018)²⁶⁴. Dans ce court texte, publié dans la revue les *Annales*, Greimas insiste sur la nécessité pour les historiens de se dégager d'une étude trop subjective du vocabulaire pour s'intéresser aux ensembles lexicaux tels que définis par la linguistique saussurienne. D'une étude de certains mots-pivots, Greimas incite les historiens à basculer progressivement vers l'étude d'ensembles lexicaux, voire, de tout le vocabulaire d'un auteur ou d'un ensemble de sources.

Cependant, les instruments de recherche manquent pour étudier les langues du Moyen-Âge. Les historiens intéressés par ces approches multiplient ainsi, dès le début des années 60, des appels aux linguistes afin de constituer de meilleurs instruments de travail pour la lexicologie. Georges Duby (1919-1996) fait, par exemple, dans un texte de 1961, un appel du pied à la collaboration entre linguistes et historiens :

*« Comment pénétrer dans la conscience des hommes de tel milieu, comment expliquer leur conduite, les relations qu'ils entretiennent, essayer de voir le monde et autrui par leurs yeux mêmes, sans bien connaître le vocabulaire qu'ils emploient [...] donc sans disposer d'un inventaire systématique et chronologique des mots ? Si bien que l'histoire des mentalités ne peut progresser sans le concours des lexicologues. Elle attend d'eux – et impatientement, en les pressant d'utiliser toutes les ressources nouvelles de la mécanographie – les listes, les dénombrements de vocables. »*²⁶⁵

Dans le contexte des années 1950 et 1960, alors que les études sur le vocabulaire médiéval se multiplient, et que les linguistes, en France, s'intéressent de plus en plus aux approches statistiques de la langue²⁶⁶, les calculateurs électroniques apparaissent comme des outils

²⁶⁴ Algirdas Julien Greimas, « Histoire et Linguistique », *Annales* 13, n° 1 (1958): 110-14.

²⁶⁵ Duby, « Histoire des mentalités », 953.

²⁶⁶ Voir Sylvain Loiseau et Jacqueline Léon, éd., *History of quantitative linguistics in France* (RAM-Verlag, 2016). Voir plus particulièrement Jacqueline Léon, « The Statistical Studies of Vocabulary in the

formidables pour poursuivre ce programme historiographique de l'histoire des mentalités en histoire médiévale.

Pourtant, si l'automatisation de certaines branches de la linguistique, et notamment la traduction automatique, est déjà bien avancée dès les années 1950 dans d'autres pays, comme les États-Unis ou le Royaume-Uni, il n'en va cependant pas de même en France. L'historienne de la linguistique Jacqueline Léon, dans un article de 2010, présente les raisons du retard français dans ce processus de « *mathématisation-automatisation* »²⁶⁷. Parmi elles, elle met l'accent sur l'absence de moyens électroniques de calcul dans les centres de recherche français avant 1955, date de l'installation d'un premier ordinateur au sein de l'Institut Blaise Pascal à Paris sous l'impulsion du CNRS. La première équipe qui travaillera sur la traduction automatique n'est par ailleurs créée au sein de ce même Institut que quatre ans plus tard, en décembre 1959²⁶⁸.

Dans cet article, Léon insiste par ailleurs sur le fait que plusieurs traditions s'affrontent au sein de cette jeune discipline qu'est la linguistique computationnelle : d'un côté une approche formelle de la langue, qui ne touche que très peu les historiens, et de l'autre une approche statistique qui se base sur le dénombrement d'éléments observables, comme les mots. À ce titre, la France fait alors figure d'exception dans le paysage de la linguistique mondiale²⁶⁹. Cette situation au sein des sciences du langage explique sûrement l'intérêt qu'ont alors eu les historiens pour les approches statistiques de la langue.

En 1973, l'historienne Régine Robin, qui travaille avec le centre de recherche de lexicologie politique de Saint-Cloud, publie un ouvrage qui offre un tableau des différentes approches inspirées de la linguistique en histoire. Ce livre témoigne, dans sa deuxième partie, de la diversité des méthodes d'approche des textes que les historiens, essentiellement des modernistes, ont mis en œuvre depuis la deuxième moitié des années

1950-60s in France. Theoretical and Institutional Issues », in *History of Quantitative linguistics in France* (RAM-Verlag, 2016), 9-28.

²⁶⁷ Jacqueline Léon, « Automatisation-mathématisation de la linguistique en France dans les années 1960. Un cas de réception externe », *2ème Congrès Mondial de Linguistique Française*, 2010, 825. Nous nous basons ici sur la version de cet article édité en tant que chapitre 7 du livre suivant : Jacqueline Léon, « Histoire de l'automatisation des sciences du langage » (Lyon, France: ENS éditions, 2015).

²⁶⁸ Léon, « Automatisation-mathématisation de la linguistique en France dans les années 1960. Un cas de réception externe ».

²⁶⁹ Léon, « The Statistical Studies of Vocabulary in the 1950-60s in France. Theoretical and Institutional Issues ».

1960 et le travail pionnier de Prost sur les proclamations électorales²⁷⁰ : statistiques lexicales, étude des champs sémantiques, analyse des énoncés, analyse sémique, auxquels il faudrait ajouter analyse de contenu pour être plus complet.

C'est dans ce contexte que certains historiens de l'université Paris 1 Panthéon-Sorbonne décident de se lancer dans la création du programme de traitement automatique des textes ALINE (pour Analyse LINGuistiqueE). En 1973, ce groupe de recherche, qui prendra en 1979 la forme d'une Équipe de Recherche Associée (ERA), suivant des modalités sur lesquelles nous reviendrons, est dirigé par le médiéviste Bernard Guenée (1927-2010). Ce groupe s'inscrit dans une structure plus large au sein de l'université Paris 1 Panthéon-Sorbonne : une Unité d'Enseignement et de Recherche (UER), dénommée « Centre de Recherche d'histoire de l'occident médiéval », constituée de deux autres équipes de recherche. Les membres de cette unité devaient, dans ce cadre, mener à la fois des missions d'enseignement et de recherche.

Dans un rapport sur la recherche à l'université Paris 1 Panthéon-Sorbonne, publié en 1973, le domaine de recherche de l'équipe dirigée par Guenée est défini en ces termes :

« l'histoire des institutions, des faits, des idées politiques et de l'historiographie aux derniers siècles du Moyen Age »²⁷¹.

²⁷⁰ Prost, *Vocabulaire des proclamations électorales de 1881, 1885 et 1889*.

²⁷¹ Ce document nous a été transmis par le service des archives de l'université Paris 1 Panthéon-Sorbonne sous une forme scannée. Précisons ici, pour la suite, que ce chapitre s'appuie également sur des sources qui nous ont été transmises directement par Jean-Philippe Genet, parmi lesquels quatre rapports d'activités de l'ERA 713 datés de 1979, 1981, 1983 et 1985, ainsi que plusieurs documents internes à l'équipe. Ajoutons qu'aucun inventaire ni classement n'a été dressé pour ces documents, et il nous est donc impossible de rendre compte aux lecteurs de leur cote d'archive. Pour garantir la possibilité de retrouver ces documents, nous indiquerons en note la nature, l'auteur et la date du document. Pour le document ici cité nous adopterons par exemple la présentation suivante : « La recherche à Paris 1 » (Paris: Université Paris 1 Panthéon-Sorbonne, 1973), 112. Je remercie encore une fois ici Jean-Philippe Genet, sans qui ce chapitre n'aurait pas pu voir le jour.

Sa particularité, par contraste avec les deux autres groupes qui travaillent au sein de l'U.E.R.²⁷², est de regrouper des chercheurs qui mettent en œuvre des moyens informatiques pour mener à bien leurs recherches. Ce rapport de 1973 nous renseigne d'ailleurs sur le premier projet de ce type initié au sein de l'équipe. Il est mené par Françoise Autrand (1932-) dans le cadre de sa thèse d'État. Il s'agit d'une

*« analyse par fiches perforées des carrières des gens du Parlement de Paris au bas Moyen Age »*²⁷³.

À ce chantier de prosopographie automatique, certains membres de cette équipe ajoutent un autre projet à teneur informatique, qui consiste en ce qu'on qualifie à cette période sous l'appellation de traitement de textes. Schématiquement, la finalité d'une telle méthode, sur laquelle nous reviendrons longuement, est de compter les mots d'un texte pour en proposer une étude basée sur des analyses statistiques. Ce sont particulièrement Claude Gauvard (1942-), qui prépare alors une thèse d'État sur le sujet « Littérature et idées politiques en France pendant le règne de Charles VI » et surtout Jean-Philippe Genet (1944-), lui aussi engagée dans une thèse d'État sur le sujet « Les idées politiques et sociales en Angleterre au XV^e siècle », qui alimentent ce projet.

Dans un document interne à cette équipe, non daté, mais fort probablement élaboré entre 1973 et 1975, Genet présente en ces termes l'objectif du programme ALINE :

« Pour permettre à l'historien de tirer le plus grand parti possible des sources de type littéraire, capitales pour le spécialiste d'histoire politique, pour faire sortir l'histoire des idées du domaine des impressions subjectives et des analyses pointillistes où elle a été trop longtemps cantonnée, et la faire entrer dans celui d'un quantitatif rigoureux par l'étude du support privilégié qu'est le texte, il nous est apparu nécessaire de construire un programme qui soit spécifiquement destiné

²⁷² Les deux groupes en question sont les suivants : 1) le centre de paléographie et sciences auxiliaires, dirigé par Robert Fossier (1927-2012) et 2) l'équipe de Pierre Toubert (1932-), dont l'intérêt est centré sur l'histoire économique, sociale et religieuse du monde méditerranéen latin.

²⁷³ « La recherche à Paris 1 », 113.

à l'historien »²⁷⁴.

Deux éléments s'avèrent significatifs dans la perspective d'une comparaison avec l'enquête du *catasto* étudiée dans le chapitre précédent. Le premier concerne le champ de recherche dans lequel s'inscrit cette étude : l'histoire des idées. En effet, nous avons vu que l'enquête de Klapisch-Zuber et Herlihy s'inscrivait dans des domaines de recherche (histoire sociale et démographie historique) dans lesquels les méthodes quantitatives étaient déjà fortement déployées depuis au moins la fin des années 1950. Au contraire, l'élaboration de méthodes quantitatives pour l'étude des idées politiques est présentée par Genet comme une véritable innovation méthodologique, ce qui apparaît, selon nous, être une représentation justifiée.

Le second élément concerne la volonté de produire un instrument de travail pour les historiens, mais d'une toute autre nature que le projet d'« édition » pour l'ordinateur de Klapisch-Zuber et Herlihy : un programme informatique. Pour comprendre l'originalité du projet ALINE dans ce contexte, il faut préciser qu'il existe, au début des années 1970, de nombreux programmes de traitement automatique des textes, comme COCOA, conçu à Cardiff au milieu des années 1960 par G. L. M. Berry-Roghe²⁷⁵, ou encore GENDEX, qui date du tout début des années 1970²⁷⁶. Il existe également des programmes plus intimement attachés à un type de corpus, comme le programme du CETEDOC de Louvain pour le traitement des textes littéraires en latin, les programmes du CRAL pour le traitement des textes diplomatiques, ou ceux du laboratoire de lexicologie politique de l'ENS de Saint-Cloud élaborés pour l'étude des textes politiques contemporains en français. Ces programmes partagent la particularité de fournir à l'utilisateur des documents de sortie comparables à ceux que souhaitent obtenir les

²⁷⁴ Le document dont est tirée cette citation nous a été transmis par Jean-Philippe Genet en 2017, et n'est donc pas référencé dans un fonds d'archives. Il porte cependant un titre : « Programme de traitement automatique des textes pour l'équipe de recherche sur l'histoire politique du bas Moyen-Âge », et nous nous y référons à partir d'ici sous la forme suivante : « Programme de traitement automatique des textes pour l'équipe de recherche sur l'histoire politique du bas Moyen-Âge », document interne à l'ERA 713, non daté.

²⁷⁵ Voir D. B. Russell, « COCOA-a word count and concordance generator », Associates Technology Literature Applications Society, 1965. Disponible en ligne : <http://www.chilton-computing.org.uk/acl/applications/cocoa/p001.htm>

²⁷⁶ O. R. Smith, « GENDEX : GENeral inDEX of Words with Context. A Concordance Generator », *Computer Studies* III (1970): 50-53.

membres de l'équipe de Paris 1 en 1973 : dictionnaires de formes, indices de fréquences, références des formes et concordances.

Genet connaît l'existence de ces programmes, puisqu'il les cite dans différents textes de présentation de ALINE dès 1974. Mais plusieurs raisons justifient qu'il ne puisse pas y avoir recours. Pour les comprendre, il faut s'appuyer sur les objectifs que ses concepteurs assignent au programme ALINE. Dans le document interne non daté précédemment cité, Genet les présente en ces termes. Le programme doit :

« être d'un maniement aussi simple que possible : le texte doit donc pouvoir être entré en machine sous la forme la moins élaborée qui soit »²⁷⁷,

et

« pouvoir être utilisé pour des textes écrits en diverses langues : les premiers projets de traitement concernent des textes rédigés en anglais, en français et, secondairement en latin. Il faut qu'aucune adaptation ne soit nécessaire pour passer d'une langue à l'autre »²⁷⁸.

Il doit encore, en fonction des opérations de lemmatisation mises en oeuvre, sur lesquelles nous reviendrons au fil de ce chapitre :

« fournir les renseignements suivants : a/ l'index de toutes les formes d'un texte [...] ; b/ l'index de tous les mots d'un texte avec leurs indices de fréquence [...] c/ un dictionnaire de concordances »²⁷⁹.

Si les « renseignements » que doit fournir ALINE sont donc assez classiques par rapport aux autres programmes existants dans ce contexte, ce sont, d'une part, les langues des corpus étudiés par les médiévistes et, d'autre part, la forme de l'entrée des données

²⁷⁷ « Programme de traitement automatique des textes pour l'équipe de recherche sur l'histoire politique du bas Moyen-Âge », document interne à l'ERA 713, non daté.

²⁷⁸ « Programme de traitement automatique des textes pour l'équipe de recherche sur l'histoire politique du bas Moyen-Âge ».

²⁷⁹ « Programme de traitement automatique des textes pour l'équipe de recherche sur l'histoire politique du bas Moyen-Âge ».

voulues par les membres de l'ERA 713 qui conditionnent la nécessité de fabriquer ce programme.

Les différents éléments dégagés dans cette longue introduction permettent de saisir d'emblée certaines des spécificités de cette entreprise, qui appellent à une comparaison avec l'enquête du *catasto*. Dans ce chapitre, nous comparerons ces deux projets sur les plans suivants : comment les acteurs s'organisent-ils pour mener à bien leur étude et la constitution d'un programme ? Qu'est-ce que ce nouveau type d'objet scientifique ? En quoi est-il original dans le contexte de l'histoire médiévale des années 1970 ? Quels raisonnements sont développés à l'aide de ce programme ? Et sur quels types de textes ? Enfin, comment ce programme a-t-il été reçu par la communauté des médiévistes ? A-t-il été réutilisé ? Et si oui, par qui ?

Ces questions, qui structureront ce chapitre d'une manière semblable au précédent, dans le but de faciliter la comparaison, permettront d'aborder une problématique plus large, celle des conséquences de l'utilisation des ordinateurs sur les pratiques de recherche de ces médiévistes.

Pour aborder ces questions, nous nous appuyerons sur un certains nombres de documents : 1) un vaste corpus d'études portant sur les relations entre histoire et linguistique ; 2) un ensemble de publications de natures diverses de Genet et Gauvard, datées entre 1973 et 1998 ; 3) des entretiens menés avec certains acteurs (et notamment Genet) ; 4) des archives de l'ERA 713, dont certains documents internes qui détaillent les objectifs et les manières qu'avaient les différents acteurs impliqués de concevoir le fonctionnement du programme ALINE ; et 5) les archives du centre de calcul de Paris 1.

Il faut néanmoins préciser dès maintenant que le programme lui-même n'a pas été conservé, comme dans le cas de l'enquête du *catasto*, et qu'il nous est donc impossible de mener une analyse sur la base de cette source, pourtant centrale. Les documents mentionnés ci-dessus permettent toutefois d'avoir un accès partiel, quoiqu'indirect, aux modalités de l'élaboration du programme ALINE et à ses fonctionnalités.

Comme dans le chapitre précédent, nous nous intéresserons dans un premier temps à la forme que prend l'organisation collective qui a conduit à la réalisation des différents éléments de cette entreprise de traitement de texte. Nous analyserons ensuite les modalités de l'édition des textes sur cartes perforées, de la sélection des écrits pertinents aux opérations matérielles de la perforation. Nous regarderons ensuite, à l'aide

des sources disponibles, les raisonnements mis en œuvre par les auteurs à l'aide de ALINE, en montrant comment l'utilisation des ordinateurs modifie les méthodes de l'histoire des idées politiques. Enfin, nous étudierons les modalités de la réception et de la circulation de ce programme, des textes édités au fil de ce travail et des publications que ces deux types d'instruments de travail ont permis de mener en histoire médiévale.

1. Organisations collectives et ordinateurs.

La comparaison entre le projet d'enquête du *catasto* et le projet ALINE met en lumière de nombreuses différences. La première concerne la forme administrative que prennent ces équipes. D'un côté, on trouve une équipe regroupée au sein d'une structure mixte du CNRS et d'une université, l'ERA 713, et, de l'autre, un projet pluri-institutionnel, mené entre trois continents, et doté d'un financement du CNRS sur projet, la RCP 181. La deuxième différence est fortement corrélée à la première. Elle concerne les modes de financement de ces deux entreprises et les disparités qui en découlent quant aux moyens humains et matériels à la disposition des acteurs pour mener leurs recherches. La comparaison présente l'intérêt de permettre, de manière simultanée, de dégager les points communs qui existent entre ces deux études, et de tenter ainsi de formuler des hypothèses sur ce qui est constitutif des entreprises qui utilisent les ordinateurs en histoire médiévale dans les années 1970.

La conduite de cette comparaison n'est pas sans poser des difficultés, qui dérivent de la disparité des sources à la disposition de l'observateur. Nous avons pu, dans le premier chapitre, nous appuyer sur la riche correspondance entre les acteurs. Par contraste, l'étude de ce second cas ne peut s'appuyer sur des sources comparables. Voilà déjà un premier indice de la différence entre ces deux projets, qui supposent, du fait des conditions de travail sur lesquelles nous allons revenir, une communication entre les membres de cette équipe tout à fait différente. Pour mener notre analyse de l'organisation collective du projet ALINE, nous nous appuierons en particulier sur 1) des archives institutionnelles de l'ERA 713 pour la période 1973-1985, récupérées par nous auprès de Genet ; 2) les archives du centre de calcul de Paris 1 et 3) un corpus d'entretiens menés, entre autres, par nos soins, entre 2003 et 2017.

1.1. Cadre administratif du projet et ressources informatiques.

Pour bien comprendre dans quel environnement institutionnel s'inscrit le projet d'élaboration de ALINE, il est nécessaire d'esquisser dans un premier temps une brève histoire de l'ERA 713, qui ne prend cette forme administrative qu'en 1977. Le rapport sur la recherche à Paris 1 de 1973, précédemment cité, constitue ici l'un des documents qui nous permet de présenter les logiques qui président au rassemblement de ses membres sous l'égide d'une structure commune. Au cœur de ce groupe de recherche se trouve le médiéviste Bernard Guenée, spécialiste des institutions médiévales de la fin du Moyen-Âge en France, et professeur à la Sorbonne depuis 1965. Il dirige en 1973 l'une des trois équipes de l'U.E.R. « Centre de recherche d'histoire sur l'Occident médiéval de l'université Paris 1 » : l'Équipe de recherche sur l'Histoire politique du bas Moyen-Âge.

1.1.1. Composition de l'équipe de recherche sur l'histoire politique du bas Moyen-Âge.

Parmi les membres de cette équipe, on compte alors dix-sept médiévistes préparant des thèses de 3^e cycle, ou d'État, sous la direction de Guenée. Comme nous avons eu l'occasion de l'évoquer, Françoise Autrand est l'une de ces médiévistes préparant une thèse d'État. Elle travaille en 1973 à la constitution d'un fichier de cartes perforées rassemblant des informations biographiques sur les gens du Parlement de Paris à la fin du Moyen-Âge, dans le but de faire une sociologie des élites politiques²⁸⁰. Au moins deux autres projets de prosopographie assistée par ordinateur sont encore en cours en 1973 au sein de l'équipe.

Hélène Millet (1944-) poursuit, quant à elle, un doctorat de 3^e cycle. Cette chercheuse y développe, en ayant recours aux ordinateurs, une analyse de la composition du groupe de chanoines exerçant à Laon pour la période 1272-1412²⁸¹. En 1977, une fois son doctorat terminé, elle participe au projet ALINE, selon des modalités que nous préciserons dans la suite de ce chapitre. Millet sera aussi, en 1981, l'une des premières

²⁸⁰ Autrand soutiendra sa thèse en 1978 sous le titre : Françoise Autrand, « Naissance d'un grand corps de l'État : les gens du parlement de Paris, 1345-1454 » (Thèse d'État, France, Université Paris 1 Panthéon-Sorbonne, 1978).

²⁸¹ Voir Hélène Millet, « Les chanoines du chapitre cathédral de Laon (1272-1412) » (Thèse 3e cycle Histoire Paris I, France, 1977).

médiévistes à publier les résultats d'une analyse factorielle dans une revue d'histoire généraliste ²⁸².

Un troisième projet à caractère prosopographique est en cours en 1973. Il compose l'un des aspects de la thèse d'État de Jean-Philippe Genet, « Les idées politiques et sociales en Angleterre au XV^e siècle ». Cette recherche, que l'auteur lance en 1967, s'articule avec le projet ALINE. Pour ce travail, Genet s'inspire particulièrement du concept de *champ*, développé par Pierre Bourdieu (1930-2002), dans le but d'explorer ce qui constitue le champ du politique en Angleterre pour une période allant de 1300 à 1600. Dans la forme finale de ce travail, publié en 1997, Genet traite son objet d'étude sous trois angles : 1) une étude systématique des auteurs engagés dans ce champ, 2) une étude du système de communication de ces auteurs et de la circulation des textes et 3) une étude des mots de ces textes²⁸³.

Ce travail de doctorat s'appuie ainsi sur deux approches complémentaires. La première consiste en l'établissement d'une base de données d'auteurs de textes politiques et historiques²⁸⁴. La seconde consiste en l'analyse lexicale des écrits de ces auteurs. C'est pour poursuivre ce deuxième objectif que Genet va initier, puis piloter le projet d'élaboration du programme ALINE dès le tout début des années 1970. Dans un entretien conduit auprès de lui en 2016, Genet nous signale qu'il s'inspirait alors du travail de Prost sur les proclamations électorales, mené, comme nous l'avons dit, avec Demonet et au contact du centre de lexicologie politique de l'École Normale Supérieure de Saint-Cloud.

Cependant, nous l'avons dit, Genet n'est pas le seul membre de cette équipe intéressé par l'analyse automatique des textes. Une autre étudiante de Guenée, Claude Gauvard, s'intéresse, elle aussi, à l'histoire des idées politiques dans le cadre de son doctorat d'État. Elle travaille alors sur une thèse intitulée « Littérature et idées politiques en France pendant le règne de Charles VI », et utilise des moyens informatiques pour analyser une source sérielle : les lettres de rémission. Par ailleurs, Guenée s'intéresse également à ce type d'approche automatique, dans le contexte d'un projet

²⁸² Hélène Millet, « La composition du chapitre cathédral de Laon au XIV^e siècle : résultats d'une analyse factorielle », *Annales* 36, n° 1 (1981): 117-38.

²⁸³ Voir Jean-Philippe Genet, « Les idées sociales et politiques en Angleterre: du début du XIV^e siècle au milieu du XVI^e siècle » (Thèse de doctorat, France, Université Paris 1 Panthéon-Sorbonne, 1997).

²⁸⁴ Pour une présentation de l'approche prosopographique de Genet voir, par exemple Jean-Philippe Genet, « Histoire sociale et ordinateur », *Publications de l'École Française de Rome* 31, n° 1 (1977): 231-37. Pour un texte plus tardif, voir encore Jean-Philippe Genet, « L'informatique au service de la prosopographie : PROSOP », *Mélanges de l'école française de Rome* 100, n° 1 (1988): 247-63.

d'enregistrement des prologues d'œuvre historiques, qu'il développe en collaboration avec l'Atelier Vincent de Beauvais du CRAL de Nancy²⁸⁵.

De fait, à partir de 1981, les projets de traitement de texte au sein de l'ERA se multiplient, comme l'indique un rapport d'activité de juin 1983. On y retrouve notamment une nouvelle application de ALINE par Marie-Anne de Mercoyrol (1950-), dans le cadre de sa thèse de 3^e cycle dirigé par Jacques le Goff (1924-2014)²⁸⁶, mais également sept projets en cours de développement au sein de l'ERA 173. Nous y reviendrons plus longuement dans la quatrième partie de ce chapitre.

Notre présentation de ces quelques membres et projets de l'équipe montre que l'ERA 713 est structurée autour de problématiques communes depuis 1973 : l'étude des institutions, des faits mais aussi des idées politiques de la fin du Moyen-Âge. Les objets de ces historiens se distinguent donc de ceux de l'enquête du *catasto*, puisqu'ils relèvent davantage de l'histoire des mentalités et de l'histoire politique que de l'histoire sociale ou de la démographie.

Un autre élément différencie encore clairement les deux équipes. Dans le cas de l'ERA 713, les médiévistes poursuivent chacun des enquêtes individuelles. Ils travaillent sur des sources différentes, avec des problématiques de recherche différentes et des objectifs différents. Dans certains cas, comme celui de l'élaboration de ALINE, les acteurs peuvent coopérer pour constituer des instruments de travail commun. Ils peuvent alors être perçus par les observateurs comme procédant d'une initiative collective, mais ils restent subordonnés à des logiques d'organisation de leur travail qui ne trouvent pas leur source dans des objectifs communs fixés par les dirigeants de l'équipe. À ce titre, l'organisation des rapports professionnels entre les membres de l'ERA 713 nous apparaît beaucoup plus horizontale que celle de la RCP 181.

Cette analyse met en lumière un autre élément sur lequel repose la cohérence de cette équipe de recherche. Il s'agit du recours aux moyens informatiques, qui permettent aux auteurs de mener des enquêtes sur la base de méthodes inspirées par d'autres sciences sociales, comme la sociologie ou la linguistique. Comme nous allons le voir, l'existence

²⁸⁵ Voir Hélène Naïs et Jean Schneider, « L'atelier Vincent de Beauvais », *Revue d'Histoire des Textes* 4, n° 1974 (1975): 438-43.

²⁸⁶ De Mercoyrol de Polo de Beaulieu soutiendra ce travail en 1984 sous le titre suivant : Marie Anne Polo de Beaulieu, « Étude et édition d'un recueil d'exempla du XIV^e siècle: la Scala Coeli de Jean Gobi » (Thèse de 3^e cycle, Paris, France, École des hautes études en sciences sociales, 1984).

de cette équipe à Paris 1 s'explique en partie 1) par la présence d'un centre de calcul dans les murs de l'université, qui rend possible, et réaliste, la mise en œuvre de ce type de projet et 2) par la nécessité qu'ont les membres de l'ERA 713 d'obtenir des crédits de calcul pour bénéficier des équipements du Centre Inter-Régional de Calcul Électronique (CIRCE) du CNRS à Orsay.

1.1.2. Le centre de calcul de l'université Paris 1 – Panthéon Sorbonne.

Pour appréhender la nature des ressources informatiques dont disposent les historiens de l'équipe de Guenée, il faut revenir rapidement sur l'histoire de la création de l'université Paris 1 Panthéon-Sorbonne, qui se substitue à La Sorbonne - avec l'université Paris 4 - après les événements de mai 1968. Dès le départ, en plus de raisons politiques, développées ailleurs²⁸⁷, l'université Paris 1 Panthéon-Sorbonne se forme sur un schéma pluridisciplinaire associant les sciences humaines, et donc l'histoire, avec le droit et les sciences économiques.

Dans un entretien que nous avons mené avec Genet en 2017, il signale que ce sont justement les économistes qui auraient permis la mise en place d'un centre de calcul à Paris 1, à partir du tout début des années 1970 (1970 ou 1971 selon les souvenirs de Genet). Dans un autre entretien, mené en mai 2015 par Sébastien Rey-Coyrehourcq, Genet présente le fonctionnement de ce centre, dirigé alors par le « *dieu tutélaire* », Claude Fourgeaud (19..-1990), économiste âgé mais très respecté, qui s'assurait « *que le centre de calcul [ait] toujours [...] les moyens qu'il lui fallait* »²⁸⁸.

Du point de vue de l'opération des machines, c'est Édouard Valensi, enseignant à Paris 1 et ancien polytechnicien, qui était en charge de l'équipement et du personnel. Il partageait alors son emploi du temps entre le centre de calcul de Paris 1 et son travail dans un centre de calcul du ministère de la Défense, où il côtoyait de nombreux informaticiens de haut-niveau. Deux opératrices-perforatrices étaient par ailleurs employées à plein temps pour produire des cartes perforées pour les différents laboratoires de l'université

²⁸⁷ Voir, par exemple, un travail déjà ancien qui traite de ce sujet : Jacques Verger, *Histoire des universités en France* (Toulouse, France: Privat, 1986).

²⁸⁸ Sébastien Rey-Coyrehourcq, « Une plateforme intégrée pour la construction et l'évaluation de modèles de simulation en géographie » (phdthesis, Université Panthéon-Sorbonne - Paris I, 2015), Annexes, 88.

Paris 1. En dehors des historiens, les archives du centre de calcul nous apprennent que leurs services sont particulièrement sollicités par les géographes et les économistes.

Plusieurs autres ingénieurs travaillent aussi au centre, dont Xavier Debanne, qui avait développé pour l'université un programme de gestion de banque de données, BDP (Banque de Données Paris 1). Ce programme fonctionnait sur l'unique ordinateur du centre, un mini-ordinateur Phillips P880, et permettait l'analyse statistique de données enregistrées sur cartes perforées, comme par exemple des tris croisés. Toujours selon Genet, le programme était simple d'utilisation puisque tout était paramétré et qu'il suffisait de suivre les instructions pour mettre en œuvre des traitements statistiques de base. Les données devaient cependant être entrées dans la machine sous forme de codage exclusivement numérique, ce qui posait de fait un problème pour le projet d'analyse automatique des textes imaginé par l'ERA 713.

Pour Genet et Gauvard, cette limite du Phillips P880 (codage exclusivement numérique) constituait un obstacle technique majeur pour mettre en œuvre leurs analyses automatiques des textes. Un autre obstacle technique dérivait de l'absence de programme informatique, au centre de calcul de Paris 1, adapté aux opérations qu'ils entendaient réaliser sur la base des textes (index de fréquences, concordances). Où trouver ce programme ? Et existe-t-il déjà ? L'une des questions qui se posent ici est de savoir pourquoi Genet et Gauvard n'ont pas récupéré, ou utilisé, un programme d'analyse automatique des textes préexistants, comme ceux du centre de lexicologie de Saint-Cloud, qui disposait alors d'une bibliothèque de programmes de ce type ? La réponse à cette question sera longuement développée dans la seconde partie de ce chapitre, mais il nous faut donner quelques éléments ici pour permettre au lecteur de nous suivre dans l'analyse.

Nous l'avons dit en introduction, l'objectif principal et la raison d'être de ALINE est de donner aux médiévistes un programme de traitement de texte qui prenne en compte les caractéristiques des textes médiévaux. Comme l'affirme Genet dans un article de présentation de ALINE paru en 1974²⁸⁹, du fait de leurs modalités d'entrée des documents en machine, les programmes existants ne sont pas adaptés aux sources qui intéressent les historiens de l'ERA 713, et notamment aux textes en langues vernaculaires du bas Moyen-Âge (moyen-français, moyen-anglais).

²⁸⁹ Jean-Philippe Genet et al., « Un programme de traitement automatique des textes : ALINE », *Bulletin du centre d'analyse du discours de l'université de Lille III*, 1974, 96-121.

Le problème que ces historiens rencontrent est avant tout un problème de diversité linguistique : les textes qu'ils étudient, et qu'ils souhaitent comparer, sont dans des langues différentes (moyen-anglais, moyen-français et latin pour Genet, moyen-français et latin pour Gauvard). Ils sont par ailleurs écrits dans des langues dont l'orthographe n'est pas fixée. Comme nous l'avons vu, certains programmes existent déjà au début des années 1970, comme celui du CETEDOC de Louvain. Cependant, ce dernier est prévu pour traiter des textes en latin, ce qui pose deux problèmes distincts.

Premièrement, les règles qui régissent le pré-codage des textes en latin, au moment de leur inscription sur les cartes perforées, ne sont pas assez souples pour permettre de traiter les nombreux problèmes linguistiques que posent les langues vernaculaires des documents étudiés par Genet et Gauvard. Il est donc nécessaire de concevoir un programme qui permette de gérer les nombreux problèmes linguistiques que posent ces textes au moment de leur entrée en machine. La solution proposée par les concepteurs de ALINE est donc de laisser à l'utilisateur le choix des caractères spéciaux qu'il emploie pour chaque difficulté qu'il rencontre (séparateur, ambiguïté sur la forme d'une lettre, lettre disparue de l'alphabet moderne, hyphénation). Dès lors, les difficultés ponctuelles peuvent être aisément surmontées au moment de l'entrée en machine par la mise en place d'une idiosyncrasie propre au texte et à l'utilisateur.

Deuxièmement, il n'existe pas de dictionnaire facilitant la lemmatisation automatique de ce type de langue ancienne. La lemmatisation est alors l'une des opérations majeures de préparation des données textuelles, puisqu'elle conditionne les résultats des opérations de statistiques lexicales comme le calcul de fréquence ou du nombre d'occurrences des mots (nous y reviendrons dans la troisième partie de ce chapitre). Elle consiste dans le regroupement des formes - les mots tels qu'ils apparaissent dans le texte - en lemmes - les mots tels qu'ils apparaissent dans le dictionnaire. Un exemple clair du résultat de cette opération consiste dans le regroupement de l'ensemble des occurrences conjuguées d'un verbe sous leur forme à l'infinitif. Pour faciliter ces opérations, les chercheurs du CETEDOC qui travaillent sur le latin disposent alors de dictionnaires, enregistrés sur cartes perforées. Or, il n'en existe pas pour les langues vernaculaires, pour des raisons qui tiennent autant au travail que cela représenterait qu'aux difficultés que posent ce type de langues.

Pour pallier ce problème, et de la même manière que pour le pré-codage, la solution trouvée par les concepteurs de ALINE est de proposer aux utilisateurs d'assurer cette tâche en partie « à la main ». La lemmatisation semi-automatique mise en place permet au chercheur de contrôler lui-même le passage des formes aux mots, pour des textes où seule la familiarité avec les sources permet de saisir la logique des problèmes orthographiques. Ces contraintes, imposées par les sources, ne sont cependant pas les seules raisons qui poussent les concepteurs de ALINE à opter pour ces solutions.

1.2. Informaticiens vacataires et nature du programme.

Les caractères des sources étudiées par les médiévistes de l'ERA 713 impliquent l'élaboration d'un programme original, adapté à leurs besoins. Toutefois, les raisons qui conduisent à opter pour une absence de règles fixes de pré-codage et une lemmatisation semi-automatique sont aussi liées à des problématiques qui concernent plus directement la structure de l'équipe et les moyens de l'ERA 713. Pour le comprendre, il faut tout d'abord souligner le fait que l'élaboration du programme ALINE nécessite le concours de spécialistes de la programmation informatique. Genet, qui a quelques connaissances de base en programmation et en mathématiques, après avoir notamment suivi les cours de Marc Barbut (1928-2011) à l'université Paris 5 en 1969, se rend vite compte que ses compétences ne sont pas suffisantes pour élaborer seul un programme de ce type.

Dans l'entretien qu'il a réalisé en mai 2015 avec Rey-Coyrehourcq, il formule la situation en ces termes :

« Quand j'ai commencé j'ai fait du FORTRAN, mais je n'ai jamais été capable de programmer des choses trop ardues. J'ai tout de suite vu que si l'on voulait programmer un logiciel d'analyse lexicale, ce n'était pas la peine d'essayer sans aide extérieure »²⁹⁰.

Genet se met donc à la recherche d'informaticiens à même de l'aider. Valensi, qui connaît ce petit milieu, lui permet de rencontrer deux personnes qui vont devenir les ingénieurs

²⁹⁰ Rey-Coyrehourcq, « Une plateforme intégrée pour la construction et l'évaluation de modèles de simulation en géographie », Annexes, 93.

informaticiens du projet ALINE : Jacques Mondelli, formé à l'école des Mines et titulaire d'un doctorat de l'université Berkeley, et François Hucher, qui est diplômé comme ingénieur de l'École Centrale de Paris²⁹¹.

Dans ce même entretien, Genet signale l'un des apports des compétences de ces deux informaticiens en ces termes :

« Ils connaissaient par exemple toute une série de tests statistiques qui permettaient de ranger les lemmes par ordre alphabétique en gagnant du temps, et surtout de l'espace mémoire, car à l'époque les contraintes étaient très grandes, donc ils avaient des tas d'astuces statistiques pour faire avancer plus rapidement la conception des logiciels »²⁹².

Dans les souvenirs de Genet, les compétences de Hucher et Mondelli présentaient un atout majeur : faire gagner du temps dans l'élaboration du programme. Cependant, les modalités administratives de la collaboration avec ces deux acteurs cruciaux pour la poursuite du projet vont entraîner toute une série de problèmes qui vont jouer sur la nature du programme ALINE.

1.2.1. Problèmes de recrutement et objectifs de travail.

Le premier problème concerne le type de contrat que l'équipe de recherche de Guenée peut proposer à Hucher et Mondelli. Dans l'entretien de 2015, Genet indique qu'ils étaient payés « *avec des vacations CNRS* »²⁹³. Ce type de contrat précaire ne permettait pas d'assurer un travail continu de la conception du programme, puisque les crédits de vacations pouvaient parfois manquer. Le travail d'élaboration du programme sera par suite très long, puisqu'il s'étendra, *a minima*, entre 1972 et 1978, période pendant laquelle trois versions différentes de ALINE seront produites. Dans un rapport préparé pour le CNRS en 1978, Genet indique que le travail avance « *à un rythme*

²⁹¹ Ces informations sont reprises d'un article récent : Léo Dumont, Octave Julien, et Stéphane Lamassé, « Articuler histoire et informatique, enseignement et recherche : le PIREH de l'université Panthéon-Sorbonne », *Humanités numériques*, n° 1 (1 janvier 2020).

²⁹² Rey-Coyrehourcq, « Une plateforme intégrée pour la construction et l'évaluation de modèles de simulation en géographie », Annexes, 93.

²⁹³ Rey-Coyrehourcq, Annexes, 96.

malheureusement trop lent, faute de crédits de vacation en quantité suffisante »²⁹⁴. Contrairement à d'autres équipes, comme celle menée par Prost pour l'étude du vocabulaire des proclamations électorales, le cadre universitaire dans lequel évolue cette équipe jusqu'en 1977, date à laquelle elle obtient le statut d'ERA, n'est donc pas idéal pour avancer au rythme voulu.

Un deuxième problème résulte de cette configuration administrative. Ne pouvant être embauchés de manière pérenne, ni par l'université Paris 1, ni par le CNRS, Hucher et Mondelli ne participent à l'entreprise ALINE qu'à temps partiel. Hucher quittera même l'aventure en cours de route. Mondelli, quant à lui, mènera par la suite une carrière bien loin de l'université ou du CNRS, en devenant cadre de la société d'informatique française Bull. Les modalités administratives de cette collaboration pèsent donc de tout leur poids sur le rythme de l'élaboration du programme. En dépit de ces difficultés, les deux ingénieurs travailleront à la mise au point du programme entre 1972 et 1976. Cependant, cette configuration de travail aura plusieurs implications sur la nature du programme que Genet et Mondelli décident d'élaborer.

La première d'entre elles est que Genet comprend dès le départ que ALINE doit être, une fois élaboré, assez simple pour être utilisé sans avoir recours à un informaticien. En effet, en raison de la difficulté d'avoir accès aux informaticiens, Genet fait de l'indépendance des historiens pour l'utilisation de ALINE l'un des réquisits essentiels qui préside à son élaboration. Cette volonté est par ailleurs renforcée par une autre raison, avancée dans le rapport de 1983 pour le CNRS : le programme créé entre 1973 et 1976 est « *lourd de maniement* », ce qui conduit Genet à préciser la surcharge de travail qu'ils en coûtaient aux membres de l'ERA 713 :

*« nous ne pouvions pas nous transformer en prestataires de service pour les autres chercheurs désireux de l'utiliser au risque de devoir abandonner notre propre programme »*²⁹⁵.

²⁹⁴ Jean-Philippe Genet, « Rapport d'activité pour le CNRS » (Équipe de recherche associée 713, 1978).

²⁹⁵ Jean-Philippe Genet, « Rapport d'activité pour le CNRS » (Université Paris 1 Panthéon-Sorbonne, 1983).

La deuxième implication de la participation à court terme des deux informaticiens concerne la définition de l'un des objectifs de cette entreprise. Dans un rapport pour le CNRS de 1978, Genet le formule en ces termes :

*« présenter un instrument de travail qui soit simple et maniable, et qui n'exige pas le recours à un programmeur ou même à un opérateur »*²⁹⁶.

Cette situation va conduire les acteurs à reprendre la première version de ALINE pour la réécrire sous la forme d'un « package » de quatre modules (ALINE 0, ALINE 1, ALINE 2 et ALINE 3), destinés chacun à un type d'opérations (catalogage et diagnostic des erreurs, référencement et création de dictionnaires, concordances et création de liste). Cette restructuration du programme doit en permettre un maniement plus simple, et même l'implantation de certains de ses modules (ALINE 0 et ALINE 1) au CIRCE dès 1979.

La configuration administrative dans laquelle se trouve l'ERA 713 conduit donc les membres de l'équipe à envisager la constitution d'un instrument de travail « clef en main », tout à fait différent des programmes qui existent alors dans d'autres institutions, comme le CETEDOC ou le CRAL. Ces derniers nécessitent en effet, pour être mis en œuvre, l'aide d'un ingénieur pour la programmation d'une part, et pour l'utilisation d'autre part. Cette volonté s'inscrit dans une tendance plus générale des évolutions de ce qu'on appelle un programme dans les années 1970. Elle nous force à remettre cette notion familière dans son contexte.

1.2.2. ALINE : un programme informatique original.

Tout d'abord, il faut rappeler que l'interface privilégiée entre l'historien et l'ordinateur est, au début des années 1970, l'informaticien, ou l'opérateur du centre de calcul. Cette configuration de travail caractérise cette organisation sociale des calculs en histoire, et se base sur le modèle des centres de calcul interuniversitaires et du CNRS, qui implique que les utilisateurs se fassent accompagner par des ingénieurs dans leur utilisation d'équipement très cher et très fragile²⁹⁷.

²⁹⁶ Genet, « Rapport d'activité pour le CNRS », 1978.

²⁹⁷ Voir, par exemple, l'entretien suivant, mené auprès de l'ancien responsable de l'informatique au CNRS pour la période 1971-1975 : Wladimir Mercouff, « Témoignage : l'informatique dans l'enseignement et

Ensuite, à la différence des logiciels commerciaux que nous utilisons désormais quotidiennement, il est très compliqué d'échanger ces programmes entre les équipes. Les chercheurs qui désirent travailler avec des outils élaborés en dehors de leurs institutions doivent soit se rendre dans les laboratoires qui les ont développés²⁹⁸, et travailler au contact des informaticiens qui les ont développés²⁹⁹, soit profiter, si elle est effective, de l'implémentation de ces programmes informatiques dans les grands centres de calculs interuniversitaires comme le CIRCE.

Dans ce contexte, le projet de ALINE nous apparaît comme tout à fait original, puisqu'il n'existe pas, à notre connaissance, d'autres projets qui avaient pour but de réaliser un programme d'analyse automatique des textes utilisable directement par des historiens dans les années 1970 en France³⁰⁰. Pour récapituler, l'équipe de l'ERA 713 proposait donc explicitement un « *instrument de travail* » pour la communauté des médiévistes, instrument qui permettait d'identifier des occurrences, de dénombrer des formes et des lemmes, de calculer des fréquences et de générer des concordances quasi-automatiquement, sans avoir besoin de recourir à un spécialiste de l'informatique. Comme le note Genet dans son article de 1974³⁰¹, il suffisait alors d'avoir quelques rudiments du langage permettant la perforation des cartes et de savoir taper à la machine pour utiliser ALINE³⁰².

au CNRS au début des années soixante-dix. », *Histoire de la recherche contemporaine. La revue du Comité pour l'histoire du CNRS*, n° Tome IV-N°2 (15 décembre 2015): 180-85.

²⁹⁸ Deux éléments, développés dans deux autres chapitres de cette thèse, méritent ici notre attention. Le premier, c'est l'existence d'une rubrique intitulée « Voyage au centre de calcul » dans le bulletin de liaison *Le Médiéviste et l'ordinateur*, publié par l'IRHT à partir de 1979. Nous le verrons dans le quatrième chapitre de cette thèse, cette rubrique encourage les historiens à se rendre dans les centres de calculs de toute l'Europe pour travailler avec les programmes et les données d'autres équipes de recherches. Le second élément a été développé dans le chapitre précédent. Nous avons vu, sur la base de la relation épistolaire qu'entretenaient Christiane Klapisch-Zuber et David Herlihy pendant le projet d'édition du *catasto* florentin de 1427, que les boîtes de cartes perforées servant de support de stockage aux programmes comme aux données étaient le plus souvent envoyées par avion dans les bagages d'Emmanuel Le Roy-Ladurie, occasionnant de grands moments d'inquiétude pour les équipes concernées.

²⁹⁹ Sur ce point, il est intéressant de prendre en compte les rapports de force et les incompréhensions qui pouvaient exister dans les équipes entre statisticiens, linguistes et historiens, par exemple à Saint Cloud. Dans un entretien mené par nous auprès d'André Salem, linguiste spécialiste des statistiques textuelles, et membre du laboratoire de lexicologie politique de Saint-Cloud pendant les années 1970 et 1980, il évoque les tensions qui ont pu exister entre linguistes et mathématiciens, mais aussi entre mathématiciens et historiens, autour notamment, de questions d'autorité sur les connaissances en statistiques.

³⁰⁰ Au Canada, un projet de ce type est en cours de développement au début des années 1970 : c'est le logiciel JEUEMO, dont nous reparlerons notamment dans le chapitre 5 de cette thèse. Voir, par exemple « Un nouveau logiciel », *Le médiéviste et l'ordinateur* 5, n° 1 (1981): 22-23.

³⁰¹ Genet et al., « Un programme de traitement automatique des textes ».

³⁰² Il faut cependant indiquer ici qu'il nous est impossible, faute de sources à notre disposition, de préciser dans quelle mesure ces compétences étaient répandues à l'époque. Cette question, qui rejoint celles sur les

Ces éléments mettent en lumière l'une des problématiques qui anime la communauté des historiens utilisant les outils informatiques dans les années 1960-1970, à savoir la question du degré de dépendance technique et matérielle qu'il faut entretenir à l'égard de ces outils et de ceux qui les conçoivent. La position défendue à travers la conception de ALINE est en ce sens très avant-gardiste pour la période. Elle se situe à mi-chemin entre une dépendance complète aux informaticiens, comme nous l'avons montré pour le cas de l'organisation du travail de l'enquête du *catasto*, et l'approche d'un autre médiéviste français de la même période, Alain Guerreau (1948-), qui plaidait, dès le milieu des années 1970, pour une indépendance matérielle et technique totale. Guerreau défendait notamment la nécessité d'une formation solide à la programmation pour les historiens et l'utilisation d'une calculatrice programmable de poche³⁰³.

Une autre originalité de ce programme était aussi de se vouloir adaptable à d'autres environnements informatiques et à d'autres données. Au contraire de programmes conçus dans d'autres contextes à la même période, comme celui de Genicot et Tombeur au CETEDOC, ou celui élaboré par Demonet dans le cadre de l'étude sur les proclamations électorales de Prost, ALINE était un programme potentiellement compatible avec d'autres environnement de travail que celui dans lequel il avait été créé.

Premièrement, le programme était utilisable sur d'autres machines, moins puissantes que celles du CIRCE (mais pour des textes plus courts, de moins de 5000 mots), en raison du choix faits par les programmeurs d'utiliser le FORTRAN, qui, s'il n'était pas idéal pour le traitement des chaînes de caractères, permettait une portabilité plus large. Deuxièmement, il pouvait s'adapter à différents types de données textuelles,

usages des ordinateurs pour les opérations de rédaction, constitue l'un des axes de recherche que nous aimerions poursuivre après cette thèse. Il faut toutefois préciser que pour la perforation et la programmation, les chercheurs en sciences humaines et sociales pouvaient bénéficier, à partir de sa création en 1969 par le CNRS, du Service de Calcul Science Humaines, devenu le LISH en 1976. On y trouvait de nombreux ingénieurs, dont certains étaient même membre d'autre laboratoire de recherche, comme l'Institut de Recherche et d'Histoire des Textes.

³⁰³ Il faut noter ici que Guerreau a toujours refusé de travailler dans les centres de calcul, et qu'il se servait à l'époque d'une calculette programmable de poche Texas Instrument, qui lui assurait une indépendance technique et matérielle presque totale. À l'aide de cet instrument, il réalisait notamment des analyses statistiques classiques et des analyses factorielles des correspondances. Nous consacrons justement le sixième chapitre de cette thèse à l'une de ces applications. Pour les plus curieux, on trouvera un exemple de l'un de ses programmes dans le numéro 2 du *Médiéviste et l'ordinateur* : Alain Guerreau, « Analyse factorielle au moyen d'une calculatrice de poche programmable », *Le médiéviste et l'ordinateur* 2, n° 1 (1979): 10-12. On découvrira encore plus largement, son diagnostic sur les relations entre informatique et histoire au début des années 1980 dans la première partie de l'article suivant, que nous analyserons dans le chapitre 5 de cette thèse : Alain Guerreau, « Analyse factorielle et analyses statistiques classiques : le cas des ordres mendiants dans la France médiévale », *Annales* 36, n° 5 (1981): 869-912.

et à différentes langues, puisqu'il offrait la possibilité d'indiquer pour chaque traitement envisagé les caractères séparateurs choisis, et, donc, d'analyser un texte, même s'il n'avait pas été prévu initialement pour être traité par ce programme.

C'était donc une vision très neuve de ce qu'est un programme informatique pour les chercheurs en sciences humaines qui était défendue par Genet, Hucher et Mondelli en 1974 : celle d'un instrument de travail utilisable et réutilisable par (presque) tous. Cette vision s'est d'ailleurs imposée plus largement avec la révolution micro-informatique, comme l'illustre d'autres logiciels partageant le même type de fonctionnalités qui apparaîtront à la fin des années 1980 dans le contexte français, comme Hyperbase³⁰⁴, Alceste³⁰⁵ ou Lexico³⁰⁶.

Malheureusement, malgré ces deux caractéristiques, on ne retrouve ni dans les archives ni dans la bibliographie associée à ALINE d'utilisation avérée de ce programme par des membres extérieurs au laboratoire. Finalement, ALINE est resté « *d'un maniement assez délicat* », comme on peut le lire dans le rapport des membres de l'ERA 713 pour le CNRS de 1978³⁰⁷. Nous reviendrons plus longuement sur ce point dans la quatrième partie de ce chapitre.

1.3. Contraintes économiques et contraintes de temps.

L'analyse du second objectif du programme ALINE, formulé par Genet en 1978, nous permet de repérer un autre aspect de l'organisation collective de l'ERA 713. Pour rappel, dans le rapport au CNRS de 1978 précédemment cité, Genet affirme en effet que l'objectif de l'entreprise est de

*« réaliser un programme d'utilisation aussi économique que possible »*³⁰⁸.

³⁰⁴ Voir Etienne Brunet, « Hyperbase: logiciel documentaire et statistique pour l'exploitation des grands corpus », *Tools for humanists*, 1989, 33-36.

³⁰⁵ Voir Max Reinert, « Un logiciel d'analyse lexicale », *Cahiers de l'analyse des données* 11, n° 4 (1986): 471-81.

³⁰⁶ Lexico est un logiciel gratuit, développé par André Salem pendant son doctorat d'État, soutenu en 1993. On pourra notamment en trouver un compte rendu ici : Annie Geoffroy, « André Salem, « Méthodes de la statistique textuelle », thèse de doctorat d'État, Université Paris 3, 1993 », *Mots. Les langages du politique* 38, n° 1 (1994): 104-8.

³⁰⁷ Genet, « Rapport d'activité pour le CNRS », 1978.

³⁰⁸ Genet.

Pour comprendre le sens de cette affirmation, il faut rappeler ici que les entreprises d'analyse automatique des textes sont, encore plus qu'aujourd'hui, des activités extrêmement chronophages pendant les années 1970. Elles le sont notamment parce qu'elles consistent en réalité en deux activités distinctes : la réalisation du programme lui-même et la saisie sur cartes perforées des textes qui seront objets des analyses automatiques.

1.3.1. Rythme et coût de la perforation des cartes.

Si nous avons vu les modalités de l'organisation du travail collectif pour la première de ces deux activités, qu'en est-il pour la seconde ? Dans le cadre de ALINE, le choix a été fait de laisser la possibilité aux historiens d'effectuer eux-mêmes la perforation sur l'une des dix machines-perforatrices du centre de calcul de Paris 1. Ce choix était corrélé aux langues des textes étudiés, que nous avons présentées plus haut. En effet, ni le personnel des entreprises de perforation privées, dont les services coûtaient cher, ni les deux opératrices-perforatrices du centre de calcul de Paris 1 n'étaient capables de lire des textes en moyen-français ou en moyen-anglais.

Le rapport d'activité de l'ERA 713 à destination du CNRS de 1981 signale que les textes sont perforés soit par l'historien lui-même, soit par « *un étudiant connaissant bien la langue du document* »³⁰⁹. La rétribution prévue était alors de 1000 francs pour 1500 cartes (soit l'équivalent d'environ 200 euros)³¹⁰, chaque carte permettant d'enregistrer 70 caractères alphanumériques. Ce mode de fonctionnement était aussi motivé par des raisons qui tenaient à la volonté de gagner du temps dans le processus d'édition des textes - sur lequel nous reviendrons longuement dans la deuxième partie de ce chapitre - puisqu'une étape de la saisie disparaissait : le moment du remplissage des bordereaux de perforations qui étaient habituellement utilisés pour transmettre les données à l'entreprise de perforation.

Le rapport de 1981 nous renseigne sur les limites de cette organisation pour la perforation des cartes. Tout d'abord, les crédits de vacations étaient insuffisants (seulement 4 000 francs pour trois ans – soit l'équivalent d'environ 800 euros -, soit

³⁰⁹ Jean-Philippe Genet, « Rapport d'activité pour le CNRS » (Équipe de recherche associée 713, 1981).

³¹⁰ La conversion a été faite au taux de change de l'année 1973, via le site internet fxtop.com. Les conversions qui suivent ont été obtenues à l'aide du même outil.

environ 420 000 caractères). Ensuite, la méthode elle-même était trop lente, puisque la moindre erreur nécessitait une nouvelle perforation de la carte en question, et parfois même des cartes suivantes. Ces opérations de perforation et de correction étaient par ailleurs rendues d'autant plus compliquées que le rapport d'activité indique que les machines-perforatrices IBM 029 du centre de calcul de Paris 1 étaient alors dans « *un état déplorable* »³¹¹.

Enfin, les opérations de vérification des données prenaient beaucoup de temps. Millet, engagée à mi-temps comme I.T.A. par l'ERA 713 à partir de 1979, prenait en charge une partie de ces opérations. Elle devait cependant participer prioritairement, du fait de ses compétences, aux autres projets menés au sein de l'équipe, et plus particulièrement aux enquêtes prosopographiques, ce qui conduisait les membres de l'ERA 713 à multiplier les demandes au CNRS pour l'augmentation des crédits vacances et le recrutement d'une autre personne à mi-temps.

1.3.2. Solutions techniques.

Dans ce même rapport de 1981, des solutions techniques étaient encore envisagées pour résoudre ce problème. D'une part, l'acquisition d'un micro-ordinateur Apple II, qui devait permettre de saisir les données directement sur disquettes, sans passer par le support de stockage que sont les cartes perforées. D'autre part, l'achat d'un lecteur optique Kurzweil Data Entry Machine, qui devait rendre possible ce qu'on appelle aujourd'hui l'océrisation³¹² des textes de façon semi-automatique.

Un autre type de contraintes est mentionné par les acteurs dans les sources dont nous disposons. Elles portent sur le temps de calcul que nécessitaient les opérations d'analyse automatique dans les centres de calculs. En effet, selon des modalités qu'il ne nous pas été possible de saisir sur la base des sources à notre disposition, les calculs étaient partagés entre le centre de calcul de Paris 1 et le CIRCE. Au CIRCE, chaque heure de calcul était facturée et payée par des crédits de calcul octroyés par le CNRS lui-même. Cette politique du temps de calcul voyait donc le CNRS injecter des crédits dans les laboratoires qui lui revenaient directement via le CIRCE. Cependant ces crédits-calculs n'étaient pas en nombre illimité pour les membres de l'ERA 713. Mieux le programme

³¹¹ Genet, « Rapport d'activité pour le CNRS », 1981.

³¹² Terme construit sur la base de l'acronyme OCR, pour Optical Character Recognition.

était écrit, et moins long serait le temps de calcul. Un autre point clef était encore la possibilité d'économiser des frais d'impression, les études de textes étant, comme le rappelait Genet dans un article de 1977, particulièrement consommatrices de papier³¹³.

Ces considérations expliquent qu'une exigence d'économie pèse sur la phase de rédaction du programme. Elle conditionnera, entre autres choses, les possibilités d'utilisation de ALINE par les membres de l'ERA 713 comme par des utilisateurs extérieurs désireux d'utiliser ce programme ou les données qui lui sont attachées. Par exemple, l'une des solutions proposées par les programmeurs de ALINE était de concevoir le programme de telle façon :

*« qu'il [soit] possible en fonction du temps machine disponible de demander les dictionnaires sous trois formes différentes : forme brève, forme élaborée, forme hiérarchisée »*³¹⁴.

Cela permettait au chercheur-utilisateur de conformer son utilisation au crédit de calcul qui lui restait, tout en évitant d'imprimer plus de résultats que nécessaire. Le programme était ainsi plus économique en temps, en argent et en papier.

Comme nous l'avons vu, les objectifs assignés à ALINE comme les solutions choisies pour les atteindre nous renseignent sur les contraintes, nombreuses, qui pouvaient exister au sein d'un projet de réalisation d'un programme informatique dans cet environnement matériel si particulier, celui des « gros systèmes » informatiques fonctionnant à l'aide de cartes perforées. Ces contraintes poussaient les acteurs à rechercher des solutions innovantes quant à la nature du programme qu'ils produisaient. Malheureusement, les années 1970 sont aussi le moment d'une révolution du matériel informatique avec l'arrivée des micro-ordinateurs.

Dès le début des années 80, les concepteurs de ALINE se retrouvent face à la nécessité d'adapter leur programme à un nouvel environnement matériel. Ce ne sera finalement jamais le cas. Tout d'abord, les informaticiens partenaires du projet n'ont plus le temps d'aider à cette transformation. Ensuite, les programmes sont désormais plus facilement échangeables. Un autre logiciel, JEUEMO, venu du Canada, et développé

³¹³ Genet, « Histoire sociale et ordinateur », 237.

³¹⁴ « Programme de traitement automatique des textes pour l'équipe de recherche sur l'histoire politique du bas Moyen Age », document interne à l'ERA 713, non daté.

par Paul Bratley, Francine Ouellette et Serge Lusignan dès 1974³¹⁵, a par exemple été rapidement adapté pour certains micro-ordinateurs, et rendait possible à peu près les mêmes opérations que celles rendues possibles par ALINE³¹⁶.

Le développement du programme ALINE s'arrête ainsi autour de 1984, après une petite dizaine d'années de développement, rattrapé par l'évolution rapide des technologies informatiques et l'indisponibilité de Hucher et Mondelli. Comme on a pu le voir, le contexte matériel de fabrication du programme pesait fortement sur sa conception, comme sur les solutions choisies pour sa réalisation. Comme nous allons le voir, ce contexte matériel pesait aussi fortement sur les analyses rendues possibles par ALINE et sur les modalités de la constitution des corpus et de l'édition des textes.

2. Constitution des corpus et éditions des textes.

Le premier ensemble d'opérations sur lequel va porter notre analyse ne concerne pas directement la production du programme ALINE, mais le processus d'édition des données textuelles sur des cartes perforées IBM à 80 colonnes. Ces opérations demeurent cependant au cœur de la logique du programme, puisque, rappelons-le, ce sont justement la nature des textes médiévaux étudiés qui conduisent Genet, Hucher et Mondelli à produire un programme original. En suivant, pas à pas, les opérations qui conduisent du choix des textes à leur édition sur cartes perforées, il sera possible de saisir les particularités de la méthode de traitement de texte mise en place pour l'édition des données, par contraste avec les pratiques analysées dans le chapitre précédent. Il nous sera encore possible de comprendre comment les médiévistes de l'ERA 713 s'inspirent des méthodes d'autres disciplines en les adaptant aux problèmes et aux questions qu'ils se posent à propos de leurs sources.

Selon quels critères les textes sont-ils choisis ? Comment ces médiévistes constituent-ils les corpus de textes qu'ils entendent étudier ? Que nous apprennent ces logiques de constitution sur les influences extra-disciplinaires des leurs méthodes ?

³¹⁵ Voir, par exemple Paul Bratley, Serge Lusignan, et Francine Ouelette, *JEUDEMO: a text-handling system* (University Press, 1974).

³¹⁶ Cette indication nous a été donnée par Genet, dans un entretien que nous avons mené auprès de lui en 2017.

Quelles sont les méthodes mises en œuvre pour les éditer sur cartes perforées ? Et quelle *modélisation* de ces textes implique leur édition sur ces cartes perforées ?

Pour répondre à ces questions, qui structureront notre analyse, nous nous appuierons sur un ensemble de documents composé d'une série d'articles publiés par Genet, Hucher et Mondelli, de deux documents internes à l'ERA 713 et d'un rapport d'activité pour le CNRS de 1978, déjà mentionnés. Ces sources ne nous permettront de nous pencher que sur les corpus de textes constitués par Genet, en raison de l'absence d'une documentation suffisante pour les autres projets d'application de ALINE. Nous mènerons par ailleurs, au fil de l'analyse, une comparaison avec le cas de l'enquête du *catasto* étudié dans le chapitre précédent, afin de saisir comment le recours aux ordinateurs conduit les médiévistes à développer de nouvelles méthodes d'édition, de nouvelles formes de modélisation des textes et de nouvelles conceptions de ce qu'est un corpus au milieu des années 1970.

2.1. Modalités de fabrication des corpus.

La première série d'opérations que pratiquent les historiens de l'ERA 713 pour conduire cette enquête sur les mentalités politiques consiste à délimiter un ensemble de textes possiblement exploitable à l'aide d'un programme de traitement automatique des textes. Cet ensemble d'opérations nous intéressent particulièrement puisqu'il nous permet de saisir comment se développe, nous le verrons, une nouvelle manière de constituer des corpus en histoire médiévale.

2.1.1. Qu'est-ce qu'un texte politique ?

Trois éléments nous autorisent à comprendre les logiques de constitution des corpus de textes de Genet pour leur étude réalisée à l'aide de ALINE. Le premier élément tient à la construction de l'objet d'étude de Genet et s'articule avec les questions suivantes : quels textes choisir pour étudier les idées politiques à la fin du Moyen-Âge ? Et, en fonction des textes choisis, qu'est-il possible d'étudier ? Nous nous concentrerons tout d'abord sur ce premier élément, et reviendrons sur les deuxième et troisième éléments dans la seconde sous-partie de cette partie 2.

Pour comprendre les difficultés que pose la définition de ce qu'est un texte politique, il nous faut faire un pas de côté par rapport à notre problématique et nous plonger dans ce qui constitue l'objet de l'étude de Genet. Nous l'avons dit, Genet reprend, pour étudier les textes politiques anglais du bas Moyen-Âge, la notion de *champ* de Bourdieu. Cette approche méthodologique implique la mise en œuvre d'une démarche que Genet définit en ces termes dans l'introduction de sa thèse d'État, publiée en 1997 :

« Nous allons donc rechercher le ou les champs où les idées sociales et politiques sont reconnues ou reconnaissables pour telles. Une fois que nous les aurons reconnus, nous y observerons d'une part les textes, puisque c'est leur mosaïque qui va nous permettre de percevoir le champ, et de l'autre les agents (les auteurs ou assimilés, mais aussi les destinataires), [...] ; enfin il sera possible d'entreprendre l'étude lexicale une fois les textes structurés en corpus »³¹⁷.

Si cette citation peut sans doute procéder d'une reconstitution de la démarche suivie *a posteriori*, plus qu'elle ne décrit la procédure effectivement suivie par Genet, elle explicite toutefois que les critères de choix des textes sont intimement liés au processus de construction des objets de recherche du médiéviste. Le critère minimal de sélection des sources défini ici par Genet consiste en l'ensemble des textes des *champs* dans lesquels les idées sociales ou politiques sont reconnues (par les acteurs) ou reconnaissables (par les médiévistes).

En proposant une telle approche, Genet étend en fait considérablement le corpus des textes dit politiques de la fin du Moyen-Âge, alors presque exclusivement basé sur les miroirs au prince³¹⁸. Cette démarche permet de ne plus se limiter à une histoire des idées politiques des élites dirigeantes, mais vise à accéder aux représentations de la plus grande diversité des milieux sociaux. Au terme de son travail, les dimensions du corpus seront d'ailleurs incroyablement vastes, puisque cette démarche conduit Genet à identifier 2 135 auteurs et plus de 35 000 textes.

Face à un tel corpus, Genet propose donc de redécouper cet ensemble d'écrits en plusieurs *sous-champs*. A ce premier critère de sélection, il faut en ajouter un second,

³¹⁷ Genet, « Les idées sociales et politiques en Angleterre », 233.

³¹⁸ Genet, 261.

défini plus loin dans l'introduction de la thèse de Genet. Après avoir dégagé l'existence de trois *sous-champs* (religieux, institutionnalisé et littéraire), au terme d'une démarche sur laquelle nous ne reviendrons pas ici, Genet présente deux autres *sous-champs* qu'il qualifie de transversaux (l'« *historique* » et le « *politique* »). Ces deux *sous-champs* intéressent particulièrement Genet puisqu'il croit possible, au regard de la nature des textes qui les composent, d'en « *mesurer le lexique* » pour comparer les textes entre eux, ce qui, nous y reviendrons, constitue pour lui « *l'objectif ultime* » de sa recherche, qui vise à repérer la structure de ces deux *sous-champs*³¹⁹.

Pour identifier les écrits qui composent ces deux derniers *sous-champs*, il propose de prendre pour base un « *genre « initial »* », qu'il qualifie de « *matrice du champ transversal* »³²⁰. C'est de cette matrice que vont naître d'autres types de textes, par hybridation et évolution successives. Pour le « *politique* », qui nous intéresse particulièrement, puisque Genet développe ALINE dans le but d'étudier un corpus de sources relevant de ce *sous-champ*, le médiéviste prend pour « *genre initial* » le miroir au prince, qui s'incarne véritablement, selon lui, dans un seul texte majeur : le *De Regimine Principum*, écrit par Gilles de Rome vers la fin du XIII^e siècle, et dont les médiévistes ont découvert plus de 250 manuscrits dans toute l'Europe.

Pour poursuivre cette démarche, il se propose de reprendre la liste de miroirs aux princes dressée par le médiéviste allemand Wilhelm Berges (1909-1978) en 1938³²¹. Un problème apparaît alors aux yeux de Genet : la définition du genre miroir aux princes que propose Berges est à la fois trop restrictive et pas assez précise. Il en va de même, selon lui, de celles avancées par d'autres médiévistes, comme Lester Kruger Born (1903-1969), Arpad Steiner ou Dora Bell³²².

Étant données les difficultés que soulève la classification des textes politiques, Genet suggère d'écarter le critère du genre et de définir d'autres points communs entre les textes pour les rassembler en corpus. Au cours des premières années du projet ALINE, entre 1970 et 1978, Genet constitue plusieurs corpus, dont il est possible d'éclairer les

³¹⁹ Genet, 262.

³²⁰ Genet, 263.

³²¹ Voir Wilhelm Berges, *Die Fürstenspiegel des hohen und späten Mittelalters* (Leipzig, Allemagne, 1938).

³²² Sur cette question, voir notamment un article récent qui reprend cette problématique : Einar Már Jónsson, « Les « miroirs aux princes » sont-ils un genre littéraire ? », *Médiévales. Langues, Textes, Histoire*, n° 51 (1 décembre 2006): 153-66.

logiques de constitutions à l'aide deux documents déjà mentionnés. Le premier est l'une des appendices du document interne à l'ERA 713, intitulé « Le choix des textes anglais »³²³. Le second est le rapport pour le CNRS de 1978 déjà présenté³²⁴. Comme le premier de ces deux documents n'est pas daté précisément, mais que nous avons pu établir qu'il avait été rédigé avant le second, il nous est possible d'étudier certaines évolutions des pratiques de sélection et de rassemblement des textes en fonction des périodes.

2.1.2. Logiques de constitution des corpus.

Le premier critère mis en avant par Genet pour l'un de ces corpus, qu'il qualifie de « Premier corpus anglais », concerne les thématiques abordées dans les textes, à savoir, un « *sujet suffisamment reconnaissable* », celui de la « *théorie* » politique, et ce « *quelle que soit la forme qu' [ils] revêtent* »³²⁵. On y trouve ainsi des traités, une lettre, la traduction d'un miroir au prince français, un sermon ou encore la traduction d'une « *sorte d'encyclopédie* »³²⁶. Genet construit donc l'un de ses corpus en suivant une logique qui ne repose ni sur le *genre* des textes, ni sur leurs cohérences archivistiques, mais sur leurs sujets.

Cependant, le fait de partager un même sujet n'est pas le seul trait qui rend ces textes comparables pour Genet, puisqu'un autre de leurs points communs est d'être écrit en anglais. La langue des textes choisis reflète l'intention d'examiner certaines problématiques, puisque l'anglais se manifeste en Angleterre à la fin du Moyen-Âge comme une langue de vulgarisation des idées scholastiques et théologique en direction des élites politiques, par contraste avec le français, langue technique du *champ* que Genet qualifie d' « *institutionnalisé* » – qui comprend notamment les textes de droit -, et le latin, langue du *champ* « *religieux* ». Ce critère nous renseigne sur le fait que la cohérence du corpus repose non seulement sur la finalité des textes étudiés, mais également sur leurs conditions de production et les publics auxquels ils pouvaient être destinés.

³²³ « Programme de traitement automatique des textes pour l'équipe de recherche sur l'histoire politique du bas Moyen Age », document interne à l'ERA 713, non daté.

³²⁴ Genet, « Rapport d'activité pour le CNRS », 1978.

³²⁵ « Programme de traitement automatique des textes pour l'équipe de recherche sur l'histoire politique du bas Moyen Age », document interne à l'ERA 713, non daté.

³²⁶ « Programme de traitement automatique des textes pour l'équipe de recherche sur l'histoire politique du bas Moyen Age », document interne à l'ERA 713, non daté.

Un troisième élément important éclaire les logiques de constitution de ce corpus. Genet y rassemble des textes écrits entre 1380 et 1509, et élabore ainsi une approche diachronique de l'étude du vocabulaire politique sur la longue durée.

Un quatrième élément est enfin absolument central pour comprendre comment Genet construit ce premier corpus. Il s'agit de la finalité de cette opération. Dans une note de bas de page d'un article de 1974, il affirme qu'il s'intéresse particulièrement à « *la lexicographie pure et à tout ce qui favorise les comparaisons de texte à texte* »³²⁷. Le raisonnement qui se dégage est donc le suivant : il s'agit de faire apparaître des différences lexicales entre ces textes, au sein d'un corpus rendu cohérent par les thématiques communes et l'utilisation de la langue anglaise.

Cependant, le rapport au CNRS de 1978 révèle que la conception du corpus paraît avoir connu un infléchissement. En effet, le corpus évoqué ci-dessus, dont la mise au point doit être antérieure à 1978, paraît avoir été ultérieurement complété par un second ensemble de textes. Par contraste avec le fait que le premier groupe rassemblait des écrits « *théoriques* », le second est appelé corpus de « *textes-programmes* », ou textes « *pratiques* ». Selon les mots de Genet, ce second corpus doit permettre « *de comparer vocabulaire théorique et vocabulaire pratique* »³²⁸. On constate ici que le raisonnement a évolué pendant la période de travail qui nous intéresse : en plus de la comparaison texte à texte au sein d'un corpus, Genet cherche aussi à comparer les vocabulaires corpus à corpus, pour identifier les différences qui existeraient dans ces deux types de discours. Le corpus devient ainsi non seulement un objet au sein duquel on compare, mais encore une unité d'observation qui peut elle-même être comparée avec d'autres.

Il faut toutefois ajouter, à ces critères intellectuels, d'autres, qui concernent le fait que le contexte matériel de l'étude, analysé dans la première partie de ce chapitre, joue sur le choix des textes. Dans le rapport de 1978, Genet présente par exemple l'un de ses corpus sous la forme d'une liste de seize textes composés en langue anglaise. En commentant cette liste, il affirme que ce corpus est constitué de « *pratiquement tous les textes « théoriques »* »³²⁹ qu'il souhaite perforer. Il ajoute cependant :

³²⁷ Genet et al., « Un programme de traitement automatique des textes », 99.

³²⁸ Genet, « Rapport d'activité pour le CNRS », 1978.

³²⁹ Genet.

« il serait trop long et trop onéreux d'inclure dans ce travail les traductions anglaises de Gilles de Rome, d'Alain Chartier et de Christine de Pisan, dont certaines sont encore inédites »³³⁰.

Ainsi, deux considérations de natures différentes entrent dans le choix des textes à enregistrer. La première concerne le coût, en temps comme en argent, de l'enregistrement sur cartes perforées des textes cités. Les textes mentionnés dans cette citation sont en effet plus longs que ceux choisis par Genet pour son corpus de textes théoriques, qui comprend, par exemple, le *Tractacus Regibus*, d'environ 34 000 mots, ou le traité *The III Considerations*, qui compte moins de 62 000 mots. Au contraire, les traductions de Gilles de Rome, de Christine de Pisan et d'Alain Chartier sont beaucoup plus longues. Le *De Regimine Principum* de Gilles de Rome compte par exemple plus de 400 pages, soit, environ, 150 000 mots³³¹.

Les éléments dégagés dans la première partie de ce chapitre éclairent les raisons qui motivent cette décision. Le manque de crédits de vacations permettant d'employer des étudiants à même de réaliser la perforation des textes ne permet pas aux médiévistes de l'ERA 713 de se lancer dans l'édition sur cartes perforées de textes aux dimensions semblables à celle du *De Regimine Principum* de Gilles de Rome. La brièveté des textes est donc, au regard de la configuration matérielle et des méthodes d'enregistrement sur cartes perforées choisies, l'une des critères de sélection des textes qui composeront les corpus. L'autre considération dont Genet tient compte, et dont notre source se fait l'écho, est le fait que les médiévistes doivent se baser sur des textes édités, ce qui évitent aux membres de ce collectif de multiplier les tâches de préparation des textes.

Au final, la thèse d'État de 1997 fait mention de l'existence de quatre types de corpus édités par Genet au cours de cette longue période : deux types dits « pratiques », qui rassemblent à la fois les textes théoriques et les textes pratiques, mentionnés dans les deux documents cités ci-dessus, un corpus poétique et un corpus en prose. Par ailleurs, il faut préciser que certains textes appartiennent à la fois à différents types de corpus, ce qui

³³⁰ Genet.

³³¹ Pour établir ce chiffre, nous avons ici choisi la valeur suivante, en nous basant sur un recueil d'édition de traités politiques anglais publié par Genet en 1977 : 375 mots par page. Cette méthode est cependant tout à fait critiquable, puisqu'elle ne se base ni sur une version originale ni sur une version éditée du miroir au prince de Gilles de Rome. Elle a toutefois le mérite de donner une idée de la différence de longueur entre ces textes.

renforce l'idée que ces opérations de constitution des corpus sont corrélées à des questions de recherches. Les corpus de Genet sont ainsi, au contraire de ce que nous avons montré pour l'édition du *catasto*, des objets scientifiques qui se recomposent au fil des problématiques de recherche. Nous reviendrons plus longuement, dans la troisième partie de ce chapitre, sur ces relations entre corpus et raisonnements pour éclairer comment Genet utilise ALINE, mais il nous faut ici insister sur ce que nous apprend cet exemple sur l'évolution des méthodes de collecte des données en histoire.

2.2. Influences extra-disciplinaires.

Que nous disent les modalités de fabrication des corpus sur le positionnement théorique de Genet ? Nous pouvons, pour examiner cette question, comparer les procédures qu'il met en œuvre avec celles qui existent alors en histoire médiévale et en linguistique. Nous verrons, dans un premier temps, les différences qui tiennent à la logique de constitution des corpus puis, dans un second, celles qui tiennent à la nature des données extraites des sources (voir 2.3.).

2.2.1. Observations sur les différences entre corpus d'historiens et corpus de linguistes.

Commençons par une comparaison avec la pratique d'élaboration du corpus étudiée dans le chapitre 1. Par contraste avec l'enquête sur le *catasto*, les médiévistes de l'ERA 713 ne fondent pas leur étude sur une source homogène d'un point de vue archivistique, mais sur des ensembles de textes disparates, qu'ils regroupent en corpus, en fonction des questions qu'ils entendent traiter.

Cette différence signale, selon nous, une forme d'évolution, induite par les méthodes linguistiques automatisées choisies par Genet, sur le rôle du corpus dans la méthode historique. Dans le compte-rendu d'une table ronde de 2016, Eliana Magnani soulève cette question en proposant un bref retour sur l'histoire du terme corpus, qui permet d'éclairer cette évolution. Au XIX^e siècle, elle montre qu'il existe dans les dictionnaires deux sens bien distincts du mot. Le premier, notamment attesté dans un dictionnaire de 1826, vient des sciences juridiques. Le *corpus juris* y est défini comme « une collection de droit romain », et le sens de ce terme s'étend progressivement vers l'idée d'une collection d'éléments de natures diverses, rassemblés dans le cadre d'une

étude. Un autre sens de ce mot est cependant attesté en Allemagne au milieu du XIX^e siècle, et est défini dans un dictionnaire de 1855 par son emploi dans les sciences linguistiques : « *ensemble de textes établis selon un critère thématique en vue de leur étude linguistique* ».

Dans ce même article, Magnani dresse le constat suivant quant à l'utilisation du terme par les médiévistes :

« [La journée d'étude] avait pour point de départ le fait que, depuis les années 1980, mais surtout les années 2000, l'usage du mot « corpus » par les historiens médiévistes s'est largement diffusé, sans que cet emploi généralisé ait été accompagné d'une quelconque réflexion sur ce qui se trouve derrière le terme, et encore moins d'un effort de conceptualisation. S'inscrivant en parallèle avec l'informatisation et la recrudescence des numérisations et des bases de données de documents médiévaux, cette situation contraste avec d'autres disciplines des sciences humaines et sociales, à commencer par la linguistique, qui depuis longtemps réfléchissent et débattent sur cette notion. »³³²

Ce bilan est complété par une étude de 127 articles de médiévistes, dans lequel elle souligne une majorité d'usages « lâche[s] » du terme, sur le modèle du corpus au sens juridique, loin du caractère « construit et délimité »³³³ qui existe chez les linguistes.

Cette remarque, formulée par une médiéviste au XXI^e siècle, fait apparaître le caractère singulier, voire avant-gardiste, de la façon dont Genet construit ses corpus. Il semble en effet, à la lumière des considérations précédentes, que sa démarche est plus proche des méthodes de certains domaines de la linguistique que de celle des historiens. Pour prendre pleinement la mesure de cette dernière affirmation, il nous faut examiner comment la démarche de Genet s'inscrit dans l'évolution de l'usage des corpus par les linguistes.

³³² Eliana Magnani, *Qu'est-ce qu'un corpus ? Compte-rendu de la journée d'études* (IRHT, 2017), 1.

³³³ Magnani, 3.

2.2.2. Fonction épistémologique des corpus constitués par Genet.

Le premier élément en ce sens concerne le statut épistémologique des corpus dans la démarche d'analyse de la langue. Au milieu des années 1950, John Rupert Firth (1890-1960) et la *London School* ont contribué à la naissance de la linguistique de corpus, en postulant que le sens d'un mot n'est pas donné *a priori* mais se situe dans ses usages avérés en contexte, d'où la nécessité de réunir des corpus de ses occurrences³³⁴. Cette approche des phénomènes linguistiques par la construction de corpus ne s'est cependant pas limitée à cette école londonienne. Les linguistes se rattachant à l'analyse de discours, domaine qui se développe notamment aux États-Unis dans les années 1960, puis en France dans les années 1970³³⁵, se sont, eux aussi, emparés de cette approche.

Dans son habilitation à diriger des recherches, soutenue en 2010, le linguiste français Damon Mayaffre revient sur la fonction du corpus en analyse de discours. S'inscrivant dans la continuité des travaux du laboratoire de lexicologie politique de Saint-Cloud, que nous avons présenté en introduction, Mayaffre définit le corpus comme

« un lieu construit où s'échafaude le sens, où se scénarise l'interprétation »³³⁶.

Il présente ainsi le corpus, pour les linguistes, comme la « *matrice du sens* »³³⁷ des phénomènes linguistiques. Par ailleurs, Mayaffre définit une autre fonction épistémologique des corpus, qu'il qualifie de « *forme maximale du contexte* », c'est-à-dire, comme l'échelle la plus grande à laquelle il est possible, de manière formalisée, d'étudier les phénomènes linguistiques³³⁸.

La démarche de constitution des corpus de Genet, que nous avons décrite au 2.1., présente ainsi plusieurs points communs avec les fonctions des corpus définies par Mayaffre en 2010. Tout d'abord, Genet construit ses corpus en fonction de ses questions

³³⁴ Voir Jacqueline Léon, « Aux sources de la « Corpus Linguistics » : Firth et la London School », *Langages* n° 171, n° 3 (2008): 12-33.

³³⁵ Cet ensemble de méthodes est introduit en France par la traduction des travaux du linguiste américain Zellig Harris (1909-1992) par Jean Dubois (1920-2015) en 1969. On se reportera ici au numéro 13 de la revue *Langages* de 1969, qui constitue le point de départ de cette histoire. Pour une approche d'observateur voir Jacqueline Léon, « Jean Dubois, un passeur », *Linx*, n° 80 (juillet 2020).

³³⁶ Damon Mayaffre, « Vers une herméneutique matérielle numérique. Corpus textuels, Logométrie et Langage politique » (Habilitation à diriger des recherches, Université Nice Sophia Antipolis, 2010), 12.

³³⁷ Mayaffre, 12.

³³⁸ Mayaffre, 13.

de recherche, et ils constituent, à ce titre, des matrices pour mener les études envisagées. Ensuite, nous l'avons vu, Genet exclut les critères plus traditionnels de classification des textes, et notamment le genre. Il fait ainsi de ses corpus « *la forme maximale du contexte* », puisque c'est au sein de ceux-ci qu'il propose de reconstruire une classification plus homogène. Cet aspect de la démarche de Genet doit, par ailleurs, être mis en relation avec la volonté de réaliser des analyses statistiques, qui ne peuvent être mises en œuvre que dans un espace de comparaison formalisé.

Pour compléter cette liste de points communs avec les pratiques de l'analyse de discours qui se développent dans les milieux académiques français depuis la fin des années 1960, on ajoutera que la nature des textes que Genet étudie rejoint celle des écrits qui intéressent ces linguistes : ce sont des textes politiques, autrefois exclus du domaine des études linguistiques³³⁹. Ce type de textes s'apparente également à ceux étudiés dans un autre domaine de la linguistique, que Genet fréquente particulièrement au début des années 1970. Dans un entretien mené en 2017, Genet nous précisait qu'il participait à cette période de façon régulière aux séminaires du laboratoire de lexicologie politique de Saint-Cloud, et que c'est là qu'il avait pris connaissance des méthodes des linguistes³⁴⁰, aussi bien que par ses lectures personnelles³⁴¹.

Toutefois, insistons sur le fait que les textes qui composent les corpus de Genet restent différents de ceux sur lesquels travaillent ces linguistes. Sur ce point, il est particulièrement intéressant de considérer une question centrale qui se pose en analyse de discours, à savoir celle de la représentativité du corpus choisi³⁴². De quoi ces corpus de textes constitués par Genet sont-ils représentatifs ? Et, plus largement, qu'est-ce qu'un corpus représentatif en histoire médiévale ?

³³⁹ Sur ces questions voir en particulier Jacques Guilhaumou, « Le corpus en analyse de discours : perspective historique », *Corpus*, n° 1 (15 novembre 2002).

³⁴⁰ Cette information nous a été donnée par Genet lui-même, dans l'entretien que nous avons mené auprès de lui en 2017.

³⁴¹ En plus du témoignage de Genet de 2017, un élément appuie cette affirmation. Dans les notes de bas de page de son article de 1974, il remercie certains membres du laboratoire de lexicologie de Saint-Cloud pour la relecture de son article.

³⁴² Voir, par exemple, Guilhaumou, « Le corpus en analyse de discours ».

À la différence des travaux de linguistique historique menés à Saint-Cloud par exemple, qui concerne des textes contemporains³⁴³ ou modernes³⁴⁴, Genet n'est pas en mesure de constituer des corpus qui peuvent prétendre à l'exhaustivité. La configuration archivistique des sources médiévales le met toujours face à des vides documentaires. Ses corpus ne sont donc représentatifs que d'une réalité documentaire, qui donne d'emblée une image déformée des idées politiques au bas Moyen-Âge en Angleterre. Les analyses statistiques qu'il met en œuvre, nous le verrons dans la troisième partie de ce chapitre, nécessitent donc de prendre en compte, au moment des conclusions, cet aspect particulier lié aux sources de la période étudiée.

La démarche de constitution des corpus mise en œuvre par Genet reflète ainsi une forte influence extra-disciplinaire. Les méthodes empruntées à la linguistique, et particulièrement le choix de constituer des ensembles de texte en corpus, destiné à opérer comme un espace de comparaison formalisé, atteste de cette influence. Nous allons voir qu'en plus de ces considérations tenant aux aspects intellectuels de ces opérations, il existe un faisceau de critères matériels qui préside au choix des textes et qui influe sur ce qu'il est possible d'y étudier.

2.3. Opérations d'édition des textes sur cartes perforées IBM à 80 colonnes.

Le deuxième ensemble d'opérations auquel nous nous intéresserons a pour finalité l'édition des textes sur cartes perforées. Ces opérations permettent, dans un second temps, d'appliquer les différents traitements prévus par le programme pour produire les index de fréquences, les dictionnaires et les concordances sur lesquels les acteurs s'appuieront pour analyser les textes. Quatre types d'activités sont associés à ce moment du travail : 1) un travail de critique des sources, 2) un travail d'édition « *traditionnel* » du texte³⁴⁵, 3) un travail de perforation des cartes, puis 4) un travail de relecture et de correction des erreurs de saisie sur le fichier des cartes perforées. Ce déroulé de la collecte des données

³⁴³ Prenons ici un exemple parmi tant d'autres, celui de l'étude sur le vocabulaire des pétitions ouvrières en 1848 : Maurice Tournier et al., « Le vocabulaire de la Révolution. Pour un inventaire systématique des textes », *Annales historiques de la Révolution française* 195, n° 1 (1969): 109-24.

³⁴⁴ Nous pensons ici particulièrement au travail de Jacques Guilhaumou sur le journal révolutionnaire *Le Père Duchesne*, mené à Saint-Cloud avec l'aide André Salem, et dont certains résultats sont présentés dans Jacques Guilhaumou, « L'idéologie du Père Duchesne: les forces adjuvantes (14 juillet-6 septembre 1793) », *Le Mouvement social*, n° 85 (1973): 81-116.

³⁴⁵ Nous reprenons ici le terme utilisé par Genet dans l'un des rapports d'activité précédemment cité.

apparaît sensiblement similaire à celui présenté dans le cadre de l'analyse de l'enquête du *catasto*, mais nous allons voir que des différences frappantes apparaissent du fait des disparités liées aux organisations collectives et à la méthode d'analyse mise en œuvre.

2.3.1. Du manuscrit à la cartes perforée.

La première étape, incontournable, est souvent passée sous silence lorsqu'on s'intéresse aux pratiques de la lexicométrie en histoire : c'est l'existence d'une première analyse « traditionnelle » des textes. Cette phase est indispensable pour le travail de perforations des textes, aussi bien que pour l'analyse des contextes de production et d'énonciation de ces écrits. Dans le cas du projet ALINE, ce travail donnera d'ailleurs lieu à deux publications, liées à deux projets personnels différents : l'une portant sur les écrits politiques de Christine de Pisan par Gauvard³⁴⁶, et l'autre sur l'un des auteurs du corpus de Genet, Sir John Fortescue³⁴⁷. Nous reviendrons dans la partie suivante sur l'une de ces deux études pour la comparer avec les analyses des textes menées avec ALINE.

Ensuite, les médiévistes doivent entreprendre une édition plus traditionnelle de ces textes, afin d'établir quelle version de chacun d'entre eux ils enregistreront pour leurs analyses. Les textes peuvent soit avoir déjà été édités par des chercheurs extérieurs à l'ERA 713, soit nécessiter une édition originale par l'auteur du projet, comme c'est le cas pour plusieurs textes qui intéressent Genet. Cet aspect du travail de Genet aboutira d'ailleurs à une édition de quatre traités politiques en 1977, qui lui vaudra une reconnaissance importante chez les médiévistes outre-Manche, plusieurs de ces textes étant jusque-là inédits³⁴⁸. Parmi ces quatre textes édités et publiés par Genet en 1977, deux font partie du corpus de textes destinés à être analysés automatiquement à l'aide de ALINE : 1) le *Tractacus de Regibus*, sorte de miroir au prince rédigé en anglais par un auteur anonyme à la fin du XIV^e siècle et 2) *The III Considerations right necessarye to*

³⁴⁶ Claude Gauvard, « Christine de Pisan a-t-elle eu une pensée politique ? A propos d'ouvrages récents », *Revue Historique* 250, n° 2 (508) (1973): 417-30.

³⁴⁷ Voir Jean-Philippe Genet, *Les idées sociales de Sir John Fortescue* (Paris, France: Publications de la Sorbonne, 1973).04/09/2021 14:18:00

³⁴⁸ Voir l'édition traditionnelle de certains textes en anglais par Genet dans Jean-Philippe Genet, éd., *Four English political tracts of the later Middle Ages* (London, Royaume-Uni de Grande-Bretagne et d'Irlande du Nord: Offices of the Royal historical society, 1977).

the good governance of a prince, traduction partielle d'un traité en français écrit en 1347 par un autre auteur anonyme.

Cette phase du travail est d'une importance capitale car elle permet l'établissement d'une édition critique des textes qui sont, dans leur forme originale des manuscrits médiévaux. Sous cette forme, ils présentent des difficultés de lecture, à la fois pour des raisons paléographiques, et parce qu'ils sont dépourvus d'orthographe stable. Cette opération, qui relève d'un travail d'édition classique, permet de faire disparaître un certain nombre d'ambiguïtés contenues dans ces textes, travail qui n'était pas nécessaire dans le cas que nous avons étudié dans le chapitre précédent, puisque Klapisch-Zuber et Herlihy entendaient proposer non une édition de chacune des quelque 60 000 déclarations des contribuables florentins, mais bien une nouvelle forme d'instrument de travail regroupant seulement certaines informations extraites de ces déclarations.

En plus de ce travail d'édition traditionnelle, la saisie des textes sur cartes perforées suppose un nouvel ensemble de procédures d'édition, différentes de celles qu'implique une édition traditionnelle : c'est la troisième étape de cette collecte de données, qui a pour finalité l'enregistrement des textes sur des cartes perforées IBM à 80 colonnes. Revenons d'abord sur les aspects matériels de ces opérations. Comme nous l'avons vu dans le chapitre précédent, chaque colonne des cartes IBM permet d'inscrire un unique caractère alphanumérique. Pour le cas d'une saisie de texte, chaque lettre occupe donc une colonne. Il est ainsi possible d'inscrire 70 caractères par carte, les cinq premières colonnes étant réservées à la numérotation des cartes et à l'inscription de trois initiales permettant de désigner le texte, et les cinq dernières étant réservées pour l'éventuelle correction des erreurs dans l'enregistrement des textes³⁴⁹.

L'autre aspect qui nous intéresse concerne les méthodes que mettent en œuvre les chercheurs de l'ERA 713 pour l'étape de la perforation. Par contraste avec le chapitre précédent, le premier élément marquant est le choix fait par les acteurs de se passer des bordereaux de saisies, technologie intellectuelle pourtant centrale dans le cas de l'enquête du *catasto*. Les raisons de ce choix sont doubles. Les premières tiennent à la fois au manque de subsides et à la nécessité d'économiser du temps, en évitant cette phase de l'enregistrement des données. L'autre raison est plus proprement liées à la nature des

³⁴⁹ Cette explication nous a été donnée par Genet au cours d'un entretien mené en 2017. Le fait de laisser cinq colonnes vides permettait, en cas d'erreur, de ne pas avoir à retaper l'ensemble des cartes qui suivait, mais de remplacer seulement la carte qui contenait l'erreur.

textes étudiés, puisque, comme nous l'avons vu, ces considérations conduisent les acteurs à ne pas avoir recours à des entreprises de perforation, mais à privilégier une perforation par les médiévistes eux-mêmes.

Toutefois, cette configuration matérielle impose également un ensemble de limites auxquelles se confrontent les médiévistes. Tout d'abord, en l'absence de petites mains similaires à celles mobilisées dans l'enquête du *catasto*, et en présence d'une gigantesque masse d'informations à enregistrer, la perforation devient vite un moment extrêmement chronophage. Pour perforer un texte de 7 000 caractères, soit environ 1 000 mots, il faut en effet produire 1 000 cartes ! Comme pour l'étude du *catasto*, le gain de temps escompté par certains utilisateurs des technologies informatiques en histoire n'est pas à rechercher dans les opérations de collecte et d'enregistrement des données.

Une fois ces opérations exécutées, une dernière étape est encore nécessaire. Il s'agit de la vérification des données ainsi enregistrées. Une fois la perforation effectuée, il devient possible d'entrer les textes en machine et de demander à l'ordinateur un texte de contrôle, qui va permettre de nettoyer le fichier, en vérifiant la fidélité des données textuelles au texte initial. C'est Millet qui s'occupe, au moins en partie, et à partir de 1976, de ce nettoyage, c'est-à-dire de la correction et de la relecture, non plus du texte, mais bien des données, sur les listings fournis par l'ordinateur. Elle met ainsi en œuvre ses qualités d'historienne et ses compétences en informatique, acquises dans le cadre de son travail de thèse sur les chanoines de Laon³⁵⁰, pour repérer les erreurs et perforer de nouvelles cartes qui remplaceront celles qui en contiennent.

Les différentes opérations que nous avons détaillées ci-dessus concourent au changement de format de ces textes, qui passent du papier des manuscrits médiévaux aux supports cartonnés des cartes perforées. C'est aussi une étape à forte valeur symbolique, qui marque le passage d'un texte à des données textuelles. Cette transformation constitue un autre élément fondamental pour comprendre la différence avec l'étude du *catasto*. En effet, la nature des données enregistrées est complètement différente dans ce deuxième projet. Les données ne nécessitent pas la définition d'un code du même type que celui établi par Klapisch-Zuber et Herlihy, mais un codage tout à fait élémentaire qui concourt à la *modélisation* des textes sous forme de chaînes de caractères. Genet et Gauvard

³⁵⁰ Millet, « Les chanoines du chapitre cathédral de Laon (1272-1412) ».

développe ainsi un tout autre rapport aux sources qui s'articule avec des objectifs de recherche et des intérêts scientifiques différents.

Pour autant, comme dans le cas de l'enquête du *catasto*, l'utilisation des cartes perforées conditionne certaines procédures de modification du texte et limite ces possibilités de modélisations. Si Genet affirme dans le rapport pour le CNRS de 1978 que le programme ne nécessite « *ni précodage, ni manipulations préalables des textes auxquels on l'applique* »³⁵¹, la réalité semble toute autre. En effet, il s'agit pour les éditeurs de rendre le texte lisible par un type de « lecteur » tout à fait particulier : l'ordinateur. De ce fait, des limites apparaissent quant aux types de caractères qu'il lui est possible de lire. Toute une série de règles de transformations du texte sont ainsi mises en place. En comparant l'édition traditionnelle de ces textes avec les règles de l'édition sur cartes perforées, il est possible de saisir plusieurs problèmes qui se posent au médiéviste et d'éclairer les méthodes d'édition des textes pour l'ordinateur auxquelles les chercheurs recourent dans le cadre de l'utilisation de ALINE.

2.3.2. Quelques problèmes spécifiques aux textes étudiés à l'ERA 713.

Le premier problème auquel sont confrontés les médiévistes de l'ERA 713 concerne le fait que l'orthographe du moyen-anglais n'est pas fixée. Par exemple, dans son édition du *Tractacus de Regibus* de 1977, Genet présente l'un des problèmes que soulève cette situation en ces termes :

*« For instance, King, subject plural, is written in full some fifteen times, under the forms Kyngis (3), Kynges (7) and Kyngus (5). I have normally chosen the most usual form, Kynges, for the extension »*³⁵².

On voit ici apparaître un problème éditorial classique, et qui est facilement géré par Genet dans le cadre de son édition traditionnelle sur papier. Cependant, la question se pose tout autrement si la finalité de l'édition est le comptage automatique des mots par l'ordinateur.

³⁵¹ Genet, « Rapport d'activité pour le CNRS », 1978.

³⁵² Genet, *Four English political tracts of the later Middle Ages*, 161.

En effet, si le chercheur considère que ces trois formes différentes sont bien le même terme, il se doit de les réunir sous un même vocable afin de pouvoir compter toutes les occurrences du terme « King », quelles que soient leurs formes dans le texte. Deux solutions existent dans ce cas : soit le médiéviste enregistre toujours le terme « King », et perd ainsi les variantes orthographiques originales, soit il met en œuvre une lemmatisation, c'est-à-dire, un traitement du texte qui lui permettra de demander au programme de compter ensemble toutes les formes du mot « King ». C'est cette seconde solution qui sera choisie dans le cas de ALINE.

Une autre difficulté du même type repose sur la langue de ces textes : certaines lettres ne sont pas de l'alphabet latin. Dans ce cas, le problème est simple à régler pour l'édition traditionnelle du texte, puisqu'il est possible d'introduire des caractères spéciaux adaptés au manuscrit. Dans le cas de l'édition sur cartes perforées, la solution choisie est la même, à la différence que le clavier alphanumérique des perforatrices ne comporte pas tous les caractères spéciaux nécessaires. Pour dépasser cette difficulté, ALINE permet au chercheur d'éviter tout pré-codage, et de choisir lui-même son idiosyncrasie : si un caractère n'existe pas sur le clavier de la perforatrice, il peut choisir n'importe quel autre caractère original pour représenter le caractère non latin, en le précisant au moment de l'entrée des données en machine.

Ces deux types d'exemple révèlent des problèmes directement liés à la langue des textes étudiés et sont intimement liés à la constitution du programme ALINE. Toutefois, ils ne posent pas de problèmes de fond insurmontables. D'autres cas posent au contraire de sérieuses questions sur les limites matérielles qu'impliquent l'utilisation des cartes perforées IBM pour l'enregistrement des textes, et attestent de l'influence de ces supports de stockage sur la modélisation des textes.

Premier exemple : comment introduire les différences qui peuvent exister entre les divers manuscrits qui servent de base à Genet pour éditer le texte ? Dans son édition de 1977 du traité *The III Considerations*, Genet précise, en note de bas de page, les variations qu'il observe entre les cinq versions du manuscrit sur lesquelles il fonde son travail³⁵³. Dans le cadre d'une édition sur cartes perforées cette opération est impossible : il faut fixer une fois pour toute la forme finale du texte. Malheureusement pour nous, en l'absence d'archives permettant de consulter les textes tels qu'édités sur cartes perforées

³⁵³ Genet.

par Genet, il est impossible de préciser les décisions qu'il a prises quant à ce premier problème d'envergure.

Une autre limite du même type apparaît lors de l'enregistrement de la structure diagrammatique de la source, que nous définissons comme l'organisation du document en colonnes, en chapitres ou en parties ³⁵⁴. Dans son édition de 1977 du traité *Tractacus de Regibus*, Genet précise au lecteur, toujours dans les notes de bas de page, le découpage en chapitres du manuscrit. Qu'en est-il de cette organisation du texte en parties dans le cadre d'une édition sur cartes perforées ? Sur ce dernier support, et en raison de l'absence de pré-codage voulue par les concepteurs de ALINE, le texte est littéralement transformé en un paquet de mots, sans possibilité d'indiquer le passage d'un paragraphe à un autre ou d'un chapitre à un autre.

Un dernier type de limites lié à cette manière d'enregistrer les données textuelles tient à l'exclusion, là encore, en raison de l'absence de pré-codage, des aspects codicologiques de ces documents (taille des marges, type de papiers utilisé, présence d'enluminures, lettrines, etc.), qui disparaissent en raison du support de stockage et des règles d'enregistrement des données choisies.

Au cours de ce processus de saisie, en plus des modifications ponctuelles du document que nous avons mentionnées plus haut, il existe donc tout un ensemble d'aspects du texte qui sont laissés de côté par les éditeurs : les variations d'un manuscrit à l'autre, la structure diagrammatique du texte et ses aspects codicologiques. Cette pratique d'éditions conduit ainsi les historiens de l'ERA 713 à transformer les textes originaux sous la forme d'un modèle particulier : une chaîne de caractères spécifiquement normalisée pour l'entrée en machine. En dépit de l'originalité de cette pratique dans les études médiévales au début des années 1970, ce type de modélisation des textes s'inscrit pleinement dans la continuité des travaux en statistiques textuelles, et en l'espèce, en lexicométrie quantitative. En effet, ces méthodes, qui se basent sur des opérations de dénombrement et de classement des mots, supposent que leurs utilisateurs conçoivent les mots comme des unités statistiques, et, donc, qu'ils réduisent les textes à certaines des informations qu'ils pourraient offrir.

³⁵⁴ Nous empruntons ce qualificatif à Karine Chemla, tel qu'elle l'a présenté dans l'article suivant : 林力娜 Karine Chemla, « On the Diagrammaticity of Ancient Texts and Its Importance for the History of Science: Based on the Example of the Early Chinese Mathematical Text The Gnomon of the Zhou », *East Asian Science, Technology and Society: An International Journal* 14, n° 2 (1 juin 2020): 279-308.

L'analyse des opérations qui conduisent à l'édition des textes sur cartes perforées nous a permis de dégager un certain nombre d'éléments qui révèlent les mutations de la pratique des sources au contact des ordinateurs. Le premier élément concerne l'évolution des méthodes de constitution des corpus, qui, sous l'influence de l'emprunt de méthodes aux statistiques lexicales, deviennent, nous l'avons dit, une instance de contrôle du parcours interprétatif des textes par l'historien, ou, en d'autres termes, la « *matrice du sens* » des phénomènes linguistiques identifiés dans ces documents. Nous reviendrons sur cette évolution dans la prochaine partie de ce chapitre, en analysant les raisonnements mis en œuvre par Genet sur la base de l'un de ses corpus.

Le deuxième élément qui ressort de notre analyse porte sur la mise en œuvre d'une méthode originale d'édition des textes dans la discipline. Nécessitant un transfert des textes du support papier vers les cartes perforées, cette méthode implique l'élaboration d'une modélisation des textes, qui prennent alors la forme de chaîne de caractères. Il est particulièrement intéressant de comparer ici les méthodes développées par Genet avec celles mises en œuvre pour l'étude du *catasto*. Dans chacun de ces deux études de cas, on constate en effet que les critères et les méthodes de l'édition de sources évoluent, mais dans des directions différentes, en fonction des objectifs de recherche visés par ces deux collectifs. D'une part, Klapisch-Zuber et Herlihy extraient d'une série de déclarations un ensemble de variables économiques et démographiques, sur la base d'un code élaboré en amont de la collecte des données. D'autre part, les médiévistes de l'ERA 713 éditent les textes sous la forme de chaînes de caractères, qui se construit en amont sur la base d'une édition réalisée de manière plus traditionnelle.

Enfin, le troisième élément concerne l'évolution des objets publiés par les acteurs, sur lesquels nous reviendrons plus en détail dans la quatrième partie de ce chapitre. D'un côté, Klapisch-Zuber et Herlihy publient une édition du *catasto* pour l'ordinateur. De l'autre, l'ERA 713 met en place une bibliothèque de textes sur cartes perforées, intitulée MEDITEXT, ainsi qu'un « package » d'analyse automatique des textes. Ces deux objets sont absolument nouveaux dans le domaine de l'histoire médiévale au milieu des années 1970, et préfigure toute une série de nouvelles formes de publication qui émerge du fait de l'utilisation des ordinateurs par les chercheurs en sciences humaines (base de données, édition électronique et programme informatique).

Pour saisir les raisonnements qui sont mis en œuvre sur la base de cette nouvelle forme d'édition de textes et, analyser comment le recours aux ordinateurs modifie les raisonnements de ces médiévistes dans leurs études des textes politiques, il nous faut nous pencher plus en avant sur l'utilisation du programme ALINE.

3. Raisonnements, programme et statistiques lexicales.

Pour saisir les raisonnements historiques développés par les médiévistes de l'ERA 713 à l'aide de ALINE, la documentation disponible est très limitée. En effet, par contraste avec le chapitre précédent, il ne nous a été possible de trouver qu'un seul article qui fasse état de son utilisation : « Automatic text processing and factorial analysis : a method for determining the lexicographical horizon of expectation », publié en 1984 dans les actes du colloque de Kalamazoo (États-Unis) de 1979³⁵⁵. De plus, comme le note l'auteur dans son introduction, cet article se concentre avant tout sur des questions méthodologiques :

*« The value of research lies chiefly in its results. The following article will, therefore, be disappointing for many, since it concentrates on methods rather than results. »*³⁵⁶

Il demeure cependant possible, en mettant en relation cet article avec d'autres documents, de saisir certains des raisonnements mis en œuvre par les médiévistes de l'ERA 713 à l'aide de ALINE. Comme pour l'analyse de l'étude du *catasto*, nous apprécierons dans cette troisième partie les apports historiographiques et épistémologiques de l'utilisation des ordinateurs dans ce contexte, en comparant ces méthodes nouvelles avec d'autres, plus traditionnelles, développées en histoire des idées politiques dans les années 1970 et 1980.

³⁵⁵ Jean-Philippe Genet, « Automatic text processing and factorial analysis: a method for determining the lexicographical horizon of expectation », *A. Gilmour-Bryson, (éd.), Computer Applications to Medieval Studies, Kalamazoo, Western Michigan University, 1984, 147-75.*

³⁵⁶ Genet, 147.

Pour ce faire, nous procéderons dans un premier temps à une comparaison la méthodologie proposée par Genet dans un article publié en 1973, pour laquelle il n'utilise pas de méthodes informatiques ou de statistiques lexicales, avec une autre méthodologie d'histoire des idées politiques développée par l'historien anglais Quentin Skinner (1940-) à la fin des années 1960. Dans un second temps, nous pourrions nous intéresser plus précisément à l'originalité des raisonnements linguistiques mis en œuvre par Genet dans ce contexte, en nous basant sur les documents de sortie que permet d'obtenir le programme ALINE (concordances, index, dictionnaires). Enfin, nous nous intéresserons de plus près aux raisonnements développés par Genet à l'aide d'une méthode d'analyse des données : l'analyse factorielle des correspondances.

3.1. La mesure du vocabulaire comme méthode pour l'histoire des idées politiques.

3.1.1. Dépasser le « *contextualisme* » de Quentin Skinner.

Dans l'introduction de sa thèse d'État publiée en 1997, Genet définit en ces termes le problème que lui posaient les méthodes développées dans les années 1960 par un autre historien des idées politiques :

« impossible de partir, même en s'entourant de toutes les précautions que recommande Quentin Skinner par exemple, d'une approche descriptive empirique de ce que pourraient être les idées sociales et politiques d'une société si ancienne qu'elle ne peut avoir, en dépit des apparences, plus grand chose à voir avec la nôtre. Les textes politiques classiques se connaissent pour la plupart et se citent comme autorités ; il n'est que trop facile pour les historiens des idées de prendre chaque idée, extraite de textes soigneusement sélectionnés grâce à un processus de raffinement où la culture et les préoccupations personnelles de l'historien des idées lui-même jouent un rôle déterminant, pour la jucher au-dessus de la précédente pour faire monter ainsi un échafaudage d'idées politiques vers un sommet ou un aboutissement. [...]. Bien sûr, cet échafaudage, cette reconstruction, donne un sens (à tous les sens du terme, soyons juste) aux textes. Mais c'est bien cela qui est gênant : il faut admettre que ces textes de la fin du Moyen-Âge doivent, dans un premier temps, nous rester radicalement

incompréhensibles si nous voulons parvenir à les comprendre, simplement parce qu'il faut que nous sachions de quoi ils parlent ; et ce dont ils parlent, ce n'est, et pour cause, ni de nous, ni de notre société, ni même (par exemple) de l'État, ou du moins de ce que nous entendons par État, puisque cet État-là n'existe assurément pas à la fin du Moyen-Âge »³⁵⁷.

Quentin Skinner, né en 1940 et historien de formation, est au début des années 1970 l'un des chefs de file de l'école dite « de Cambridge ». Il propose de renouveler la manière de faire l'histoire des idées politiques, incarnée notamment, dans le monde anglo-saxon, par les approches idéalistes de Arthur O. Lovejoy (1873-1962), Georges Sabine (1880-1961) ou Leo Strauss (1899-1973). Ces derniers, dont les écrits ont été largement diffusés en Angleterre comme aux États-Unis, développent une histoire basée sur la recherche des « idées élémentaires » (Lovejoy), des « problèmes universels » (Sabine) et/ou proposent une reconstruction anhistorique et linéaire des progrès politiques (Sabine, Strauss)³⁵⁸.

Pour dépasser ces approches idéalistes, Skinner développe une nouvelle méthode d'approche des textes basée sur ce que les observateurs ont appelé un « contextualisme ». Au contraire des historiens idéalistes, il soutient que l'analyse des textes de philosophie politique et morale doit se fonder sur une analyse des « contextes discursifs »³⁵⁹ dans lesquels les œuvres s'inscrivent. Pour comprendre le texte, l'historien doit donc prendre en compte non seulement sa signification, mais aussi sa place comme forme d'action sociale : comment interagit-il avec d'autres textes ? À qui l'auteur répond-il ? Quelle position défend-il et devant qui la défend-il ? Cette approche conduit non seulement à une relecture de certaines grandes œuvres de la philosophie politique, comme le *Leviathan* de Thomas Hobbes³⁶⁰, mais aussi à « élargir le corpus traditionnel d'auteurs reconnus par la tradition philosophique »³⁶¹.

³⁵⁷ Genet, « Les idées sociales et politiques en Angleterre », 53.

³⁵⁸ Voir, par exemple : Arnault Skornicki et Jérôme Tournadre, « I. L' « école de Cambridge » et le contextualisme », *Repères*, 12 mai 2015, 9-32.

³⁵⁹ Skinner utilise notamment ce terme dans un entretien mené par Florent Guénard, « Ce que la philosophie veut dire », *La Vie des idées*, 3 mai 2019, <https://laviedesidees.fr/Ce-que-la-philosophie-veut-dire.html>. On pourra encore regarder Nicolas Brisset, « Qu'est-ce qu'un contexte ? Quentin Skinner et l'histoire des idées », *Economia. History, Methodology, Philosophy*, n° 10-3 (1 septembre 2020) : 581-92 ou Sophie Marcotte-Chénard, « Le contextualisme de Quentin Skinner à l'épreuve du cas Machiavel », *Methodos. Savoirs et textes*, n° 13 (23 avril 2013).

³⁶⁰ Quentin Skinner, *Reason and rhetoric in the philosophy of Hobbes* (Cambridge (GB), Royaume-Uni de Grande-Bretagne et d'Irlande du Nord: Cambridge university press, 1996).

³⁶¹ Skornicki et Tournadre, « I. L' « école de Cambridge » et le contextualisme ».

Il apparaît, à la lecture de la citation de Genet reproduite plus haut, que Skinner fait figure pour lui de modèle autant que de contre-modèle. En effet, le travail de Genet incarne lui aussi cette volonté d'étudier les idées politiques en contexte. Nous l'avons vu dans la partie précédente, il propose même une approche plus radicale en prenant en compte la position sociale des auteurs ainsi que les conditions de la production et de la circulation des textes – ce qu'il appelle dans sa thèse les « *conditions de communication* »³⁶². Par ailleurs, comme Skinner, son travail conduit à élargir le corpus d'œuvres considérées comme politiques pour la période concernée.

Cependant, la citation reproduite ci-dessus nous montre que Genet entend dépasser ce modèle. Pour préciser comment la critique de Genet envers Skinner se manifeste en pratique dans le travail d'analyse des textes, il nous est possible d'examiner un article publié par Genet en 1973. L'article en question s'intitule « Les idées sociales de Sir John Fortescue »³⁶³, et s'inscrit dans le travail de début de thèse de Genet. Ce texte nous intéresse particulièrement puisqu'il est contemporain de l'élaboration de ALINE, et qu'il contient des indices qui nous permettent d'y saisir les raisons de l'intérêt que porte Genet aux méthodes de la statistique lexicale, à la lexicométrie et aux ordinateurs.

3.1.2. Une analyse des idées politiques de Sir John Fortescue.

Pour Genet, l'enjeu de l'article est de montrer comment l'œuvre de Sir John Fortescue, avocat, juge, juriconsulte et théoricien politique anglais du XV^e siècle, doit permettre aux historiens des idées politiques de :

*« dépasser la sempiternelle discussion du problème de la « monarchie royale et politique » »*³⁶⁴.

Ce « *problème* », qui anime les discussions sur la genèse de l'État moderne en Europe, peut être résumé en ces termes : l'État moderne est-il né par le haut, d'une domination du peuple par la force d'un chef (modèle de la « *monarchie royale* »), ou par le bas, d'une

³⁶² Voir notamment l'introduction de la publication de son travail de thèse : Jean-Philippe Genet, *La genèse de l'État moderne: culture et société politique en Angleterre* (Paris, France: Presses universitaires de France, 2003).

³⁶³ Genet, *Les idées sociales de Sir John Fortescue*.

³⁶⁴ Genet, 448.

adhésion du peuple au pouvoir du souverain (modèle de la « *monarchie politique* ») ? Pour contribuer au dépassement de cette question, Genet propose une analyse que nous estimons proche de la méthode développée par Skinner.

Plusieurs indices montrent que Genet insiste sur des éléments du contexte de production des œuvres de Fortescue pour comprendre les idées de cet auteur. La première partie de l'article est consacrée à l'exposé d'éléments biographiques. La carrière politique et juridique de Fortescue, auprès des familles royales anglaises (Henri VI, Marguerite d'Anjou), de Louis XI en France (auprès duquel il vient chercher du soutien en 1462), conduit Genet à établir que Fortescue est un « *témoin bien placé* »³⁶⁵ pour lui permettre de saisir les idées des élites politiques en Angleterre au début du XV^e siècle.

Dans une seconde partie, Genet développe les grands traits de la pensée politique de Fortescue. Le médiéviste examine notamment les références que mobilise ce théoricien politique, parmi lesquelles, l'œuvre de Saint-Augustin. Il s'intéresse encore aux publics auxquels sont destinés les différents textes écrits par Fortescue au cours de sa vie. Genet poursuit son étude en illustrant, sur la base d'un texte qu'il juge particulièrement important, *The Governance of England*, l'originalité de la pensée de Fortescue à cette époque en Europe de l'Ouest. Il insiste, en particulier, sur l'intérêt que porte ce penseur anglais à la question des finances publiques, à travers une comparaison entre la monarchie française et la monarchie anglaise. L'analyse de Genet se termine enfin par l'exposé de la cohérence de la pensée politique de Fortescue, qui mêle rigoureusement les caractères géographiques de l'Angleterre, avec les structures politiques, économiques, sociales et culturelles.

Il faut faire remarquer ici que Genet, dans cet article, étudie les idées de Fortescue en s'appuyant sur une méthode d'analyse des textes tout à fait classique. Le médiéviste fait référence aux œuvres de l'auteur anglais à travers des citations, qu'il commente pour dégager les idées politiques et sociales de l'auteur du *Governance of England*. L'objet de l'étude est certainement neuf, en raison du personnage qu'elle concerne, mais la méthode mise en œuvre révèle tous les biais dénoncés par Genet dans l'introduction de sa thèse en 1997. À la lumière de l'écart chronologique entre cette introduction et cet article de 1973, il semble vraisemblable que Genet n'ait pas encore formulé ici clairement les critiques

³⁶⁵ Genet, 454.

qu'il adressera par la suite à la méthode de Skinner. Cependant, la troisième partie de cet article nous renseigne sur les limites qu'il perçoit déjà quant à ce type d'analyse.

Dans cette troisième partie, Genet s'intéresse plus particulièrement aux termes des catégories sociales utilisés par Fortescue : « *great men of the londe* », « *franklins* », « *yeomen* », mais aussi « *chevalier, écuyers, householders ou tenanciers* »³⁶⁶. La diversité des usages de ces termes conduit Genet à se poser la question de l'articulation entre certaines des catégories sociales qu'ils désignent, et, en particulier, à la différence que Fortescue établit entre *gentry* et noblesse. Pour Genet, il existe ici une « *ambiguïté dans la pensée de Fortescue* », puisque la *gentry*, pour laquelle il utilise un terme spécifique, est, selon les historiens, fortement liée, sinon confondue à la noblesse pour des raisons aussi bien économiques que culturelles. Genet conclut :

*« Il y a donc là une ambiguïté dans la pensée de Fortescue, mais que nous devons respecter, car les travaux des historiens de l'économie et de la société anglaises tendent à démontrer qu'elle est amplement justifiée par la réalité »*³⁶⁷.

Pour aller au-delà de cette ambiguïté, Genet suggère la nécessité d'une analyse d'un autre type, qu'il présente en ces termes :

*« L'œuvre de Fortescue pourrait toutefois nous aider à en préciser les conséquences politiques : ce serait ici à une étude serrée de l'emploi du mot « Commons » qu'il faudrait se livrer ; mais les occurrences du mot sont nombreuses, ses contextes variables et ses sens différents : nous attendons pour éclaircir ce point les résultats d'une enquête en cours portant sur l'analyse exhaustive du vocabulaire anglais de Fortescue »*³⁶⁸.

Un autre sens du terme contexte apparaît ici, qu'il nous faut être associer non plus au contextualisme de Skinner, mais bien aux travaux des linguistes. Malheureusement pour notre analyse, aucun article de Genet ne rend véritablement compte d'une « *étude serrée de l'emploi du terme « Commons »* » dans l'œuvre de Fortescue. On sait cependant que

³⁶⁶ Genet, 456.

³⁶⁷ Genet, 457.

³⁶⁸ Genet, 457.

le texte *Governance of England* a bien été analysé à l'aide de ALINE dans un article publié en 1984, mais sans attention particulière pour un terme ou un autre (nous y reviendrons au 3.3.).

Cette affirmation de l'existence d'une « *enquête en cours* » portant sur « *les occurrences du mot* », « *ses contextes variables et ses sens différents* » nous renseigne cependant sur l'une des directions de recherche qu'envisage alors Genet pour dépasser l'approche de Skinner. C'est ici, en analysant les possibilités qu'offrait le programme ALINE pour l'analyse des emplois d'un terme, qu'il est possible de reconstituer les raisonnements élaborés par Genet au contact des ordinateurs.

3.2. Comptages, tris et ordinateurs.

Trois documents nous permettent de mieux comprendre les manipulations des données textuelles rendues possibles à l'aide du programme ALINE : 1) un article co-écrit par Genet, Hucher, Mondelli et Valensi en 1974, qui présente précisément les fonctionnalités du programme³⁶⁹; 2) un document interne déjà cité à plusieurs reprises, non daté, destiné aux utilisateurs de ALINE et intitulé « Programme de traitement automatique des textes pour l'équipe de recherche sur l'histoire politique du bas Moyen-Âge » et 3) l'article, déjà mentionné, de 1984, rédigé en anglais, qui constitue l'unique publication de résultats d'une étude menée à l'aide de ALINE³⁷⁰. Ces trois documents nous permettent de reconstituer comment avait pu se mettre en place une « *étude serrée* » d'un terme du corpus de Genet. Ce petit corpus nous éclaire encore sur les fonctionnalités du programme et sur la fonction épistémologique des opérations qu'il rend possible.

3.2.1. Fonctionnalités du programme ALINE.

Dans l'article de 1984, publié plus de dix ans après le début de l'aventure ALINE, Genet se base sur un corpus de six textes anglais, choisis selon des modalités qu'il précise en ces termes : ils appartiennent à des périodes différentes (1300-1480), traitent de sujets variés (relations entre le pouvoir spirituel et laïc, conseil au prince, problèmes pratiques de la politique publique), mais appartiennent aussi, selon les critères de classification

³⁶⁹ Genet et al., « Un programme de traitement automatique des textes ».

³⁷⁰ Genet, « Automatic text processing and factorial analysis ».

traditionnels, à des genres différents (miroir au prince, sermon parlementaire, discours ou texte polémique).

La première série d'opérations faite sur la base de ces textes, après leur édition sur cartes perforées, consiste à calculer (« *computed* »), à l'aide de l'un des modules du programme ALINE, deux index lexicaux : un index de fréquences (« *frequency index* ») et un index de fréquences relatives (« *relative frequency index* »)³⁷¹. Ces deux documents, nous le verrons, sont les pièces centrales car nécessaires aux différents raisonnements des médiévistes. Dans l'article de Genet de 1984, ils sont à la base du raisonnement, et leur élaboration constitue la condition *sine qua non* d'une approche des textes par les méthodes quantitatives, nous y reviendrons. Ces opérations sont également centrales pour Genet afin de poursuivre l'étude d'un mot telle qu'il l'avait envisagée dans son article de 1973, puisqu'elles permettent de compter le nombre d'occurrences d'un terme et, par conséquent, d'établir un indice quantitatif permettant la comparaison de son utilisation d'un texte à l'autre.

Pour produire ces deux index, plusieurs opérations de traitement des données sont nécessaires. La première a déjà été abordée sous un autre angle dans la première partie de cet article. Elle consiste dans le travail de lemmatisation, c'est-à-dire dans l'opération de regroupement des *formes* des textes en mots. Pour ce faire, dans un premier temps, l'ordinateur fournit à l'utilisateur un dictionnaire de toutes les formes présentes dans le texte. Le travail de l'utilisateur consiste alors en trois sous-opérations : la différenciation des homographes, la réduction des orthographes irrégulières et la neutralisation des flexions (accords, déclinaisons, etc.).

Pour chacune de ces sous-opérations, les médiévistes doivent résoudre un certain nombre de problèmes capitaux pour la suite. Dans l'article de 1974 mentionné plus haut, les quatre auteurs (Valensi, Hucher, Mondelli, Genet) explicitent l'importance de « *se fixer des règles* » pour faciliter le travail³⁷². Dans le cadre de l'étude du corpus de textes en anglais, ces règles se rapportent à quatre questions principales : 1) « *doit-on distinguer les différents sens d'un mot ?* »³⁷³ ; 2) « *doit-on accorder une rubrique spéciale à*

³⁷¹ Genet, « Automatic text processing and factorial analysis », 156.

³⁷² Genet et al., « Un programme de traitement automatique des textes ».

³⁷³ L'exemple que donne Genet est le suivant : faut-il distinguer sens (les cinq sens) et sens (le sens de l'histoire) ?

certaines formes verbales ? »³⁷⁴ ; 3) « *doit-on tenir [...] compte des phénomènes de contraction et d'hyphénation ?* » et enfin 4) que faire des mots grammaticaux ?

L'examen de ces questions suggère deux remarques. Il nous faut tout d'abord commenter ce choix de lemmatiser, qui n'est pas neutre. La majeure partie des équipes de recherche utilisant des méthodes issues de la statistique lexicale dans les années 1970 considèrent cette question comme épineuse³⁷⁵. Mais la langue des textes analysés par Genet modifie radicalement la pertinence de cette question. En effet, au contraire des corpus étudiés à Saint-Cloud, composés d'écrits de la période contemporaine, les variantes orthographiques des sources médiévales sont bien plus nombreuses. Il existe, par exemple, dans l'un des corpus de Genet, 91 formes différentes du mot « *himself* ». Dans son cas, on peut donc légitimement se demander si l'absence de lemmatisation est véritablement une option.

Ensuite, il faut insister sur l'importance de l'opération de définition des mots grammaticaux, qui peuvent être présentés sommairement comme l'ensemble des conjonctions, des articles, des pronoms, etc. Genet prend ici encore pour modèle le travail des linguistes de Saint-Cloud, qui les excluent de certains comptages. Cependant, dans le cadre de la mise en œuvre de ALINE, la solution est laissée à la discrétion de l'utilisateur, qui peut donc choisir ou non de les prendre en compte.

Ces différentes options possibles, à propos des homographes, des variantes orthographiques, des flexions et des mots grammaticaux, vont avoir un impact capital sur les opérations qui sont au centre de l'analyse lexicométrique. En effet, le choix de regrouper les formes qui apparaissent dans le texte sous l'égide de lemmes détermine les résultats des comptages des occurrences des mots. Ainsi, le choix des mots grammaticaux à exclure va influencer sur les résultats du calcul des indices de fréquences absolues et relatives, puisqu'il y aura moins de mots au total, mais que les nombres d'occurrences des lemmes seront les mêmes.

³⁷⁴ L'exemple choisi ici par Genet est celui de l'emploi des participes présent en anglais. Faut-il distinguer au moment de la lemmatisation *keeping* de sa forme verbale *to keep* ?

³⁷⁵ Les débats sur cette question sont intenses dans les années 1970 au sein de la communauté des linguistes comme des historiens. Pour certains il est indispensable de lemmatiser, tandis que pour d'autres, cette modification des données brutes entraîne une perte de certains résultats. La position médiane consiste à dire que cela dépend de la nature des textes étudiés, comme des objectifs de recherche. Voir par exemple Étienne Brunet, « Qui lemmatise dilemme attise », in *11^e Rencontres linguistiques en pays rhénan*, éd. par Publications de l'Université Marc Bloch (Strasbourg, France: L Kosé, A. Theissen, 1999), 7-32.

Une fois ces règles posées par les utilisateurs commence véritablement la phase de calcul des indices de fréquences. Pour comprendre la nature de ces opérations, nous regarderons en particulier la méthode de calcul des indices de fréquence relative et de fréquence absolue, détaillée dans l'article de 1974 présenté plus haut.

Pour calculer l'indice de fréquence absolue, les concepteurs de ALINE définissent les quantités numériques suivantes : 1) le nombre de réalisations d'un mot, c'est-à-dire le nombre global d'occurrences des formes différentes attachées à ce mot ; 2) le nombre total des réalisations de tous les mots du texte. L'indice de fréquence absolue d'un mot donné est ainsi la valeur du nombre de réalisations de ce mot, divisé par le nombre total de réalisations de tous les mots du texte. L'indice de fréquence relative est quant à lui le quotient du nombre de réalisations d'un mot considéré comme significatif par le nombre total de réalisations de tous les mots considérés comme significatifs, c'est-à-dire des mots qui ne sont pas considérés comme des mots grammaticaux.

L'intérêt de ALINE pour les médiévistes, au-delà du calcul automatique de ces deux types de fréquences, consiste en la possibilité qu'offre le programme d'obtenir, automatiquement là-aussi, une liste classant les mots du corpus en fonction de leurs fréquences (relative ou absolue). Ce document de travail est lui aussi fondamental pour Genet, puisqu'il lui permet, nous le verrons, de repérer des ensembles de termes qui seront au cœur de ses analyses.

Figure 13. Exemple de dictionnaire de fréquences produit à l'aide de ALINE.

TEXTE THE III CONSIDERACIONS				
DICTIONNAIRE DES FORMES SELECTIONNEES				
FORME	FREQUENCE	FREQUENCE RELATIVE	FREQUENCE CORRIGEE	STATUT
*****	*****	*****	*****	*****
AND	1089	.0875	.0875	0000
THE	606	.0487	.0487	0000
OF	507	.0408	.0408	0000
TU	320	.0257	.0257	0000
THAT	310	.0249	.0249	0000
IN	275	.0221	.0221	0000
HIS	267	.0215	.0215	0000
HE	247	.0199	.0199	0000
IS	216	.0174	.0174	0000
AS	154	.0124	.0124	0000

Source : Lucie Fossier et Jean-Philippe Genet, « Un type d'exploitation : le traitement de texte », *Le Médiéviste et l'ordinateur* 1, n° 1 (1979): 7.

C'est sur la base de cette liste de mots les plus fréquents que Genet va développer différents types d'études : une approche sémantique (voir 3.3) et une approche de classification des textes (voir 3.4). Toutefois, avant d'entrer plus précisément dans l'analyse de ces deux types de démarches, il nous faut revenir sur ces opérations de calculs de fréquence, dans le but d'établir quels sont les apports de l'automatisation pour le déploiement de ce type de calcul.

3.2.2. Apports de l'automatisation de l'établissement des dictionnaires de fréquences.

Deux sujets se doivent d'être abordés pour comprendre les apports des ordinateurs à ces opérations d'analyse du vocabulaire dans ce contexte : 1) les raisons qui poussent les linguistes, et notamment les linguistes du laboratoire de lexicologie politique de Saint-Cloud, à mettre en œuvre de nouveaux outils de calcul pour mener leurs recherches et 2) les emprunts de Genet aux méthodes qui existent en alors en lexicométrie. En effet, comme l'expose Régine Robin dans son livre de 1973 *Histoire et linguistique*, les linguistes de Saint-Cloud établissent des indices de fréquence lexicale depuis les années 1950, sans avoir pour autant recours à des moyens informatiques. Que change l'ordinateur dans la mise en œuvre de ces opérations pour les linguistes ? Comment le recours aux moyens informatiques influencent-ils leurs raisonnements ? Et comment la reprise de ces pratiques par les historiens modifie-t-elle leurs raisonnements historiques ?

Le chapitre V de l'ouvrage de Robin donne un aperçu des méthodes de lexicométrie du laboratoire de lexicologie de l'ENS de Saint-Cloud depuis 1954. Après avoir évoqué laconiquement leurs pratiques manuelles d'indexation des corpus, elle cite un document produit en 1970, et intitulé « Traitement automatique des textes », dans lequel les auteurs affirment :

« À partir d'une certaine ampleur, la consultation des index et surtout le tri des fiches ne se révèlent maniabiles qu'avec lenteur. La machine nous est donc apparue comme le seul moyen efficace de systématiser et d'objectiver

l'inventaire. »³⁷⁶

Au moins trois éléments importants pour notre problématique ressortent de cette citation. Le premier concerne le rôle déterminant joué par l'« ampleur » des index, c'est-à-dire par la masse de données extraites des corpus qu'il est possible de trier manuellement. Comme pour l'étude du *catasto*, il existe donc un lien entre la taille des corpus qu'il est possible d'étudier et l'intérêt d'avoir recours aux ordinateurs, notamment du fait de la difficulté qui naît des opérations de tris nécessaires à l'étude d'une telle quantité de mots.

Pour le cas des travaux menés au sein de l'ERA 713, il faut cependant rappeler que les textes choisis par les médiévistes sont plutôt brefs, en raison, notamment, du manque de moyens pour la perforation. On peut donc se demander si la manipulation des index à la main n'aurait pas été possible dans leur situation, au moins pour certaines des études envisagées, comme celle sur le terme « *Common* » évoquée par Genet dans son article de 1973. Sur la base de ce dernier exemple, il ne semble en tout cas pas si évident d'affirmer que l'utilisation des ordinateurs se justifie par la taille du corpus, à l'inverse du cas du *catasto* par exemple.

Deux autres éléments importants apparaissent dans la deuxième phrase de cette citation : l'ordinateur permet de « *systématiser* » et « *d'objectiver* » l'inventaire des mots. Ces deux autres raisons avancées pour justifier de l'utilisation des ordinateurs constituent deux autres aspects des motivations qui incitent les acteurs à employer cet instrument dans leurs recherches. Sur la seule base de cette citation, il demeure cependant difficile d'en saisir véritablement le sens. Est-ce que cela rejoint l'une des conséquences que nous avons pu observer dans le chapitre précédent, à savoir, le fait que l'utilisation de l'ordinateur permet d'éviter au maximum les erreurs de comptage, et donc, de rendre les opérations de calculs plus fiables ? Ou est-ce que ces deux verbes rendent compte d'autres évolutions, différentes de celles observées dans le chapitre 1 ?

Même sans avoir nécessairement de réponse tranchée à ces dernières questions, les trois éléments dégagés ici nous permettent de remarquer que les méthodes employées par Genet, et notamment le calcul des fréquences des mots, préexistent à son propre travail. Genet s'inscrit ainsi dans la continuité des travaux des spécialistes de la

³⁷⁶ Robin, *Histoire et linguistique*, 125. Le document cité par Robin, qu'il ne nous a pas été possible de retrouver, est le suivant : Annie Geoffroy, Pierre Lafon, Gabrielle Muc, Maurice Tournier, *Traitement automatique des textes*, I. : *Perforation* (polycopié de l'ENS de Saint-Cloud, mars 1970): 7.

lexicométrie de Saint-Cloud, qui ont déjà éprouvé l'intérêt des ordinateurs pour ce type d'approche des textes. Ce constat nous permet de formuler l'hypothèse suivante : les raisons qui poussent Genet à utiliser les ordinateurs sont du même ordre que celles qui ont poussé les linguistes de Saint-Cloud à utiliser ces instruments.

Plusieurs éléments permettent d'ailleurs d'établir plus précisément ce parallèle entre les méthodes développées au laboratoire de lexicométrie de l'ENS de Saint-Cloud et celles de Genet. Dans le chapitre V de son livre, Robin propose une généalogie de ces méthodes, selon le schéma suivant. Pierre Guiraud (1912-1983), linguiste français avait, dès 1954, fondé un objet nouveau en s'intéressant aux « *caractères statistiques du vocabulaire* »³⁷⁷. Il s'intéressait alors particulièrement à la fréquence des mots³⁷⁸. Charles Muller (1909-2015), autre linguiste français, avait ensuite « *approfondi, précisé et diversifié* »³⁷⁹ cette approche, en publiant notamment une *Initiation à la statistique linguistique* en 1968³⁸⁰. C'est dans la continuité des travaux de Muller que s'inscrivent, dès 1964, les travaux du laboratoire de lexicométrie de l'ENS Saint-Cloud, composé notamment d'Annie Geoffroy, Maurice Tournier (1933-2013) et Pierre Lafon³⁸¹.

Si l'on compare, en se basant sur ce chapitre de Robin, les concepts définis par Guiraud et repris par les membres du laboratoire de Saint-Cloud, le fait est que la première étape de l'analyse de Genet dans son article de 1984 suit la méthode des linguistes. Le premier élément qui va dans ce sens concerne l'identification de ce que Guiraud a qualifié de « mots-thèmes », c'est-à-dire les mots qui permettent à l'auteur d'organiser sa pensée³⁸². Dans son article Genet, s'intéresse en effet aux 40 mots les plus fréquents de son corpus, en excluant ainsi les termes que Guiraud appellent les « *mots de base* », qui, toujours dans les termes de Robin, « *constituent la substance du discours* »³⁸³, mais également les « *mots de caractérisations* »³⁸⁴, plus nombreux mais moins fréquents, qui sont d' « *une grande restriction de sens donc très précis* »³⁸⁵.

³⁷⁷ Robin, 124.

³⁷⁸ Pour une étude plus poussée, par un observateur, du travail de Guiraud et de sa contribution aux statistiques lexicales, voir Gabriel Bergounioux, « How Statistics Entered Linguistics: Pierre Guiraud at Work. The Scientific Career of an Outsider », *Glottometrics* 33 (2016): 45-55.

³⁷⁹ Robin, *Histoire et linguistique*, 124.

³⁸⁰ Charles Muller, *Initiation à la statistique linguistique* (Paris, France: Librairie Larousse, 1968).

³⁸¹ Tournier, « Le centre de recherche de lexicologie politique de l'E.N.S. de Saint-Cloud ».

³⁸² Robin, 128.

³⁸³ Robin, 128.

³⁸⁴ Robin, 128.

³⁸⁵ Robin, 128.

Plusieurs différences apparaissent cependant si l'on poursuit la comparaison. Une autre notion importante pour les lexicomètres de Saint-Cloud consiste en ce que ces chercheurs qualifient de mots-clés. Ils les définissent comme les mots « *dont la fréquence s'écarte de la normale* »³⁸⁶. Genet, qui travaille sur des textes écrits dans des langues anciennes peu étudiés par les linguistes (moyen-anglais, moyen-français), ne peut ici suivre le type d'analyse développé à Saint-Cloud, puisqu'il ne possède pas de point de comparaison qui lui indiquerait cette norme. En effet, il lui faudrait pour cela un inventaire de tous les mots de la période qui l'intéresse, travail déjà incroyablement ambitieux pour une période récente, mais tout bonnement impossible pour la période médiévale, en raison des lacunes des sources.

Une autre limite du même type apparaît encore sur la base de la présentation des méthodes lexicométriques par Robin. Dans son chapitre, l'historienne pointe vers un autre enjeu que les linguistes attachent à l'usage des ordinateurs :

*« L'inventaire manuel, en outre, ne permet pas de déterminer les caractères statistiques complexes du texte considéré. »*³⁸⁷

Un autre intérêt de l'automatisation de ces opérations apparaît ici, distinct des considérations sur le gain de temps et l'objectivation/systematisation dans le maniement des index. Il tient aux capacités de calculs et de mémoire des ordinateurs, qui permettent de construire des indicateurs plus complexes que les fréquences absolues et relatives. Les linguistes de Saint-Cloud développent, par exemple, toute une série de coefficients qui portent sur l'ensemble du texte : coefficient de fonctionnalité, de lexicalité, de redondance générale, de variété, etc. Pour des raisons qu'il nous est impossible d'expliquer, faute de documents, le programme ALINE ne permettait pas de calculer ce type de coefficient.

3.3. Automatisation de l'étude sémantique.

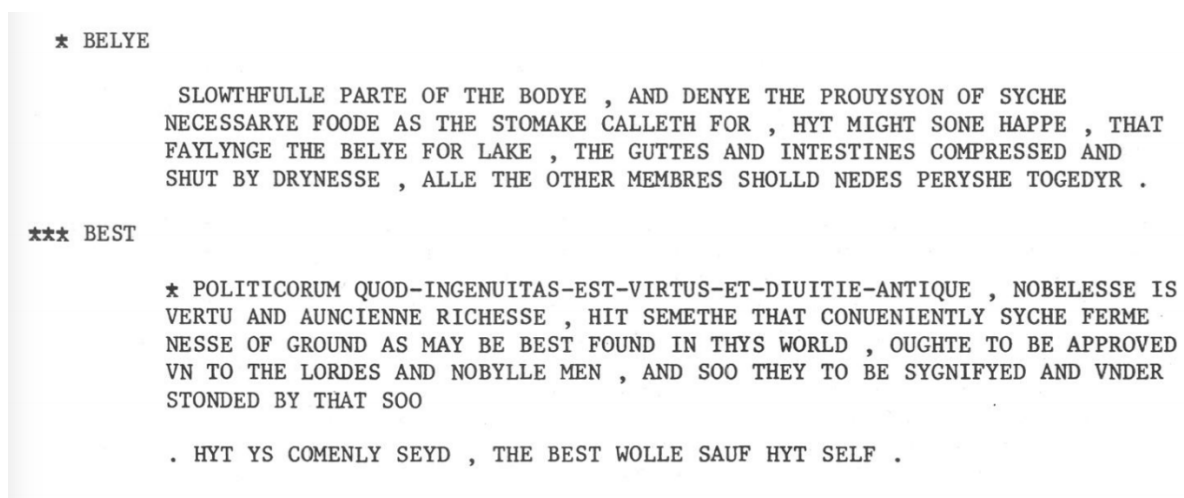
Un autre type de document de sortie du programme ALINE doit encore être présenté pour éclairer l'apport des méthodes de la linguistique et de leur automatisation

³⁸⁶ Robin, 130.

³⁸⁷ Robin, 125.

dans le travail de Genet : les concordanciers. Ces documents se présentent comme une liste d'occurrences d'un terme, sélectionné par le chercheur dans son contexte textuel, c'est-à-dire aux côtés des termes qui le précèdent et qui le suivent. Le programme ALINE permet, par exemple, de sélectionner le nombre de mots qui entourent le mot-pôle choisi (entre 5 et 30 mots)³⁸⁸.

Figure 14. Exemple d'un concordancier produit à l'aide de ALINE.



Source : Fossier et Genet, « Un type d'exploitation : le traitement de texte », 8.

3.3.1. Comment attribuer un sens à un mot ?

Pour analyser plus précisément les usages de ces concordanciers dans les raisonnements de Genet, il convient à présent de nous pencher sur un article publié en 1979 dans le bulletin de liaison *Le Médiéviste et l'ordinateur*, dans lequel Genet aborde le type de raisonnement fondé sur ces documents de travail. Il est possible, sur la base de ce texte, et en l'absence d'articles présentant des résultats dérivant de cette approche, de dégager schématiquement la démarche que suppose ALINE. Dans un premier temps, les utilisateurs établissent, à l'aide des indices de fréquences relatives et absolues, quels sont les mots les plus utilisés par les auteurs. Dans un second temps, ils analysent les différents emplois de ces mots-cibles en contexte, à l'aide des concordanciers. C'est cet aller-retour entre des indicateurs quantitatifs et les concordances qui est au centre de la méthode de

³⁸⁸ « Programme de traitement automatique des textes pour l'équipe de recherche sur l'histoire politique du bas Moyen-Âge », document interne à l'ERA 713, non daté.

la lexicométrie.

Ce type de raisonnement est remis en cause au début des années 1970 par certains acteurs de ce champ de recherche. Deux critiques principales sont mises en avant par les linguistes et les historiens pour en illustrer les limites. Dans un article de 1977, publié dans les actes du colloque « Histoire médiévale et informatique »³⁸⁹, Genet expose dans un premier temps les difficultés d'interprétation que rencontrent les chercheurs qui ont recours aux seules fréquences pour leurs analyses. Pour ce faire, il prend l'exemple d'une critique formulée par le linguiste Frédéric François (1935-2000) à propos travail de son collègue Jean Roche, sur l'utilisation des pronoms personnels par les candidats aux élections présidentielles de 1969.

Dans ce travail, Roche remarque que les candidats de gauche utilisent majoritairement le « nous », tandis que les discours des candidats de droite présentent une majorité de « je ». Une exception lui pose cependant problème : le candidat de gauche Jacques Duclos utilise majoritairement le « je ». Pour expliquer ce phénomène, Roche propose plusieurs interprétations de « *bon sens* »³⁹⁰ : Duclos a une forte personnalité ; Duclos utilise le langage de la droite pour se faire élire ; Duclos est, au fond, un homme de droite. Le passage d'une approche formalisée à une interprétation par le « *bon sens* » soulève selon Genet le problème suivant :

*« le traitement statistique d'un texte peut-il se faire à partir d'une seule unité, le mot ? »*³⁹¹

La seconde critique soulevée par Genet à l'encontre de ce type d'approche complète la première, mais concerne plus directement les approches sémantiques qui nous intéressent ici. Elle concerne un problème théorique qui repose sur le problème de la « transparence » de l'usage du mot. La plus belle manifestation de ce débat chez les historiens se trouve dans la critique que Robin adresse, dans son livre de 1973, au travail

³⁸⁹ Jean-Philippe Genet, « Ordinateur, lexique, contexte », *Publications de l'École Française de Rome* 31, n° 1 (1977): 297-317. Pour une analyse de l'importance de ce colloque dans le développement des utilisations des ordinateurs en Europe voir le troisième chapitre de cette thèse.

³⁹⁰ Genet, 301.

³⁹¹ Genet, 300.

de Prost sur les proclamations électorales de 1881, 1885 et 1889³⁹². Genet la résume ainsi, dans son article de 1977 :

*« elle [Robin] accuse [Prost] de faire du mot l'indice d'un comportement politique, en lui accordant une place exagérée, en excluant délibérément tout ce qui est extérieur au texte »*³⁹³.

Ce que reproche Robin à son collègue, c'est de prendre un mot comme indice, et d'assumer son univocité de sens. Au contraire, Robin, comme d'autres spécialistes comme Jean Dubois (1920-2015) ou Michel Pêcheux (1938-1983), propose de dépasser cette approche en étudiant le sens des mots en fonction de leur place dans une phrase, dans un texte, dans un corpus.

Genet est lui aussi convaincu par cette critique, et met en garde l'historien qui analyse les fréquences des mots de ne pas « *se contenter d'un dangereux « raccourci théorique et méthodologique* » »³⁹⁴. Cette erreur commune repose, selon lui, sur un manque de formation à la linguistique des historiens :

*« N'ayant pas une formation spécifique en ce qui concerne les problèmes de la syntaxe et de la sémantique, ils portent leur attention sur le lexique. D'où une tentation qui est de postuler la « transparence » du mot »*³⁹⁵.

Pour éviter cette erreur théorique, et « *sortir* » du ghetto du texte³⁹⁶, Genet propose dans son article de s'inspirer d'un nouvel ensemble d'approches développées pour l'étude des textes politiques développées par des linguistes à partir de la deuxième moitié des années 1960 : l'analyse de discours. À ce titre, Genet rappelle notamment la

³⁹² Prost, *Vocabulaire des proclamations électorales de 1881, 1885 et 1889*.

³⁹³ Genet, « Ordinateur, lexique, contexte », 301.

³⁹⁴ Genet, 301.

³⁹⁵ Genet, 301.

³⁹⁶ Genet, 301. L'expression utilisée ici par Genet n'est pas attestée, à notre connaissance, dans d'autres contextes. Elle renvoie cependant à un type de démarche répandu parmi les linguistes au milieu des années 1970, qui consiste, d'une part, à opposer au texte autonome la valeur du texte intégré à un corpus considérée en tant que construction et, d'autre part, à la délinéarisation du texte en énoncés élémentaires susceptibles d'être mis en paraphrase, approche notamment développée par Pêcheux.

définition que donnent deux linguistes, Gérard Gayot (1941-2009) et Pêcheux, de cette nouvelle notion en 1971 :

« le terme discours désigne un matériel historique en tant qu'il est relié à ses conditions de production. On dira que, les conditions de production d'un discours étant fixées, il leur correspond un processus de production de ce discours, ou processus discursif. Effectuer l'analyse sémantique d'un matériel reviendra donc... à décrire méthodiquement le processus discursif dont ce matériel est la réalisation »³⁹⁷.

L'une de ces méthodes, qui vise à décrire le processus discursif, implique justement le recours aux ordinateurs, tout en étant opposé aux analyses quantitatives du types de celles proposées par les lexicomètres. Il s'agit de l'Analyse Automatique de Discours (AAD), développées par Pêcheux à partir de la fin des années 1960³⁹⁸, et qui représente la pointe de ce programme de recherche. Pêcheux et son équipe ont même élaboré un programme informatique, disponible au centre de calcul de l'université de Bologne³⁹⁹, qui contribue à l'élaboration de ce que Pêcheux appelle, dans un article plus tardif, « *l'objectivisme quantitatif* »⁴⁰⁰.

Sans qu'il soit utile de rentrer dans le détail de ces travaux, on remarque que, de la même manière que pour le cas de la lexicométrie, on retrouve, dans cet autre programme d'études linguistiques, la volonté d'utiliser les outils de calcul pour objectiver l'analyse sémantique. La principale différence entre analyse de discours et lexicométrie sur ce point vient du fait que l'analyse du discours à la Pêcheux ne repose pas sur des indicateurs quantitatifs, mais sur des critères syntaxiques et discursifs, et le découpage des discours en « *énoncés élémentaires* »⁴⁰¹. En dépit de l'intérêt qu'il porte au programme informatique développé par l'équipe de Pêcheux, Genet exclut la possibilité

³⁹⁷ Genet, 302

³⁹⁸ Voir Michel Pêcheux et al., « Présentation de l'analyse automatique du discours (AAD69) : théories, procédures, résultats, perspectives », *Mots. Les langages du politique* 4, n° 1 (1982): 95-123.

³⁹⁹ Cette information est précisée par Genet dans son article. Genet, « Ordinateur, lexic, contexte », 302.

⁴⁰⁰ Michel Pêcheux et al., « Présentation de l'analyse automatique du discours (AAD69) : théories, procédures, résultats, perspectives », 96.

⁴⁰¹ On trouvera une explication de cette procédure et des exemples dans Michel Pêcheux et al.

de recourir à cet instrument de travail, en raison de la difficulté que représente le découpage en énoncé élémentaire :

« il faut voir que tout le travail de transformation des phrases en énoncés canoniques doit être fait manuellement : or, il requiert des compétences linguistiques qui en rendent l'emploi délicat pour les historiens »⁴⁰².

Les limites des compétences en linguistique des membres de l'ERA 713 vont donc pousser Genet à se tourner vers une autre méthode, l'analyse automatique des contextes, qui doit elle aussi permettre de dépasser l'obstacle théorique soulevée par Robin en 1973⁴⁰³.

3.3.2. Analyse automatique des contextes.

Dès les premières années de la conception du programme ALINE, et plus probablement au cours de l'année 1974, Genet envisage de faire programmer un module d'analyse automatique des contextes. Dans une note de bas de page de la troisième partie de son article de 1977, Genet indique explicitement le modèle dont il s'inspire. Il s'agit d'une des méthodes développées par une équipe de linguistes de Saint-Cloud dans un ouvrage collectif publié en 1975, *Des Tracts en Mai 68. Mesures de vocabulaire et de contenu*⁴⁰⁴.

Le principe de ce type d'analyse est de calculer les fréquences de co-occurrences des mots situés dans l'entourage d'un mot-pôle choisi. Un test statistique, appelé test z, permet de différencier les mots voisins du mot-pôle choisi des termes co-occurents à ce mot-pôle. Schématiquement, ce test

⁴⁰² Genet, « Ordinateur, lexique, contexte », 302.

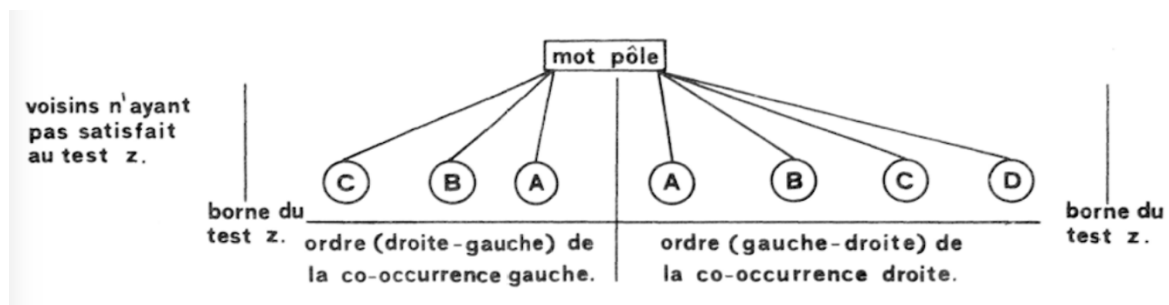
⁴⁰³ Il faut faire remarquer que les spécialistes de l'analyse de discours font, au milieu des années 1970, leur propre auto-critique, et notamment sur le plan pratique, en réfléchissant aux moyens de faciliter la phase de pré-codage, qui nécessitait alors l'aide d'un spécialiste de la grammaire.

⁴⁰⁴ Michel Demonet, Annie Geffroy, et Jean Gouazé, *Des tracts en mai 68 : mesures de vocabulaire et de contenu* (Paris, France: Fondation nationale des sciences politiques : A. Colin, 1975).

« permet de montrer que l'item en question intervient plus fréquemment dans le contexte du mot pôle que ne le laisserait prévoir son nombre d'occurrences total dans le texte. »⁴⁰⁵

On peut alors construire la table de co-occurrences qui permettra de construire divers graphes en vue de comparer l'utilisation de ces mots co-occurents au sein du même texte, comme d'un texte à l'autre. Pour ce faire, un certain nombre d'indices complexes doivent être calculés : le coefficient de proximité, le coefficient de voisinage, l'écart des co-fréquences et le coefficient de co-occurrence. La finalité de la méthode est de construire un nouveau type de représentation des relations entre les mots-pôles et les termes qui leurs sont co-occurents : le lexicogramme.

Figure 15. Schéma d'un lexicogramme.



Source : Genet, « Ordinateur, lexique, contexte », 305.

Deux éléments de nature tout à fait différente sont ici particulièrement intéressants pour notre sujet. Le premier c'est que la programmation du module d'analyse automatique des contextes de ALINE n'a pas encore commencé en 1977, et que Genet présente donc dans cet article des résultats de calculs d'indices faits à la main. L'intérêt pour les méthodes de l'analyse automatique des contextes n'est donc pas seulement lié à son automatiser, et à la possibilité de traiter de plus grandes quantités de données, mais bien directement au raisonnement que cette méthode suppose. L'enjeu est en effet d'introduire de la mesure dans l'analyse des contextes, là où l'usage des concordanciers restait dans une perspective exclusivement qualitative.

⁴⁰⁵ Genet, « Ordinateur, lexique, contexte », 303.

Le deuxième élément concerne plus directement le fait que Genet ait choisi cette méthode, et non une autre. En effet, le médiéviste emprunte ici l'une des innovations méthodologiques toute récente de la recherche en linguistique pour dépasser les limites qu'il perçoit dans les méthodes des historiens des idées politiques. Genet se nourrit ainsi d'emprunts directs aux pratiques les plus récentes, qui influencent son positionnement théorique.

Ces différentes opérations sur les mots prévues par le programme ALINE (calcul des fréquences, établissement de concordances et analyse automatique des contextes) nous renseignent sur le modèle linguistique sur lesquels se base Genet. Avec la conception de ce programme, et dans les conditions matérielles qui sont celles de son enquête, Genet essaie de s'approprier les techniques les plus pointues de la lexicométrie développées à Saint-Cloud. Les bénéfices que les linguistes dérivent du recours aux ordinateurs paraissent ainsi être transposés à l'histoire médiévale : gain de temps pour l'établissement des index et des concordances, systématisation/objectivation des procédés d'élaboration de ces documents de travail indispensables, possibilités d'établir des indices ou des coefficients plus fins et plus complexes, et nouvelles formes de représentations des relations entre les mots d'un corpus.

Nous pouvons cependant affiner notre analyse en montrant que ces apports sont en fait de trois natures différentes qu'il faut ici distinguer. Il existe, tout d'abord, des apports que nous avons qualifiés, dans le chapitre précédent, d'apports « quantitatifs ». Le recours aux ordinateurs permet d'exécuter, à une nouvelle échelle, une série d'opérations qui préexistaient à l'utilisation de ces machines : 1) établissement d'index sur des corpus de dimension plus grande et 2) établissement d'indices ou de coefficients plus rapidement. Cependant, ces apports ne modifient pas réellement les raisonnements des acteurs par comparaison avec ceux mis en œuvre sans ordinateur, puisqu'ils s'inscrivent dans la continuité de méthodes qui préexistent à leur automatisation.

Ensuite, l'emploi d'ordinateurs modifie le rapport que les acteurs entretiennent avec les données, en accroissant la confiance qu'ils leur accordent. Comme nous l'avons vu dans le chapitre précédent, le recours aux ordinateurs permet également, pour les projets de traitement automatique des textes, de traiter les données avec plus de fiabilité, en s'assurant par-là de l'absence d'erreurs dans les opérations de calcul et de classement. C'est d'ailleurs certainement à ce type de phénomènes que renvoyait la citation

précédente de Robin sur l'intérêt des ordinateurs pour « *objectiver et de systématiser* » la constitution des index. Par-delà ses apports quantitatifs, il apparaît que le recours aux ordinateurs permet également un bond qualitatif dans la réalisation de ces opérations, en raison de l'absence d'erreur que l'ordinateur garantit au moment des comptages.

Le troisième type d'apport des ordinateurs aux raisonnements historiques que nous avons mis en lumière dans cette sous-partie tient quant à lui d'un tout autre ordre : ces nouveaux instruments permettent la transformation d'un certain nombre d'opérations de calculs. Pour les linguistes de Saint-Cloud, les moyens électroniques avaient notamment rendu possible le calcul d'une nouvelle série de coefficients permettant de caractériser les textes (en fonction de la richesse du vocabulaire, etc.). De la même manière, les ordinateurs autorisent la mise en œuvre de la méthode d'analyse automatique des contextes, par la possibilité qu'ils offrent de réaliser le test z dans des corpus de grande dimension et le calcul des indices complexes qui caractérisent les relations entre les différents mots du texte. Ce troisième type d'apport, qui repose sur les capacités de calcul et de mémoire des ordinateurs, ainsi que sur la possibilité de programmer des suites d'opérations complexes, constitue ce que nous appellerons désormais les apports qualitatifs de l'utilisation des ordinateurs.

Toutefois, pour saisir plus précisément la nature de ce troisième type d'apport, fondamentale pour notre sujet, il nous faut nous tourner plus précisément vers d'autres raisonnements mis en œuvre par les acteurs. En effet, dans le cadre du travail de Genet, l'évaluation des coefficients ne peut être pris comme exemple puisque les calculs sont faits à la main, et qu'aucune archive n'atteste de la réalisation de ce module d'analyse automatique des contextes dans le package ALINE. L'exemple des raisonnements mis en œuvre par Genet à l'aide de ALINE dans son article de 1984 constitue ainsi l'unique source à même de nous permettre de saisir précisément les apports qualitatifs de l'utilisation des ordinateurs, à travers la combinaison des opérations d'établissement des fréquences précédemment décrites et de d'un autre type de procédure d'analyse des données : l'analyse factorielle des correspondances.

3.4. Approche classificatoire et analyses factorielles des correspondances.

Dans son article de 1984, intitulé « Automatic Text Processing and Factorial

Analysis: a Method for Determining the Lexicographical Horizon of Expectation », Genet présente l'étude d'un corpus de six textes politiques anglais. Comme il l'affirme dans son introduction, la finalité de cette recherche doit être comprise en rapport avec une méthode définie par l'historien de la littérature allemand Hans Robert Jauss (1921-1997) dans un livre de 1967⁴⁰⁶, que Genet cite sous la forme suivante :

« (1) *To understand how a literary work has been received, one must be able to reconstitute the « horizon of expectation » of the public at a given time (i.e., the previous experiences in the literary genre of the work, discrepancies between poetical and practical languages, etc.)*

(2) *The gap between the « horizon of expectation » and the actual work is the « aesthetically deviation », which must be measured by the study of the reactions of the public. »*⁴⁰⁷

Afin de poursuivre cette approche pour les textes politiques anglais de la fin du Moyen-Âge, Genet se propose, dans la continuité d'un autre médiéviste, lui aussi inspiré par Jauss, Stephen G. Nichols (1936-), de se concentrer sur les « *basic constituents* »⁴⁰⁸ des textes, et en particulier sur la fréquence des mots⁴⁰⁹. Cette approche doit permettre, selon Genet, de reconstituer « *the lexicographical horizon of expectation* » des textes de son corpus.

Pour ce faire, Genet affirme qu'il est nécessaire d'avoir soit un bon dictionnaire, ce qui, nous l'avons vu, n'existe pas pour les langues des textes qu'il étudie, soit un ordinateur. Il présente ensuite, dans la première partie de son article, les fonctionnalités du programme ALINE, sur lesquelles nous ne reviendrons pas. Dans la seconde partie, qui nous intéresse ici plus particulièrement, Genet propose :

⁴⁰⁶ Hans Robert Jauss, *Literaturgeschichte als Provokation der Literaturwissenschaft* (Konstanz, Allemagne: Im Verlag der Druckerei und Verlagsanstalt Konstanz, Universitätsverlag, 1967). La date donnée par Genet dans sa bibliographie est celle de la quatrième réédition en allemand, publiée en 1974. Une traduction française a par ailleurs été publiée en 1978 : Hans Robert Jauss, Claude Maillard, et Jean Préfacier Starobinski, *Pour une esthétique de la réception* (Paris, France: Gallimard, 1978).

⁴⁰⁷ Genet, « Automatic text processing and factorial analysis », 148.

⁴⁰⁸ Genet, 149.

⁴⁰⁹ Voir, plus particulièrement, la référence citée par Genet dans son article : Stephen G. Nichols, « A Poetics of Historicism? Recent Trends in Medieval Literary Study in Transformation and Continuity », *Medievalia et Humanistica Cleveland, Ohio*, n° 8 (1977): 77-101.

« a method for text classifying based upon the discrepancies between the occurrences of the word in the texts »⁴¹⁰.

Pour mener cette étude, Genet s'appuie sur des textes qui présentent de nombreuses différences les uns par rapport aux autres. Ces différences portent à la fois 1) sur leurs dates de rédactions, qui se situent entre le XIII^e et le XV^e siècle ; 2) sur le fait que les textes abordent des sujets variés, quoi que tous liés au politique, comme les relations entre les pouvoirs spirituel et laïque, la bonne gouvernance du prince ou des problèmes moraux liés aux pouvoirs politiques et 3) sur le genre des textes étudiés (sermon, discours, pamphlet).

3.4.1. Mise en œuvre des analyses factorielles des correspondances.

La première opération que Genet applique à ce corpus consiste à établir automatiquement, à l'aide du programme ALINE, la liste des termes avec leur fréquence. Suivant la méthode que nous avons présentée plus haut, il isole alors les 40 mots les plus fréquents du corpus (les mot-thèmes), et dresse un tableau de leurs nombres d'occurrences dans chacun des six textes. C'est à partir de ce tableau qu'il va développer deux méthodes de classification, qui reposent sur deux procédures développées en analyse des données : l'analyse factorielle de correspondances et la classification automatique. Le choix de ces méthodes représente à lui seul une innovation par rapport aux travaux des lexicométriciens de Saint-Cloud, puisque ces derniers étudient la répartition des termes dans un corpus à l'aide d'une autre démarche, basée sur l'établissement de tableaux de contingence et la réalisation d'un test du khi-deux⁴¹¹.

Deux remarques sur les modalités de circulation de ces méthodes permettent d'éclairer pourquoi Genet choisit cette option. La première dérive du fait, nous l'avons dit, que l'étude de Prost constitue un modèle méthodologique pour Genet. Or, Prost est justement l'un des premiers historiens en France à utiliser les analyses factorielles des

⁴¹⁰ Genet, 149.

⁴¹¹ Cette méthode, sur laquelle nous ne revenons pas ici, est détaillée dans Robin, *Histoire et linguistique*, 130-31.

correspondances pour l'étude de données textuelles, à la fin des années 1960. Pour saisir comment cette procédure d'analyse des données fait irruption dans les bagages de cet historien, il est nécessaire de revenir dans un premier temps sur les raisons de la collaboration entre le statisticien Demonet et l'historien Prost.

L'analyse des données est une branche des statistiques qui se développe notamment en France sous l'impulsion du mathématicien Jean-Paul Benzécri (1932-2019). Au milieu des années 1960, à une date difficile à préciser, Prost rencontre Benzécri au cours d'un voyage en train entre Orléans et Paris. Prost envisage alors de recourir aux ordinateurs pour mener ses recherches, et Benzécri lui propose un étudiant pour l'aider dans un projet d'étude de ce type : Michel Demonet⁴¹². Dans la boîte à outils statistiques de Demonet se trouvent alors plusieurs méthodes d'analyse des données développées par son professeur, et notamment, l'analyse factorielle des correspondances. C'est ainsi que, dans le cadre de l'enquête de Prost sur les proclamations électorales, des méthodes d'analyse des données sont introduites pour l'étude de textes politiques⁴¹³.

On peut déceler une autre raison pour rendre compte du fait que Genet se tourne vers ces méthodes. En effet, ALINE ne permettait pas de mettre en œuvre des analyses factorielles de correspondances. Cela n'était cependant pas nécessaire puisque certains disciples de Benzécri mettaient à disposition, depuis le milieu des années 1960, des programmes d'analyse des données⁴¹⁴. Au début des années 1980, l'accès à ces programmes est même grandement facilité par la création de l'Association pour le Développement et la Diffusion de l'Analyse des Données (ADDAD), qui met à la disposition des chercheurs une bibliothèque de programmes informatiques permettant l'application de ces méthodes. Pour son article de 1984, Genet profite justement de ces programmes de l'ADDAD pour réaliser des analyses factorielles des correspondances et des opérations de classification automatique sur la base des index de fréquences fournis par ALINE.

Ces éléments de contexte nous renseignent sur l'articulation entre différents programmes informatiques dans la mise en œuvre de la démarche de classification des

⁴¹² Cette anecdote nous a été racontée par Demonet, dans un entretien que nous avons mené auprès de lui en 2017.

⁴¹³ Prost, *Vocabulaire des proclamations électorales de 1881, 1885 et 1889*.

⁴¹⁴ Sur ce sujet voir, par exemple, Philippe Bonnet, *Chapitre 2. Pour une histoire sociale de l'analyse des données, La méthodologie de Pierre Bourdieu en action* (Dunod, 2015).

textes qu'élabore Genet. La mise en œuvre des analyses factorielles des correspondances repose ainsi sur les tableaux lexicaux des 40 mots-thèmes obtenus à l'aide de ALINE (nombre d'occurrences, fréquence absolue, fréquence relative). La première étape de cette procédure d'analyse des données consiste ensuite à construire, pour chacun de ces tableaux, un tableau des profils pour chaque ligne (les mots) et pour chaque colonne (les textes) du tableau. Ces profils peuvent être considérés comme un indice du poids relatif de chaque mot, dont la procédure d'analyse factorielle conduira à calculer les écarts par rapport à un profil moyen, qui correspond à la situation d'indépendance entre chacune de ces lignes et de ces colonnes⁴¹⁵.

Pour construire ces profils-lignes et ces profils-colonnes, plusieurs méthodes de transformation existent. La plus répandue est aujourd'hui celle qui consiste à construire le tableau des écarts à l'indépendance, qui résulte de la différence entre le tableau théorique, qui correspondrait à la situation d'indépendance entre les variables en lignes (ici les mots) et les variables en colonnes (ici les textes), et le tableau des valeurs observées.

Dans le cas de l'article de Genet de 1984, la procédure de transformation des valeurs du tableau lexical en profils repose sur un autre calcul. Genet y indique que ce choix demeure à la discrétion du statisticien qui met en œuvre l'analyse factorielle des correspondances, ce qui laisse entendre qu'une collaboration avec un spécialiste s'impose :

*« It must be stressed that there are several possible other transformations and that a skillful statistician will be able to select the formula which will suit the data best »*⁴¹⁶.

Sans qu'il soit nécessaire de revenir précisément sur les modalités de sélection de cette formule, on saisit, ici encore, les limites des compétences de l'historien dans la mise en œuvre de procédures statistiques complexes.

⁴¹⁵ Pour une explication plus poussée de cette procédure, et une définition de certaines de ces notions clefs, comme l'indépendance ou l'inertie, nous conseillons de se référer à Philippe Cibois, *L'analyse factorielle: analyse en composantes principales et analyse des correspondances* (Paris, France: Presses universitaires de France, 2000).

⁴¹⁶ Genet, « Automatic text processing and factorial analysis », 160.

Une fois les profils établis, le rôle de l'analyse factorielle des correspondances est de définir les positions de ces profils sur des axes que le programme informatique permet de dégager. La finalité de ce type de méthode est de visualiser des différences entre les profils-lignes et les profils-colonnes de manière géométrique, à travers la construction d'un nuage de points des profils, et de rendre compte de leurs coordonnées qui sont responsables des plus grandes différences. Pour ce faire, la construction suit un principe théorique simple : le profil moyen est placé dans ce nuage et sera considéré comme son origine. S'il y a indépendance entre les variables considérées (ici les mots-thèmes qui apparaissent dans les textes), tous les profils seront confondus avec le profil moyen, et le nuage sera donc réduit à un point : l'origine des axes. Au contraire, plus les données s'écartent de la situation d'indépendance, plus les profils s'écarteront de cette origine.

La difficulté de ce type de représentation réside dans le fait que l'enjeu de cette analyse est de comparer toutes les variables en même temps. Le nuage est donc nécessairement multidimensionnel. Pour le représenter, les statisticiens qui ont développé cette méthode ont opté pour une projection sur un plan à deux axes. Il en existe cependant une infinité. Une nouvelle procédure de calcul permet ainsi d'établir lesquelles de ces projections sont à même de représenter le maximum d'informations du nuage de points. C'est là qu'interviennent nécessairement des outils de calculs électroniques (ordinateurs, calculatrices programmable). Ils permettent de calculer une suite d'axes orthogonaux, classés de façon à rendre compte de l'inertie⁴¹⁷ de façon décroissante.

Un programme d'analyse factorielle donne à son utilisateur un tableau des valeurs propres, qui lui indique les axes offrant le meilleur taux d'inertie de représentation du nuage sur un plan à deux axes. Il est alors possible de demander au programme la position de chaque profil-colonne et de chaque profil-ligne sur les deux axes ainsi définis. Pour ce faire, l'une des particularités de l'analyse factorielle est encore de proposer comme méthode de calcul des distances, non la distance euclidienne, mais la distance du khi-deux. Il est dès lors possible de représenter le nuage des points-colonnes et de points-lignes sur le même plan. Nous le verrons dans le cas de l'analyse de Genet, les analystes multiplient alors les représentations en faisant varier le choix des axes, ce qui permet de vérifier si la représentation du premier plan, d'inertie maximum, est bonne, et de découvrir si, en changeant de plan, la disposition des points-colonnes, c'est-à-dire la

⁴¹⁷ Cibois, *L'analyse factorielle*.

forme du nuage, révèle de nouvelles différences entre les profils-lignes et les profils-colonnes.

3.4.2. Exploitation des graphiques.

Dans le cas de l'étude de Genet, quatre graphiques d'analyses factorielles sont publiés. Deux sont construits sur la base du tableau des occurrences des 40 termes les plus fréquents des six textes du corpus, et deux autres sur le tableau des fréquences relatives de ces termes. Pour chacun de ces deux tableaux, deux combinaisons d'axes factoriels sont prises en compte (axes 1 et 2, puis axes 1 et 3). À eux trois, ces axes représentent 85,8% de l'inertie (71,6% pour les deux premiers et 14,2% pour le troisième), et permettent donc de visualiser les principales oppositions qui existent dans le jeu de données.

L'analyse du second graphique, basé sur le tableau des occurrences des 40 termes, et reproduit ci-dessous (figure 16) conduit Genet à formuler la conclusion suivante :

*« It is obvious in Figure 4 [notre figure 16], based upon the numbers of occurrences of the words, that the chief opposition in this group of texts is between the two tracts with predominantly religious occupations – the Dialogus [représenté par le sigle **DMC** sur le graphique] and the Tractacus [représenté par le sigle **TDR** sur le graphique] – and all the others. These two tracts also appear to be closely associated with maximum values for words such as « Christ », « law », « power », « God », « man », « manner » and « reaume » (kingdom) »⁴¹⁸.*

La procédure d'analyse factorielle conduit Genet à identifier, sur la base du nombre d'occurrences des mot-thèmes du corpus, une première opposition entre, d'une part, deux traités et, d'autre part, le reste des textes du corpus.

⁴¹⁸ Genet, « Automatic text processing and factorial analysis », 162.

Figure 16. Graphique d'analyse factorielle basée sur le tableau du nombre d'occurrences (axes 1-2).

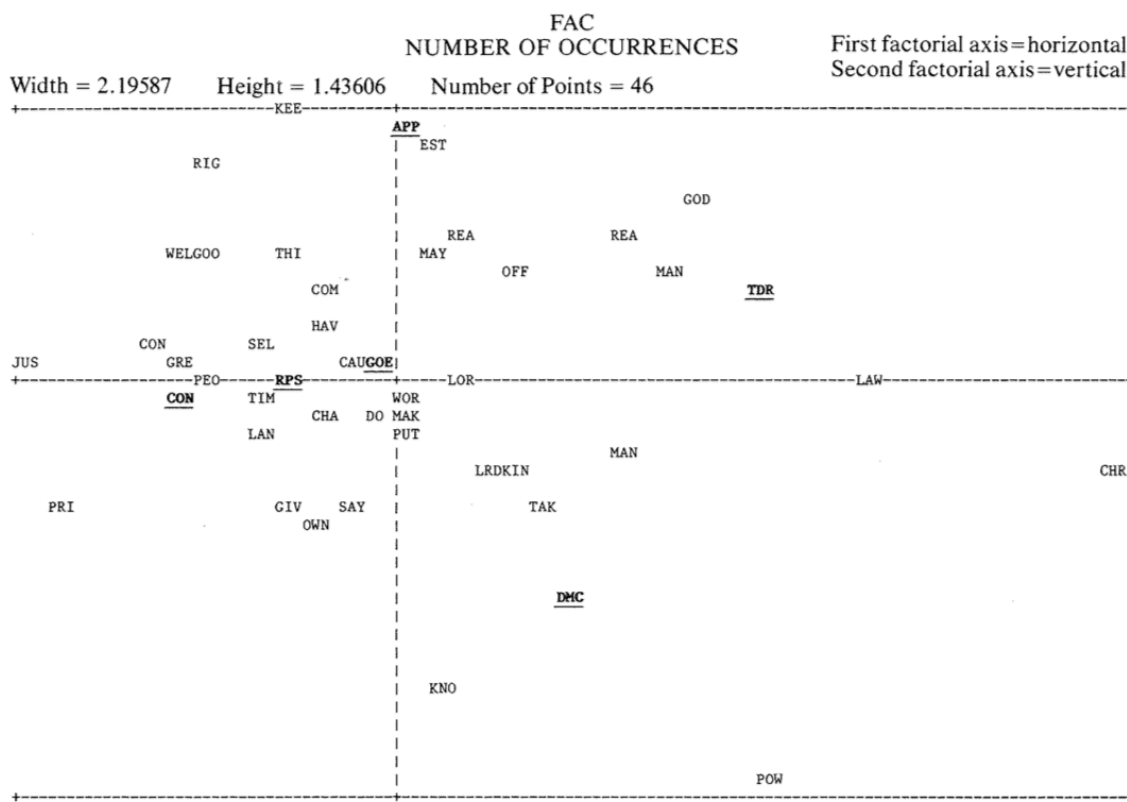


Figure 4

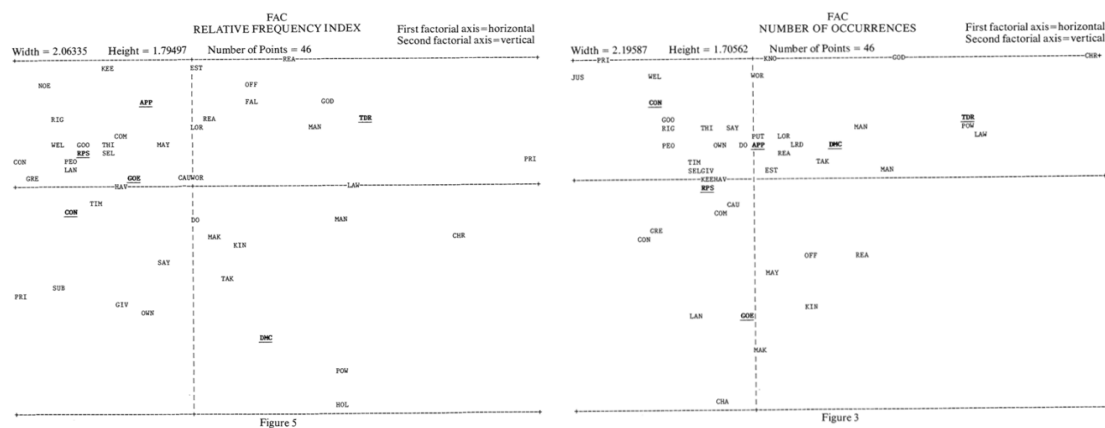
Source : Genet, « Automatic text processing and factorial analysis », 169.

L'analyse d'un autre graphique, basé sur le tableau des indices de fréquences relatives, et reproduit ci-dessous (graphique de gauche sur la figure 17), confirme cette opposition majeure entre ces deux groupes de textes au sein corpus. Cependant, au contraire du graphique précédent, les oppositions mises en lumière par le deuxième axe sont d'une autre nature. Genet commente cette seconde visualisation en ces termes :

« The second axis is completely different from that in Figure 3 [graphique de droite sur la figure 17], since the Governance of England [**GOE**] is no longer alone at the negative end of the axis but exactly at the middle of the axis, not far away from The III Consideracions [**CON**] and Russell's parliamentary sermon [**RPS**]. It is the Dialogus [**TDR**] which stands alone opposite all the other tracts. »⁴¹⁹

⁴¹⁹ Genet, 163.

Figure 17. Deux graphiques d'analyse factorielle des correspondances basé sur deux types de données (tableau des fréquences relatives à gauche et tableau du nombre d'occurrences à droite).



Source : Genet, « Automatic text processing and factorial analysis », 170.

La comparaison entre de ces deux graphiques permet à Genet de dégager l'existence d'une opposition fondamentale entre le *Dialogus* et le *Tractacus* d'une part, et les autres textes d'autre part, mais aussi de nuancer la ressemblance entre les deux textes qui paraissent s'opposer aux autres :

« In fact, the change of data had made clear that the two religious tracts, though still opposed to all the others, are themselves quite different and that the originality of Fortescue's vocabulary (and concepts) should be examined more seriously »⁴²⁰.

Ce raisonnement, à visée classificatoire, basé sur les graphiques d'analyse factorielle des correspondances, appelle plusieurs remarques importantes pour la problématique qui nous intéresse.

La première concerne la fonction heuristique de cette procédure d'analyse des données, dans la mesure où cette démarche permet à Genet d'observer sous un nouveau jour les textes de son corpus. Contrairement aux raisonnements analysés dans le chapitre précédent, le recours aux analyses factorielles des correspondances ne permet pas seulement d'affiner la description des sources, mais également de révéler à l'historien des

⁴²⁰ Genet, 163.

oppositions entre les textes, qui suscitent de nouvelles questions pour la suite de la recherche (ici sur les concepts de Fortescue par exemple). L'analyse factorielle est ici utilisée comme un outil d'exploration des mesures basées sur le vocabulaire des textes, et donne une image du champ politique sur la base de ce petit corpus. Cette application des analyses factorielles en vue de proposer une représentation du champ étudié par l'auteur fait écho à l'utilisation d'un autre type d'analyse factorielle, l'Analyse des Correspondances Multiples (ACM), par Pierre Bourdieu (1930-2002) dans des études contemporaines de celles de Genet.

En effet, comme nous l'avons fait remarquer auparavant, Genet s'inspire grandement du travail de Bourdieu en lui empruntant notamment son concept de *champ*. En cherchant à poursuivre l'étude du champ politique, Genet s'empare également de la procédure d'analyse des données centrale de son programme sociologique : les analyses factorielles. Comme ont pu l'établir de nombreux observateurs, il existe en effet une corrélation entre la conception de *champ* proposée par Bourdieu à cette période et la représentation en deux dimensions que produit l'analyse factorielle. Pour le sociologue Julien Duval, spécialiste de Bourdieu, les analyses factorielles représentent :

*« la production de représentations visuelles des espaces que sont les champs et l'espace social »*⁴²¹.

Si Bourdieu a utilisé majoritairement un autre type d'analyse factorielle, les analyses de correspondances multiples (ACM), il semble cependant qu'on retrouve chez Genet la même corrélation entre outil et concept.

La deuxième remarque porte sur le fait que ces visualisations produites par la procédure d'analyse factorielle se fonde sur un corpus de textes défini au préalable par le chercheur. Comme dans l'étude du *catasto*, on remarque que l'analyse ne se fait plus à travers une comparaison entre un ensemble d'objets d'étude dont les limites resteraient floues, mais au sein d'une unité d'observation : le corpus préalablement construit. Cependant, là où les 60 000 déclarations du *catasto* représentaient, pour Klapisch-Zuber et Herlihy, un corpus homogène du fait de ses conditions de production par

⁴²¹ Julien Duval, « L'analyse des correspondances et la construction des champs », *Actes de la recherche en sciences sociales* N° 200, n° 5 (2013): 110-23.

l'administration florentine, Genet constitue quant à lui un corpus dont l'homogénéité n'est pas garantie par des raisons historiques, mais par les modalités de la construction de son objet d'étude.

Toutefois, de la même manière pour ces deux projets, les mesures mises en œuvre sur les données ne peuvent être intelligibles qu'au sein de ces corpus. Par exemple, aucun texte extérieur ne peut être comparé avec ceux du corpus de Genet, à moins qu'il n'y soit justement intégré, ce qui modifierait de fait non seulement la liste des mots-thèmes mais encore la position des profils-lignes et des profils-colonnes sur les graphiques d'analyses factorielles.

La troisième remarque s'appuie sur l'observation du fait que Genet ne se contente pas d'une analyse factorielle, mais réplique cette procédure pour découvrir des oppositions entre les textes, pour vérifier et contre-vérifier la pertinence de ces oppositions. C'est certainement l'ordinateur qui permet le développement de cette possibilité de répliquer les analyses, déjà observé dans le cas de l'enquête du *catasto*. Ce type de pratique modifie en profondeur l'épistémologie des acteurs, puisqu'il conduit les historiens à affiner la description des objets d'étude par la multiplication des angles d'observations. Dans ce même article de 1984, Genet propose d'ailleurs une autre méthode pour repérer les oppositions entre les profils lexicaux des textes : la classification automatique. Cette volonté de multiplier les représentations des sources atteste d'une autre facette de cette expérimentation dans les pratiques d'observations en histoire, qui semble liée à la possibilité de multiplier les analyses statistiques qu'offrent les ordinateurs.

La quatrième remarque concerne l'utilisation des visualisations, qui revêtent, dans les deux études de cas analysées jusqu'ici, un rôle assez similaire en dépit de leurs natures différentes. En effet, nous avons vu que dans le cas du *catasto*, les cartes avaient pour objet d'observer et de représenter des corrélations entre, d'une part, des variables économiques ou démographiques, et d'autre part des variables géographiques. Dans le cas de cette étude de Genet, les graphiques d'analyses factorielles ont eux aussi une double fonction : observer et représenter.

Sur ce point, certaines différences entre ces deux études doivent toutefois être soulignées. Pour la représentation, la principale distinction se manifeste dans la nature des espaces représentés : les cartes figurent un espace réel, la Toscane, tandis que les

graphiques d'analyse factorielle figurent un espace théorique, celui du champ politique, constitué par l'ensemble des textes qui font partie du corpus. Cette différence fondamentale se corrèle avec les fonctions différentes qui sont dévolues aux visualisations, puisque les cartes permettent à Klapisch-Zuber et Herlihy d'accéder à la réalité historique de la distribution des variables sociales ou démographiques de la région, tandis que les graphiques d'analyses factorielles permettent à Genet d'observer la structure des données de son corpus. Les deux types de visualisations produites par ces deux collectifs ne permettent ainsi pas d'observer le même type d'objet, ce qui atteste d'une différence épistémologique majeure entre ces acteurs, sur laquelle nous reviendrons dans la conclusion de cette thèse.

La combinaison entre ALINE et les logiciels permettant la mise en œuvre d'analyses factorielles représente donc une véritable innovation pour les médiévistes, mais aussi, plus généralement, pour les historiens des idées politiques. Dans un rapport d'activité de l'ERA 713 de 1983, Genet souligne en ces termes les apports de cette méthode pour la recherche :

« la combinaison de ALINE et de l'analyse factorielle fournit en effet un instrument extrêmement précieux au chercheur pour analyser et comparer entre elles des listes de vocabulaires : l'historien peut ainsi, à partir du moment où il travaille sur une collection de textes assez étendue, étudier sur la longue durée l'évolution du stock lexical dans un domaine conceptuel choisi (ou pour l'ensemble du lexique), et ce de manière exhaustive à l'intérieur de son corpus, ce qui n'a jamais encore été entrepris sur une telle échelle »⁴²².

Il justifie ainsi la plupart des remarques formulées plus haut, et atteste qu'il est bien conscient qu'il se joue là une véritable évolution des méthodes de la critique des textes, en lien avec le recours aux ordinateurs.

Pour comprendre plus avant la portée de ces évolutions dans la critique des sources et les pratiques de l'observation historique, il nous faut regarder comment ces méthodes ont circulé et ont été reprises par d'autres chercheurs depuis le début des années 1970.

⁴²² Genet, « Rapport d'activité pour le CNRS », 1983.

4. Réutilisation et devenir de ALINE et de MEDITEXT.

La nature du projet d'étude automatique des textes mis en œuvre par l'ERA 713 aboutit, comme nous l'avons vu, à la réalisation de trois types d'objets dont il est possible de suivre la circulation : le programme ALINE, la banque de textes enregistrés pour l'ordinateur MEDITEXT et quelques articles de recherche. Plusieurs démarches permettent d'étudier cette circulation, puisqu'il est possible de suivre les citations des articles et les réutilisations de MEDITEXT ou de ALINE. Cependant, les sources à notre disposition pour ce faire sont, dans ce cas, d'une nature différente de celles que nous avons pu employer pour le chapitre précédent. Pour étudier la circulation du programme ALINE et de MEDITEXT, les rapports d'activités de l'ERA 713 constituent l'unique documentation disponible. En dépit des limites de ces documents, les informations qu'ils contiennent nous permettent cependant de saisir assez précisément comment ces deux objets ont été utilisés et reçus par des chercheurs au sein de l'ERA 713 comme par des chercheurs extérieurs à ce collectif.

Le problème est plus délicat pour les articles de recherche publiés, qui n'ont, à notre connaissance, et au contraire du livre sur l'enquête du *catasto*, fait l'objet d'aucune recension ou d'aucune critique publiée. L'article de Genet de 1984, « Automatic text processing » est bien cité dans un article de 1993, mais cette référence ne permet pas de saisir la réception des raisonnements de Genet dans ce contexte. La publication de la thèse de Genet sous le titre *La genèse de l'État moderne* en 1997 donne lieu, quant à elle, à des réactions au sein de la communauté des médiévistes, mais elles sont fortement postérieures au contexte qui nous intéresse ici.

Pour comprendre les conséquences des innovations de la méthode de recherche développée dans le cadre du projet ALINE sur l'historiographie, il nous faut nous tourner vers deux éléments qui incarnent le devenir de ces méthodes : leur enseignement et la poursuite des recherches de ce type à l'aide d'autres logiciels, et sur d'autres types d'équipement, les micro-ordinateurs.

4.1. Circulation du programme ALINE.

Dès le début de l'entreprise, ALINE est conçu comme un instrument de travail à

destination des historiens, et en particulier des médiévistes. Dès 1979, nous l'avons vu, les créateurs de ce programme le transforment en package afin d'en permettre une utilisation plus large. Cette première opération permet notamment d'implémenter le programme au CIRCE ce qui va permettre à des chercheurs extérieurs à l'ERA 713 de l'utiliser. Cependant, aucune source ne permet, à notre connaissance, de saisir si cette implémentation a bien donné lieu à des réutilisations via le CIRCE. Une autre collaboration de ce genre est encore prévue dans le rapport pour le CNRS de 1983. Elle consiste à coupler ALINE avec les programmes distribués par l'ADDAD. Là encore, aucune source ne permet à notre connaissance d'affirmer si cette opération a été menée à son terme.

D'autres éléments permettent encore d'attester de la volonté des membres de l'ERA 713, et particulièrement de Genet, de diffuser ce programme. Plusieurs publications, que nous avons croisées au fil de ce chapitre, participent de cette dynamique. Entre 1974 et 1984, Genet publie ainsi quatre articles qui ont pour objet de présenter les fonctionnalités du programme : 1) le premier est publié dans un bulletin spécialisé en analyse de discours, le *Bulletin d'analyse du discours de l'université de Lille*⁴²³, 2-3) deux autres paraissent dans un autre bulletin de liaison à visée informative et didactique, *Le Médiéviste et l'ordinateur*⁴²⁴ et 4) l'article de 1984 est, quant à lui, publié dans les actes d'un colloque qui a eu lieu à Kalamazoo en 1979, haut lieu de rencontre pour les historiens utilisant des ordinateurs dans les années 1970 et 1980. Une autre communication de ce type, n'ayant pas donné lieu à une publication, est encore attestée dans les rapports d'activité. Il s'agit de la présentation du programme ALINE au séminaire, dirigé par le médiéviste Peter Lewis (1931-2014), consacré à la France du XV^e siècle au All Souls College à Oxford en mars 1981.

Cependant, si ces activités attestent de la volonté de Genet de faire circuler ALINE, elles ne suffisent pas pour comprendre comment le programme a effectivement franchi les murs de l'ERA 713. La communication à Oxford a bien donné lieu à la réalisation d'un mémoire de DEA par une étudiante dont il ne nous a pas été possible de retrouver le nom sur la base des informations des rapports d'activité. Il semble toutefois que cette dernière ait essentiellement travaillé à l'université Paris 1 pendant cette période.

⁴²³ Genet et al., « Un programme de traitement automatique des textes ».

⁴²⁴ Les chapitres 4, 5 et 6 de cette thèse sont justement consacrés à la fonction de ce bulletin de liaison au sein de la communauté des médiévistes.

Dans les faits, les rapports d'activités attestent plutôt des limites de la circulation du programme en dehors du laboratoire. Quatre échelles de circulation effective peuvent cependant être dégagées : 1) la première concerne le développement de l'utilisation de ALINE au sein de l'ERA 713, avec le développement de nombreux projets impliquant le programme par des médiévistes rattachés à l'équipe entre 1981 et 1985 ; 2) la seconde regarde l'utilisation de ALINE par des chercheurs extérieurs à l'ERA 713, mais qui sont très vite rattachés à l'équipe au titre de « chercheurs associés » ; c'est par exemple le cas de Marie-Anne de Mercoyrol, nous y reviendrons ; 3) la troisième touche l'utilisation de ALINE pour l'enseignement au sein de l'UER d'histoire de l'université Paris 1 et 4) la quatrième concerne la mention d'une utilisation en dehors des frontières de la discipline, par un membre de l'UER de Science politique de l'université de Paris 1, Pierre Gaborit, à propos duquel nous n'avons pas pu trouver de plus amples d'informations.

4.1.1. Utilisations par les médiévistes de l'ERA 713.

En 1971, au début de l'aventure de traitement automatique de textes au sein de l'équipe, seuls Genet et Gauvard envisagent l'utilisation de ALINE pour leurs recherches. Cette situation perdure visiblement au moins jusqu'en 1981, puisque le rapport d'activités pour le CNRS mentionne alors seulement trois « applications en cours ». Un tournant apparaît cependant pendant la période 1981-1983, puisqu'on constate dans le rapport de 1983 que pas moins de cinq autres chercheurs de l'équipe développent alors des projets d'applications de ce type de méthode. En 1985, huit chercheurs sur seize ont des projets de traitement automatique de textes. Les méthodes d'enregistrement, comme les objectifs scientifiques que les médiévistes se proposent de poursuivre, sont extrêmement variés. Il n'empêche que l'essor des applications du traitement de textes aux sources médiévales est incontestable au sein de l'équipe. Sur ces huit chercheurs, et à l'aide des sources disponibles, il n'est toutefois possible d'affirmer que pour deux d'entre eux qu'ils utilisaient effectivement le programme ALINE.

Le premier est le médiéviste Gérard Rippe (1944-1999), qui prépare à partir de 1981 un doctorat d'État au sein de l'ERA 713. Il travaille alors sur un *catasto* padouan du XIII^e siècle, et s'intéresse à la fois à la structure foncière, à la reconstitution des familles et à des problématiques onomastiques. Pour ce dernier aspect de sa recherche, il fait le choix d'enregistrer les noms du cadastre en plein texte afin de les analyser à l'aide

de ALINE. Dans un article de 1989, Rippe présente ce travail qu'il a mené depuis le début des années 1980⁴²⁵. Il y justifie le recours à l'utilisation de ce programme par la masse des données qu'il entendait exploiter, à savoir, des informations sur plus de 30 000 familles.

Comme le note Genet dans le rapport d'activité pour le CNRS, cette utilisation est cependant tout à fait originale. Premièrement, elle ne s'applique pas à des textes dits « littéraires », comme dans les travaux de Genet ou de Gauvard, mais bien à une source dite « documentaire », qui rappelle le *catasto* florentin étudié par Klapisch-Zuber et Herlihy. Deuxièmement, Rippe utilise ALINE essentiellement pour compter et classer les occurrences des prénoms et des patronymes, c'est-à-dire en poursuivant des problématiques onomastiques, tenant à l'histoire sociale, et non des questions sur le vocabulaire politique comme le faisaient Genet et Gauvard. Rippe propose donc un tout autre usage du programme que celui initialement prévu par ses concepteurs.

Le second cas qui nous intéresse ici est celui de l'utilisation de ALINE par le médiéviste Michel Sot (1942-), membre de l'ERA 713 et maître de conférences à l'université Paris X. Contrairement à Rippe, Sot exploite ALINE d'une façon tout à fait similaire à celle de Genet, puisque, comme nous l'indique le rapport d'activité de l'ERA 713 pour le CNRS de 1985, ce médiéviste cherche à dégager « *les notions qui reviennent le plus souvent* », « *le système des lieux* » et la « *liste exhaustive* »⁴²⁶ des acteurs mentionnés dans un texte du chroniqueur du X^e siècle Flodoard de Reims : l'*Historia Ecclesiae Remensis*. Sot projette aussi, toujours selon le même document, de développer « *des études comparatives de vocabulaire entre le texte de Flodoard et les nombreux documents qu'il y insère* »⁴²⁷. Comme Rippe, son intérêt pour ALINE repose donc sur les possibilités de classement et de calcul du programme. Elle s'articule toutefois aussi avec un réemploi de la méthode développée par Genet pour comparer les textes les uns avec les autres.

⁴²⁵ Gérard Rippe, « L'onomastique dans la Vénétie du XIII^e siècle : l'exemple de Monselice », *Publications de l'École Française de Rome* 120, n° 1 (1989): 269-85.

⁴²⁶ Jean-Philippe Genet, « Rapport d'activité pour le CNRS » (Université Paris 1 Panthéon-Sorbonne, 1985).

⁴²⁷ Genet.

4.1.2. Utilisation par une chercheuse « associée » à l'ERA 713.

À ces usages par des chercheurs affiliés à l'ERA 713 viennent s'ajouter d'autres entreprises, menées par des « membres associés » à cette équipe. Ce statut peut s'appréhender à travers l'exemple de Marie-Anne de Mercoyrol de Beaulieu. Cette chercheuse, que nous avons mentionnée plus haut, prépare une thèse de 3^e cycle à l'EHESS sous la direction de Le Goff, mais est rattachée à l'ERA 713 et participe donc aux activités de recherche de cette unité. Elle n'est cependant pas rémunérée comme un membre de l'équipe et ne possède, par exemple, pas de droit de vote aux assemblées générales. Par-delà ces considérations administratives, les raisons de cette association sont de première importance pour notre sujet. En effet, de Mercoyrol de Beaulieu est l'une des principales utilisatrices du programme ALINE, qu'elle met en œuvre dans le cadre de son étude sur le vocabulaire du recueil d'*exempla* du XIV^e siècle de Jean Gobi, la *Scala Coeli*. Les raisons de sa collaboration avec l'ERA 713 reposent alors autant sur sa volonté d'utiliser ce programme que sur la possibilité de profiter de certains équipements du centre de calcul de Paris 1.

Dès 1981, elle développe ainsi une méthode d'analyse des textes originale qui s'inspire des méthodes de la lexicométrie mais y ajoute une dimension. En effet, de Mercoyrol de Beaulieu cherche aussi à étudier des « catégories de mots » : noms propres, noms de lieu, personnages historiques, œuvres, auteurs, entités théologiques, mais également des classes grammaticales : adjectif, substantif et adverbe⁴²⁸. Par comparaison avec la méthode développée par Genet, elle doit donc mettre en place un pré-codage beaucoup plus élaboré pour annoter ces textes avant de les perforer. Elle utilise ainsi, pour le traitement de ces données, non seulement ALINE mais aussi deux autres programmes, JEUEMO et SAINCHEF, qui permettent le traitement des données sur les catégories de mot et les classes grammaticales.

Les apports de de Mercoyrol de Beaulieu à l'évolution des pratiques de l'analyse des textes dans ce contexte ne s'arrêtent cependant pas à la mise au point de cette articulation entre ces trois programmes. En effet, son travail avec les chercheurs de l'ERA 713 permet aussi de faire évoluer les conditions matérielles de ce type d'étude. De Mercoyrol de Beaulieu est par exemple la première à expérimenter, entre 1981 et 1983,

⁴²⁸ Genet, « Rapport d'activité pour le CNRS », 1983.

la saisie directe des textes sur le nouveau micro-ordinateur du centre de calcul de Paris 1, un MITRA 525, via un « *terminal écran* ». Selon Genet, cette solution « *apparaît positive tant sur le plan de la rapidité de la saisie que sur celui des coûts de fonctionnement* »⁴²⁹. La collaboration de de Mercoyrol de Beaulieu avec l'ERA 713 nous éclaire donc sur les modalités de la reprise des méthodes développées par Genet, aussi bien que sur les évolutions que peut entraîner le réemploi de ALINE à d'autres fins.

4.1.3. Utilisation dans le cadre d'activités de formation.

En plus de cette circulation dans le cadre d'activité de recherche, le programme ALINE bénéficie par ailleurs d'un autre type d'utilisation, qui touche un tout autre public. Le rapport d'activités de 1981 signale en effet que le programme est utilisé pour la formation des étudiants de la licence d'histoire de l'université Paris 1 dans le cadre de l'Unité de Validation « Informatique et histoire ». Dans le rapport d'activité de 1985, Genet précise même que le programme est utilisé pour mener une étude sur « *un jeu de données constitué par les discours télévisés des candidats à la présidence de 1981* »⁴³⁰.

Cet élément nous éclaire sur les relations entre recherche et enseignement, et témoigne de l'importance que peut revêtir l'existence de ce programme comme outil de formation. Il nous permet encore de mettre en lumière la longévité de ce projet d'intégration des méthodes informatiques en histoire médiévale. Le Laboratoire de Médiévistique Occidentale de Paris (LAMOP), créé en 1998, et successeur de l'ERA 713, est aujourd'hui encore l'un des principaux acteurs du développement et de la formation à ce type de méthodes en histoire médiévale en France. Le Pôle Informatique de Recherche et d'Enseignement en Histoire (PIREH) créé l'année suivante, en 1999, et dont les membres sont aujourd'hui majoritairement issus du LAMOP⁴³¹, poursuit le développement de cette relation entre recherche et enseignement à travers le développement d'une « *pédagogie centrée sur la création et l'exploitation des bases de données et de corpus textuels qui met l'étudiant en situation de chercheur* »⁴³².

Il faut signaler qu'il existe, au début des années 1980, d'autres types

⁴²⁹ Genet.

⁴³⁰ Genet, « Rapport d'activité pour le CNRS », 1985.

⁴³¹ Voir Dumont, Julien, et Lamassé, « Articuler histoire et informatique, enseignement et recherche ».

⁴³² Citation tirée du site de présentation du PIREH : <https://recherche-pireh.pantheonsorbonne.fr/bienvenue-site-pireh>. Dernière consultation le 17 mai 2021.

d'enseignement prévus à l'aide de ALINE, non plus à destination des étudiants mais d'un autre type de public. Le rapport d'activité pour le CNRS de 1983 nous indique notamment la mise en place, entre 1980 et 1981, d'un « *stage de sensibilisation à l'informatique* »⁴³³, organisé par André Zysberg (1947-) et Jean-Marie Mayeur (1933-2013). Si on ne connaît ni la teneur exacte ni le programme de ce stage, il est probable que ALINE y ait été présenté et utilisé dans le cadre d'une sensibilisation à l'intérêt des méthodes de traitement de texte automatique. Peut-être est-ce par ce canal que certains chercheurs se rapprocheront de l'ERA 713 dans les années suivantes ?

4.1.4. Utilisation en dehors de l'ERA 713.

Pour conclure, tournons-nous vers la dernière échelle de circulation mentionnée plus haut. Elle nous montre que le programme ALINE franchit les murs de l'ERA 713 et de l'UER d'histoire de l'université. Dans le rapport d'activité de 1983, Genet mentionne en effet l'utilisation de ALINE par un membre de l'UER de sciences politiques, Pierre Gaborit. Malheureusement, aucune publication ne nous permet de saisir la nature de cette utilisation.

Ces premiers éléments révèlent l'intérêt que les médiévistes ont éprouvé envers le programme ALINE, ainsi que l'utilisation en dehors de cette discipline de cet outil. Cependant, ces exemples illustrent surtout les limites de l'exportation du programme en dehors de l'université Paris 1. Les raisons qui pourraient rendre compte de cette situation sont à rechercher dans les conditions matérielles de la recherche assistée par ordinateur dans les années 1970 et 1980. En effet, pour utiliser ALINE il fallait non seulement en connaître l'existence, mais aussi saisir les modalités de son usage. Les utilisations de ALINE sont ainsi toutes fortement liées au lieu dans lequel le programme avait été créé, c'est-à-dire le centre de calcul de Paris 1, où l'ingénieur Marc Turquet assurait une assistance aux utilisateurs intéressés. En dépit de la volonté des membres de l'ERA 713 de faire de cet outil un instrument de travail pour toute la communauté des médiévistes, son attachement à un environnement de travail spécifique rendait la circulation de ce programme extrêmement difficile.

Un autre élément relatif aux environnements de travail informatique explique

⁴³³ Genet, « Rapport d'activité pour le CNRS », 1983.

encore cette faible circulation en dehors des murs de l'université Paris 1. Depuis la fin des années 1970, le développement de la microinformatique modifie en effet complètement le rapport aux centres de calcul. Pour utiliser un ordinateur, plus besoin de passer par les lourdes configurations de ces équipements. De nouvelles institutions ayant vocation à accompagner les chercheurs en sciences humaines et sociales dans la conduite de projets à teneur informatiques sont par ailleurs développées par le CNRS, comme le Laboratoire d'Informatique pour les Sciences de l'Homme (LISH), situé à la Maison des Sciences de l'Homme. Nous verrons dans les chapitres 3, 4 et 5 de cette thèse que ce sont également les réseaux d'information et de formation qui se développent à partir de cette période.

De plus, comme le note Genet dans son rapport de 1985 :

« de nouveaux programmes de traitement de texte vont arriver sur le marché, notamment l'ensemble des programmes mis au point par Pierre Lafon pour l'équipe que dirige Maurice Tournier au CNRS »⁴³⁴.

Mis à part ces programmes issus du travail des linguistes de l'ENS de Saint-Cloud, plusieurs autres programmes apparaissent pour la micro-informatique au début des années 1980, par exemple Lexico et Hyperbase. Les études sur le vocabulaire de Genet publiées après 1984 font justement état de ce changement d'instrument de travail, tout en attestant d'une continuité dans les méthodes de la lexicométrie et de l'analyse des données employées. Dans ce contexte, ALINE devient complètement obsolète, et le temps de travail que nécessiterait une adaptation de ce programme pour la micro-informatique conduit finalement Mondelli et Genet à abandonner la programmation de ALINE à partir de 1985.

4.2. Circulation et devenir de la cartothèque MEDITEXT.

Le deuxième instrument de travail qui nous intéresse ici est composé de l'ensemble des textes édités sur cartes perforées par les membres de l'ERA 713 et les chercheurs ayant utilisé ALINE. En effet, entre 1971 et 1985, plus de 40 textes, soit près

⁴³⁴ Genet.

de 160 000 cartes perforées ont été édités, au sens d' « édition » discuté plus haut. En 1983, Genet affirme même dans le rapport d'activité au CNRS que cette équipe est la seule à avoir commencé en France une « *bibliothèque de ce type* »⁴³⁵, alors qu'il en existe déjà une à Louvain, à Liège, à Montréal ou en Angleterre.

Cependant, comme le répète à de nombreuses reprises Genet dans la série de rapports d'activité rendus entre 1981 et 1985 :

« *la cartothèque MEDITEXT est malheureusement très difficile à utiliser pour des chercheurs extérieurs à l'équipe* »⁴³⁶.

4.2.1. Un processus de transfert de données extrêmement lourd (1972-1989).

Un cas illustre de manière particulièrement intéressante ces problèmes d'échanges de textes entre différents collectifs. Dans le rapport de 1983, Genet mentionne l'intérêt du Oxford University Computing Service pour un texte édité sur cartes perforées par l'ERA 713, *The Governour* de Thomas Elyot. À la lumière du temps de travail que peut nécessiter une édition au format cartes perforées (voir partie 2), on comprend vite l'intérêt d'une telle demande : un échange des données textuelles enregistrées sur un support de stockage informatique (cartes perforées, bandes magnétiques, disquette, etc.) permet en effet de court-circuiter la lourde phase d'édition des textes. Deux problèmes de nature différente se posent cependant aux acteurs dans ce contexte. Le premier est d'ordre éthique : il s'agit de s'assurer que les résultats obtenus sur la base de ces données ne seront pas publiés en premier lieu par l'équipe qui reçoit le texte. Mais ce problème est en fait tout à fait mineur.

Le problème majeur qui se pose dérive de contraintes techniques. En effet, comme nous le signale Genet dans ce même rapport de 1983, il est nécessaire, pour transmettre le texte de : 1) « *faire l'acquisition d'une bande magnétique* » ; 2) « *recopier dessus le contenu des cartes perforées, en prenant certaines précautions au stade du libellé de la bande de façon à ne pas gêner les utilisateurs d'Oxford* » ; 3) « *faire parvenir la bande à*

⁴³⁵ Genet.

⁴³⁶ Genet.

Oxford »⁴³⁷.

Ces trois opérations ne représentent qu'une partie du processus de transfert, qui est de ce fait particulièrement lourd. Il faut y ajouter les contraintes liées aux conditions de réalisation de ces opérations : 1) des problèmes liés à la nécessité de passer par le CIRCE pour faire le transfert des cartes aux bandes magnétiques ; 2) il faut encore gérer la conservation de ces bandes, qui doivent « *tourner au moins une fois par an* »⁴³⁸ pour être réutilisables. La nécessité de dupliquer les fichiers pose ainsi toute une série de problèmes matériels qui demandent du temps, et qu'en l'absence de moyens humains, les membres de l'ERA 713 ne peuvent assurer.

Ce manque de moyens de l'ERA 713 limite ainsi la possibilité de faire circuler ce nouveau type d'éditions de textes. Il n'est donc pas étonnant de retrouver, dans chacun des rapports d'activité adressés par l'ERA 713 au CNRS entre 1981 et 1985, des demandes concernant l'embauche d'un I.T.A. Ce nouvel agent servirait non seulement à accroître le rythme de perforation des textes, nous l'avons déjà évoqué, aussi bien qu'à favoriser une meilleure circulation de ces textes vers des chercheurs d'autres collectifs.

4.2.2. Devenir de MEDITEXT après 1989.

En l'absence de trace de circulation des textes dans ce contexte des années 1970-1980, on pourrait penser que l'échec de MEDITEXT est encore plus flagrant que celui de ALINE. Cependant l'intérêt de cet objet pour notre sujet réside dans le fait qu'au contraire de ALINE, cette bibliothèque de textes a été pérennisée et est encore utilisable aujourd'hui. En effet, elle subsiste sous une forme augmentée grâce à l'obtention, en 2009, d'un financement ERC qui a conduit au projet intitulé *Signs and States : les vecteurs de l'idéal*.

Dans un article de présentation de la plateforme PALM, datée de 2013, les auteurs présentent la mission de cette entreprise en ces termes :

« examiner les multiples manières par lesquelles le développement et l'application de nouvelles formes de gouvernance et de pouvoir étatique à la fin

⁴³⁷ Genet.

⁴³⁸ Genet.

du Moyen-Âge ont transformé les sociétés européennes, non seulement socialement et économiquement, mais également culturellement »⁴³⁹.

Le travail des chercheurs engagés dans ce projet, dont Genet était le principal artisan, consistait notamment dans l'élaboration d'une plateforme d'analyse linguistique médiévale, PALM, qui permettait de faciliter les trois tâches suivantes : 1) la composition d'un corpus ; 2) la lemmatisation des textes et 3) la mise en forme du corpus pour le traitement automatique des textes à l'aide de logiciel comme Lexico 3, développé par André Salem, HyperBase développé par Étienne Brunet ou TXM, développé par l'équipe de Serge Heiden à l'université de Lyon.

La plateforme PALM se base sur une bibliothèque de textes au titre familier, MEDITEXT, composée de près de 600 textes dont les 40 textes enregistrés sur cartes perforées entre 1971 et 1985 par les membres de l'ERA 713. La migration des supports anciens aux supports numériques contemporains a été menée selon le modèle suivant : les textes sur cartes perforées ont été scannés à partir des cartes elles-mêmes, scans qui ont ensuite été exploités à l'aide d'un logiciel de reconnaissance optique de caractère (OCR). Cette méthode reposait sur le fait que les textes édités sur cartes perforées ne posaient pas les problèmes inhérents aux limites de reconnaissance automatique des caractères des textes anciens. En effet, les logiciels d'OCR fonctionnent très mal pour les langues médiévales, en raison de l'absence de dictionnaires pour contrôler l'existence des formes, mais aussi en raison de considérations paléographiques.

Ce sont encore les moyens humains et financiers des projets ERC qui ont permis la pérennisation de ces textes. Une linguiste spécialiste du moyen-français, Naomi Kanaoka, a par exemple travaillé pendant un an sur le corpus de texte français. De nombreux ajouts de textes ont encore pu être réalisés, comme par exemple le *De Regimine Principum* de Gilles de Rome qui avait été, en raison du manque de moyens, écarté du corpus au moment du projet ALINE. Pour les textes anglais, deux autres linguistes sont encore embauchés à cette période, Rachel Moss puis Chloé Morgane, ce qui a permis l'ajout au corpus de nombreux textes, et notamment de la traduction du miroir au prince de Gilles de Rome.

⁴³⁹ Le document en question est disponible ici : <https://archive-2013-2016.lamop.fr/IMG/pdf/PALM-Meditext.pdf>. Voir page 1 pour la citation. Dernière consultation le 18 mai 2021.

Concernant les opérations de lemmatisation, le projet *Signs and States* a aussi pu bénéficier du développement des méthodes menées depuis 1985 en d'autres lieux. Pour les textes en moyen-français, les chercheurs ont notamment pu se baser sur un lemmatiseur créé à l'université de Nancy entre 2009 et 2010, LGerM, qui permet l'automatisation d'une partie des opérations de lemmatisation⁴⁴⁰. Pour les textes en latin, il était possible de s'appuyer sur un dictionnaire électronique finalisé en 2009 à l'École des Chartes, et constitué sur la base du lexique du glossaire Du Cange⁴⁴¹. Cependant, en 2013, il n'existait toujours pas de solution complètement automatique pour la lemmatisation des langues médiévales. Le projet reposait ainsi sur un modèle semi-automatique amélioré par rapport au modèle élaboré dans le cadre de ALINE, qui a permis la mise en place d'une interface d'annotation qui facilitait les tâches de lemmatisation.

Un autre élément de cette plateforme semble encore intéressant pour illustrer l'évolution des méthodes employées par les historiens pour annoter les corpus. Dans le cadre de PALM se développe une interface d'étiquetage morpho-syntaxique, c'est-à-dire une interface permettant de caractériser automatiquement chaque terme d'une phrase par sa fonction grammaticale. Le projet *Signs and States* a ainsi été de pair avec la mise au point de nouveaux types d'instruments de travail pour les langues du Moyen-Âge qui, contrairement au programme ALINE, permettait d'étudier automatiquement les aspects syntaxiques du texte.

Ces différents éléments attestent d'une pérennisation des données textuelles produites par l'ERA 713 dans les années 1970 et 1980 et d'une reprise de certains principes du programme ALINE plus de trente ans après sa première version. De la même manière que pour l'enquête du *catasto*, on remarque ainsi que dans le temps long, les acteurs attribuent plus de valeur aux éditions de leurs données qu'à leurs programmes. Cet état de fait peut être expliqué par la nature différente des compétences nécessaires pour produire chacun de ces types d'objets. Pour les programmes, la présence d'informaticiens est indispensable. Au contraire, pour les données, ce sont les spécialistes

⁴⁴⁰ Gilles Souvay et Jean-Marie Pierrel, « LGeRM Lemmatisation des mots en Moyen Français », *Revue TAL* 50, n° 2 (2009): 21.

⁴⁴¹ Voir Frédéric Glorieux, « Pourquoi informatiser un vieux glossaire ? », *Ela. Etudes de linguistique appliquée* n° 156, n° 4 (2009): 417-28.

des sources et de leurs langues (historiens et linguistes) qui sont en mesure de produire en toute indépendance des outils pertinents.

Conclusion.

Au terme de cette longue analyse, quelles évolutions liées aux usages des ordinateurs est-il possible de mettre en lumière ? Et comment ces conclusions se confrontent-elles aux précédentes, formulées dans le chapitre portant sur le projet d'enquête du *catasto* ? Répondre à ces deux questions suppose non seulement 1) de reprendre chacun des quatre grands axes de l'analyse développée dans ces deux derniers chapitres (organisation collective, méthode de collecte et d'édition des données, raisonnements mis en œuvre à l'aide des ordinateurs et conséquences historiographiques), mais nécessite encore 2) d'insister sur les limites de cette comparaison, qui viennent non seulement de la nature des projets mais aussi de la structure de la documentation disponible.

Le premier élément sur lequel il semble important d'insister concerne les objectifs, sensiblement similaires, que poursuivent les deux collectifs que nous avons étudiés jusqu'ici. Il s'agissait en effet, pour tous les deux, 1) de produire de nouveaux types d'instruments de travail pour la communauté des médiévistes (une édition pour les uns et un programme pour les autres), tout en 2) poursuivant leurs propres recherches à l'aide de ces instruments. L'identification de ces deux dimensions, que nous avons respectivement qualifiée d'heuristique et de documentaire, constitue un premier résultat issu de la comparaison entre ces deux cas. Nous verrons dans le chapitre suivant que cette façon de se représenter le rôle des ordinateurs s'inscrit, dès le début des années 1970, comme l'une des conséquences d'une ambition qui prend forme, plus largement, au sein de la discipline « histoire médiévale », en lien avec l'apparition des moyens électroniques de calcul.

Toutefois, derrière ce premier point commun, il faut noter que les collectifs qui se constituent dans ces deux contextes sont de dimension et de composition très différentes. D'un côté, Genet travaille pour la conception d'ALINE au sein d'une petite équipe, composée d'une dizaine de chercheurs, dont seuls quelques-uns participent activement à

la confection du programme. De l'autre, nous l'avons vu, les deux médiévistes qui dirigent l'enquête sur le *catasto* sont entourés de nombreux collaborateurs et de « petites mains », qui forment une équipe de plus d'une trentaine de personnes, à cheval sur deux continents et trois pays (France, États-Unis, Italie). Cette situation permet, du fait de la dispersion géographique de ses membres, et de la correspondance écrite qui en résulte, de saisir précisément les modalités du travail collectif qui se met en place.

Il faut ajouter à ce premier ensemble de considérations le fait que les moyens, matériels, financiers et humains, dont disposent ces deux équipes sont tout à fait disparates. L'équipe du *catasto*, intégrée au Centre de Recherches Historiques, bénéficie de la culture interdisciplinaire de l'École Pratique des Hautes Études, et donc, de ses moyens humains et matériels (statisticien à temps plein, laboratoires de mécanographie, de démographie, de cartographie). Par contraste, l'ERA 713 rencontre tout au long de la période étudiée des difficultés liées à la disponibilité limitée des informaticiens et au manque de personnel pour contribuer à l'édition des textes sur cartes perforées. On pourrait encore y ajouter l'absence de véritable spécialiste de la linguistique, discipline pourtant au cœur du projet, qui tranche avec l'aide précieuse apportée par les démographes à l'enquête du *catasto*.

Ce sont encore les relations entre les chercheurs au sein de ces deux collectifs qui s'organisent selon des modèles très différents. Les membres de l'ERA 713 développent des relations horizontales, d'égal à égal, puisque l'équipe est constituée essentiellement d'étudiants poursuivant des projets de recherche individuels, dans le cadre de leurs doctorats. Dans ce contexte institutionnel, le développement du programme ALINE apparaît comme un projet qui vise à mettre en commun des méthodes, empruntées à la lexicométrie. Par contraste, l'équipe de l'enquête du *catasto* se constitue autour d'un objectif de recherche commun, qui consiste en l'étude et l'édition pour l'ordinateur du *catasto* florentin de 1427. Les relations entre les membres de ce second collectif sont plus hiérarchisées, et le travail s'organise de façon presque pyramidale.

Ces trois éléments, liés à l'organisation de ces collectifs et à la division du travail qui s'y met en place, ont de fortes implications sur la nature des instruments de travail qu'ils produisent. À ce titre, le projet de développement du programme ALINE constitue un exemple particulièrement intéressant, puisque les difficultés matérielles et financières que ses initiateurs rencontrent, combinées au fait que ces chercheurs ne travaillent pas

tous sur les mêmes sujets, les conduit à développer un programme informatique original, puisqu'utilisable pour différents types de corpus, et de façon presque autonome par les médiévistes. Cet exemple montre le lien qui existe entre l'organisation collective des équipes et la nature des instruments de travail qu'elles produisent. Il nous permet de plus de préciser les modalités de cette relation, somme toute évidente *a priori*, en insistant sur le fait que deux éléments y apparaissent déterminants : 1) l'indépendance technique des médiévistes et 2) l'unicité ou la multiplicité des objectifs de recherche poursuivis à l'aide de ces instruments de travail.

Il faudrait, pour établir des conclusions plus générales, voire même, pour établir une typologie de ces relations, étudier aussi précisément les organisations collectives et les instruments produits par d'autres collectifs. En attendant la suite de ces recherches, notons seulement ici qu'en France, dans les années 1970 et 1980, rares sont les équipes organisées sous le modèle de l'enquête du *catasto* (pour rappel, une Recherche Coopérative sur Programme) dans le champ des études médiévales assistées par ordinateur, puisque la forme la plus commune est en effet celle de l'Équipe de Recherche Associée⁴⁴².

De nombreux points communs entre ces deux études de cas peuvent encore être mis en lumière quant aux méthodes de collecte et d'édition des données. Tout d'abord, on constate un élargissement des corpus que traitent les historiens, selon des modalités différentes en fonction des projets. Le bond quantitatif est cependant évident, du moment que l'on se concentre sur les unités d'analyse mobilisées (les familles pour l'enquête du *catasto*, les textes et les mots pour le travail de Genet). Ensuite, on remarque dans ces deux cas l'influence décisive d'autres sciences sociales sur l'élaboration de ces méthodes. D'une part, la démographie, qui joue un rôle central dans l'élaboration du code pour Klapisch-Zuber et Herlihy et, d'autre part, la linguistique et la sociologie, dont Genet recycle les méthodes pour développer non seulement un nouveau procédé de constitution de corpus en histoire des idées politiques mais également une nouvelle approche pour classer les textes politiques du Moyen-Âge.

Il résulte de ces influences extra-disciplinaires une conséquence sur ce que nous avons qualifié de *modélisation* des textes, et qui peut être défini comme la manière par

⁴⁴² Fossier et al., « Vingt ans d'informatique en histoire médiévale ».

laquelle les chercheurs conçoivent ce qui, dans un texte, mérite d'être analysé, en fonction de leurs problématiques, des méthodes qu'ils développent et de leurs différents champs de recherche. Mais cette évolution de la modélisation des textes n'est pas uniquement due au recyclage des méthodes des sociologues, des démographes et des linguistes.

Elle tient tout d'abord à la contrainte théorique qu'implique la mise en œuvre de méthodes quantitatives, qui suppose de constituer des corpus composés de données de même nature, dans le but de les comparer. Sur le modèle des méthodes de l'histoire quantitative, que nous avons présentées dans le chapitre précédent et sur lesquelles nous ne reviendrons donc pas ici, la constitution des corpus devient, dans le même mouvement, une opération de constitution d'un espace de comparaison. Sur ce point, nous avons notamment vu que la méthode de construction des corpus de Genet apparaissait, pour l'histoire des idées politiques, comme une véritable évolution théorique.

L'évolution de la *modélisation* des textes repose encore sur la nature des équipements informatiques de cette période, qui impliquaient nécessairement l'utilisation d'une technologie intellectuelle très contraignante : la carte perforée IBM à 80 colonnes. L'une des conséquences de l'utilisation des ordinateurs qui se dégage, sur la base de ces deux études de cas, est donc la nécessité que rencontraient les acteurs de formaliser explicitement la représentation de ce qu'était, pour eux, un texte. Si cette problématique préexiste, certainement, à l'apparition des ordinateurs, le recours à ce nouvel instrument nous permet, en tant qu'observateur, d'en voir apparaître, dans les archives des historiens, les logiques et les manifestations de façon précise.

Cette formalisation a par ailleurs des effets sur la nature des instruments de travail produits par ces collectifs. En effet, le choix des données à extraire de la source diffère en fonction des problématiques historiques que poursuivent les acteurs, ce qui, nous y reviendrons plus largement dans le chapitre 3, limite la portée des instruments de travail qu'ils entendent constituer. Cette tension entre la volonté de mettre en commun les initiatives d'informatisation (sous la forme d'édition pour le *catasto* ou de programme pour ALINE) et la multiplicité des types de modélisation possible se posera ainsi comme l'une des discussions centrales pour la communauté des médiévistes utilisateurs des ordinateurs à la fin des années 1970⁴⁴³.

⁴⁴³ Nous renvoyons ici au troisième chapitre de cette thèse, qui porte en grande partie sur un colloque organisé à l'École Française de Rome en 1975, moment clef de ces échanges entre les différents collectifs engagés dans l'informatisation de la discipline.

Une autre des questions centrales qui traverse cette première partie de la thèse est de comprendre quels sont les effets de l'usage des ordinateurs sur la pratique des statistiques en histoire médiévale. L'analyse de l'étude du *catasto* nous avait permis de conclure sur un constat globalement attendu, avec, notamment, un recours plus intensif aux opérations statistiques basiques de l'histoire quantitative, comme les tris à plats, les tris croisés et les classements, qui portaient également, dans ce cadre, sur une masse de données plus grande. On retrouve les mêmes éléments dans l'utilisation des ordinateurs que propose les membres de l'ERA 713 : élargissement des dimensions des corpus, gain de temps sur les opérations simples de manipulation des données (tris et classement), mais également sur les opérations de calcul des fréquences.

Deux autres apports des ordinateurs, soulignés par ces deux groupes d'acteurs, sont directement lié au précédent. Il s'agit tout d'abord de ce que Robin qualifiait de systématisation des opérations de traitement des données, et que nous avons analysées dans le chapitre précédent à travers les propos du démographe Louis Henry (1911-1991). Le recours aux ordinateurs était en effet présenté par ce dernier comme une contrainte théorique essentielle qui conduisaient à élaborer des raisonnements reproductibles systématiquement. Ce phénomène constitue, selon nous, l'une des évolutions majeures liée au recours aux ordinateurs, qui se manifeste dans le contexte des sciences sociales et des sciences humaines dès la fin des années 1960. Cette évolution renforce, sur un autre aspect que celui décrit plus haut, la formalisation des raisonnements dans ces disciplines.

Cette conséquence est par ailleurs corrélée à un autre apport des ordinateurs mentionné par Robin, auquel elle attribue la vertu d'« *objectiver* » les méthodes de traitement des informations. Si cette expression n'est pas clairement définie par Robin, nous l'avons, au cours de notre raisonnement, rapprochée à la fois du phénomène soulevé par Klapisch-Zuber, à savoir que le recours aux ordinateurs évitait les erreurs de calculs, et de l'idée plus large que le recours aux statistiques et aux méthodes informatiques limitait le recours à la subjectivité de l'historien.

L'un des éléments de cette conséquence peut encore, selon nous, être repéré dans ce qui constitue le changement le plus frappant dans les méthodes d'analyses mises en œuvre par ces médiévistes : les nouvelles formes d'images qui se développent dans ces deux contextes. Dans le premier chapitre, nous avons notamment montré l'apparition

d'une nouvelle fonction des cartes, qui, du fait de leur production selon des procédés automatiques et des manipulations que cela permettait, devenaient des outils heuristiques pour la recherche, et non plus de simples outils d'illustration pour le lecteur des études.

Dans ce deuxième chapitre, les graphiques d'analyses factorielles des correspondances s'inscrivent dans un registre similaire, et ceci de manière encore plus flagrante que dans le cas précédent. Ces nouvelles représentations des corpus de données constituent indubitablement l'une des conséquences de l'utilisation des ordinateurs sur l'épistémologie des acteurs, sur lequel nous reviendrons dans les chapitres 4 et 6 ainsi que dans notre conclusion générale.

Les conséquences de l'utilisation des ordinateurs que nous mettons ici en lumière nous permettent de conclure que le recours aux moyens électroniques de calcul s'inscrit, dans ce contexte et à cette période, dans une volonté plus large de faire évoluer les méthodes historiques en s'appuyant sur de nouvelles formes d'images, à même de représenter les relations entre plus de deux variables à la fois (cartes de Bertin, graphiques d'analyse factorielles), et de développer ainsi de nouvelles formes d'observation et de critique des sources.

À ce titre, les conséquences historiographiques de l'étude du *catasto* sont certainement les plus facilement mesurables, à partir, notamment, des recensions et des très nombreuses citations du livre. La précision et la pertinence des analyses en ont fait, comme nous l'avons vu, un étalon pour les historiens comme pour les démographes. Mais les grandes innovations méthodologiques et épistémologiques en histoire médiévale sont plutôt à rechercher du côté du travail de Genet qui, comme d'autres médiévistes de sa génération, sont véritablement au cœur de l'évolution des méthodes quantitatives, qui aboutira au milieu des années 1980 avec la création de la revue *Histoire & Mesure*. Dans l'éditorial du premier numéro, Genet, qui en est l'un des fondateurs, présentait en ces termes le changement profond qui s'était produit en histoire sous l'effet des applications de l'informatique :

« L'école historique française, de fait, s'est vite taillée une réputation dans le domaine de l'histoire quantitative. Mais aujourd'hui la mode n'est plus guère à ce genre de préoccupations, et d'autres débats agitent apparemment la communauté

scientifique. Pourquoi donc créer, aujourd'hui en 1985, une revue intitulée Histoire & Mesure ?

Le besoin, l'urgence sont là. Ils sont révélés en histoire comme dans les autres sciences humaines et sociales, par l'ordinateur, qui a profondément transformé la pratique de nombreux historiens. Tout d'abord, il a obligé à travailler sur un ensemble de données, extraites de la source, qui sont construites et pourvues nécessairement d'une structure finie. Il a contraint l'historien à chercher les outils optimaux d'analyse et d'interprétation et à effectuer des choix décisifs dans l'arsenal sans cesse renouvelé que lui offre l'analyse des données. Dès lors il n'est plus d'histoire quantitative ou d'histoire qualitative. Tout se mesure : les flux, les évolutions et les ruptures, les équilibres et les déséquilibres peuvent être dégagés, circonscrits, dessinés. »⁴⁴⁴

Les raisons de ce changement de paradigme sont à rechercher, nous semble-t-il, dans deux mouvements communs et contradictoires du début des années 1980. Le premier, c'est l'essoufflement des programmes historiographiques lancés par les travaux Camille-Ernest Labrousse (1895-1988), Georges Duby (1919-1996) ou encore Pierre Goubert (1915-2012) en France dans les années 1950 et 1960, et qui se donnaient pour objet d'étudier les couches populaires des sociétés du passé par l'exploitation des grandes sources statistiques. À ce titre, l'étude du *catasto* est certainement un exemple emblématique de la réussite de l'informatisation de ces ambitions historiographiques, mais demeure cependant une exception en histoire médiévale. En effet, et comme nous l'avons vu, les structures de recherche adaptées à ce type d'entreprise demeuraient finalement très rares chez les médiévistes. Ce sont donc d'autres formes d'organisations collectives, plus légères, et plus indépendantes, que sont venues des innovations méthodologiques et épistémologiques décisives.

De la même manière que les méthodes développées par les historiens de la *microstoria*, qui permettait d'atteindre les mentalités des couches populaires des populations sans avoir recours aux immenses sources statistiques⁴⁴⁵, les médiévistes de ces équipes comme l'ERA 713 devaient donc inventer des méthodes dans des contextes

⁴⁴⁴ « Éditorial », *Histoire & Mesure* 1, n° 1 (1986): 5.

⁴⁴⁵ Sur ce point voir notamment la préface de la première édition de Carlo Ginzburg, *Il Formaggio e i vermi: il cosmo di un mugnaio del'500* (Torino, Italie: G. Einaudi, 1976)..

plus contraignants, sans pouvoir se lancer dans l'étude de sources de cette envergure. C'est d'ailleurs ces mêmes limites matérielles qui rendaient nécessaires pour des historiens comme Genet ou Guerreau⁴⁴⁶ une autre approche de l'informatique, qui passait par une formation solide en statistique et en programmation et qui contribuait de ce fait à leur indépendance technologique. La commercialisation massive des micro-ordinateurs au début des années 1980 finira par donner raison, de l'extérieur de la discipline, à ces historiens qui seront au cœur du renouvellement des méthodes dites quantitatives en histoire à partir des années 1990.

⁴⁴⁶ Pour une analyse des pratiques de recherche et des méthodes de Guerreau, nous renvoyons ici au chapitre 6 de cette thèse.

PARTIE II

Chapitre 3 : Une nouvelle ambition collective pour l'histoire médiévale : les calculateurs électroniques (1970-1979).

La première partie de cette thèse avait pour objet de comparer les effets de l'utilisation des ordinateurs via l'étude de deux collectifs de chercheurs en histoire médiévale (l'équipe de l'enquête du *catasto* et l'Équipe de Recherche Associée 713 (ERA 713)). Dans ces deux premiers chapitres, nous ne nous sommes concentrés que sur les cultures de travail locales qui se sont développées dans ces contextes. Cependant, comme nous l'avons vu, les *instruments de travail* élaborés par ces collectifs (édition de source, éditions de textes, programme) avaient vocation à circuler plus largement au sein de la communauté des médiévistes. Cette circulation posait toute une série de difficultés matérielles, mais aussi intellectuelles, que nous avons aperçues dans les dernières parties de ces deux chapitres.

La deuxième partie de cette thèse vise justement à poursuivre cette analyse en examinant comment certains des nombreux collectifs impliqués dans l'utilisation des ordinateurs en histoire médiévale se sont aménagés des espaces de discussions et d'échanges à l'échelle de la discipline. L'objet de ce troisième chapitre est d'analyser les relations entre les premières initiatives qui attestent de cette dynamique méta-collective en Europe.

Depuis 1965, plusieurs formes d'initiatives de ce type ont été mises en œuvre à l'échelle des sciences humaines. Dans un article de 1972, Christian Wenin (1928-1987), professeur de philosophie médiévale à la Faculté de Philosophie et de Lettres de l'université catholique de Louvain, propose par exemple un inventaire - non-exhaustif - des rencontres et des publications en lien avec ce sujet dans le domaine de la philosophie⁴⁴⁷. Il dénombre onze rencontres (congrès, colloques, tables rondes), presque toutes centrées sur des champs de recherche ou des disciplines particulières (études des classiques, philologie, lexicologie, linguistique, textes philosophiques). Elles ont eu lieu

⁴⁴⁷ Christian Wenin, « L'informatique au service de la philosophie. Réalisation et projets », *Revue Philosophique de Louvain* 70, n° 6 (1972): 177-211.

en France (4), en Allemagne (1), en Belgique (1), en Italie (1), en Tchécoslovaquie (1), en Écosse (1), aux États-Unis (1) et au Canada (1).

Dans cet article, Wenin présente aussi sept publications qui ont vocation à recenser et à faire circuler les travaux de ce type. Elles sont, pour cinq d'entre elles, basées aux États-Unis et traitent de toutes les disciplines des sciences humaines et sociales. Deux autres sont plus spécialisées : 1) la plus ancienne, la revue du Laboratoire d'Analyse Statistiques des Langues Anciennes (LASLA), lancée en 1965 en Belgique, porte plus précisément sur les approches linguistiques et 2) le bulletin d'information *Computer and Medieval Data Processing (CAMDAP)*, publié à partir de 1971 par l'Institut d'Études Médiévales de l'université de Montréal, qui se consacre comme son nom l'indique aux études médiévales.

Cet inventaire dressé par Wenin en 1972 nous permet de relever l'existence de plusieurs modes opératoires qui ont vocation à organiser les activités des nombreux collectifs qui utilisent des ordinateurs. D'une part, certains chercheurs s'impliquent dans le recensement des collectifs et des projets. L'article de Wenin participe d'ailleurs de cette dynamique puisqu'il propose, en plus de l'inventaire des rencontres et des revues, une liste de quarante-deux collectifs engagés dans ce type de projets en philosophie et dans les champs de recherches affiliés (philologie, études littéraires). À une autre échelle, la revue *Computer and the Humanities* fait figure de modèle, puisqu'elle publie deux fois par an, depuis 1966, un *Directory of Scholars Active*, qui recense les collectifs et les acteurs impliqués dans l'utilisation des ordinateurs à l'échelle mondiale, et pour l'ensemble des sciences humaines.

D'autre part, l'organisation entre ces collectifs passe aussi par des initiatives dont la visée n'est pas le recensement mais la mise en relation de chercheurs confrontés à des problèmes communs. Les onze rencontres présentées par Wenin dans son article nous donnent quelques exemples des vecteurs de réunion qui se mettent en place entre 1965 et 1972. Les chercheurs se regroupent par exemple autour 1) de types d'objectifs, comme l'étude du vocabulaire ou des citations, 2) de type de corpus, comme les textes philosophiques, ou 3) plus largement autour d'un champ de recherche comme l'étude des classiques. Il faut cependant insister sur le fait que ces deux modes opératoires se combinent et se recoupent, et structurent donc le champ de recherche à plusieurs niveaux.

A l'échelle de l'histoire médiévale, trois principales initiatives naissent au début des années 1970. La première prend forme en Amérique du Nord à travers la création du bulletin d'information *CAMDAP* évoqué plus haut. En 1970, le comité permanent de la « Medieval Academy of America », la plus grande organisation dans le champ des études médiévales aux États-Unis⁴⁴⁸, décide de la création d'une sous-commission chargée « *de coordonner et de diffuser les informations sur le travail mené à l'aide d'ordinateurs concernant des textes médiévaux* »⁴⁴⁹. Ce sont alors quelques chercheurs de l'Institut d'Études Médiévales de l'université de Montréal qui prennent en charge l'édition et la distribution d'un bulletin gratuit, rédigé en anglais, qui a pour objet de faire circuler « *des informations sur les travaux en cours, sur le matériel et les méthodes utilisées* »⁴⁵⁰.

La deuxième se dessine en Europe, et trouve sa source dans le champ de la philosophie médiévale. En 1972, au V^e Congrès de la Société internationale pour l'étude de la philosophie médiévale tenue à Madrid, une commission intitulée « *Utilisation de l'électronique au service de la publication et de l'étude de textes médiévaux* » avait élaboré un projet d'organisme nommé Ordinateurs et Recherches Médiévales (ORM), dont la fonction était de recenser, d'informer et de coordonner les recherches de ce type à l'échelle de l'Europe⁴⁵¹. Cette commission était dirigée par Paul Tombeur (1936-), directeur du Centre de Traitement Électronique des Documents (CETEDOC) créé en 1968 à l'université catholique de Louvain pour le « *traitement automatique de documents médiévaux* »⁴⁵². Ce dernier avait notamment lancé en 1970 une enquête qui avait pour but de recenser les collectifs utilisant des ordinateurs dans ce domaine, et sur laquelle nous reviendrons dans la troisième partie de ce chapitre.

Un an après le colloque de Madrid, un proche collaborateur de Tombeur, le médiéviste Léopold Genicot (1914-1995), publiait un premier manifeste en faveur d'une initiative similaire dans le domaine de l'histoire médiévale. En 1973, Genicot est au faite de sa carrière académique. Il est professeur d'histoire médiévale à l'université catholique de Louvain depuis 1947, mais aussi membre de l'Académie royale des Sciences, des

⁴⁴⁸ Voir par exemple David Herlihy, « The American Medievalist: A Social and Professional Profile », *Speculum* 58, n° 4 (octobre 1983): 881-90.

⁴⁴⁹ Wenin, « L'informatique au service de la philosophie. Réalisation et projets », 183.04/09/2021 14:18:00

⁴⁵⁰ Wenin, 183.

⁴⁵¹ Christian Wenin, « Le Ve Congrès international de philosophie médiévale », *Revue Philosophique de Louvain* 71, n° 9 (1973): 129.

⁴⁵² Paul Tombeur, « Le Centre de Traitement Électronique des Documents, au service des disciplines ecclésiastiques », *Revue Théologique de Louvain* 1, n° 4 (1970): 433.

Lettres et des Beaux-Arts de Belgique depuis 1960 et co-fondateur du Centre belge d'Histoire rurale, créé en 1963⁴⁵³. Il est par ailleurs engagé, comme nous l'avons vu dans le chapitre précédent, dans une entreprise d'édition sur bandes magnétiques des sources littéraires composées en latin sur le territoire de l'actuelle Belgique avant 1200⁴⁵⁴.

Dans le premier numéro de la revue internationale *Francia*, consacrée à l'histoire du Nord de l'Europe, il lance un appel intitulé : « Pour une organisation de la recherche en histoire médiévale »⁴⁵⁵. Pour réaliser ce programme, il suggère en conclusion de son article la création d'un « *secrétariat* » qui aurait pour fonction de :

« *centraliser et répandre toutes les informations en matière d'emploi des machines électroniques en histoire médiévale* »⁴⁵⁶.

Dans les deux dernières pages de ce texte, on découvre que vingt-six autres médiévistes, occupant tous des positions importantes dans des institutions impliquées dans des projets utilisant des ordinateurs, soutiennent l'initiative de Genicot, et se déclarent prêts à « *recevoir des suggestions et des adhésions d'autres médiévistes* »⁴⁵⁷. Si la France (11), l'Allemagne (5) et la Belgique (3) sont surreprésentées, la liste comprend des signatures de chercheurs exerçant en Angleterre, en Écosse, en Italie, en Espagne, en Tchécoslovaquie ou même en Californie⁴⁵⁸.

⁴⁵³ Voir par exemple André Joris, « Léopold Genicot (1914-1993) », *Cahiers de Civilisation Médiévale* 39, n° 153 (1996): 173-75.

⁴⁵⁴ Pour une brève présentation de l'un des résultats de ce travail voir Léopold Genicot, « Index Scriptorum Operumque Latino-Belgicorum Medii Aevi. Nouveau répertoire des œuvres médiolatines belges publié sous la direction de L. Genicot et P. Tombeur. Première partie : VIIe-Xe siècles par A. Stainier (= Académie royale de Belgique. Comité national du dictionnaire du latin médiéval), Bruxelles, 1973 », *Bulletins de l'Académie Royale de Belgique* 60, n° 1 (1974): 29-30.

⁴⁵⁵ Léopold Genicot, « Pour une organisation de la recherche en histoire médiévale », *Francia* 1 (1973): 692-98.

⁴⁵⁶ Genicot, 697.

⁴⁵⁷ Genicot, 697.

⁴⁵⁸ Nous reviendrons plus longuement sur cette liste au fil de notre analyse. Indiquons cependant ici quelques noms pour illustrer notre propos (avec les positions qui y sont accolées) : Georges Duby (1919-1996), professeur au Collège de France, Jean Glénisson, directeur de l'Institut de Recherche et d'Histoire des Textes (IRHT), Jean Schneider (1903-2004), professeur à l'université de Nancy II, Edmond-René Labande (1908-1992), directeur du Centre d'Études Supérieures de Civilisation Médiévale à Poitiers, Yves Lefèvre (1918-1987), président de l'université de Bordeaux ou encore Philippe Wolff (1913-2001), professeur à l'université de Toulouse, Emilio Saez (1919-1988), directeur de l'Institut d'Histoire Médiévale de l'université de Barcelone, Helmut Beumann (1912-1995), de l'Institut für mittelalterliche Geschichte, geschichtliche Hilfswissenschaften und geschichtliche Landeskunde de l'université Philipps de Marbourg, Carlrichard Brühl (1925-1997), du séminaire historique de l'université de Giessen, Eugen Ewig (1913-2006), de l'université de Bonn, Karl Ferdinand Werner (1924-2008), directeur de l'Institut Historique

La troisième initiative qui nous intéresse prend une toute autre forme. Elle ne naît ni au sein d'une académie ni d'une société savante comme les deux précédentes, mais s'inscrit dans la continuité des activités d'une historienne centrale pour notre sujet. Au début des années 1960, Lucie Fossier (1924-), archiviste-paléographe de formation, est attachée de recherche à l'Institut de Recherche et d'Histoire des Textes (IRHT). À partir de 1966, elle est détachée de l'IRHT au Centre de Recherche et d'Applications Linguistique (CRAL), en tant que responsable de la section des textes historiques. Elle y mène notamment une entreprise d'indexation documentaire - méthode sur laquelle nous reviendrons longuement - d'actes diplomatiques du XV^e siècle, dont les premiers éléments sont publiés en 1970⁴⁵⁹. En 1971, tout en poursuivant ses activités au CRAL, elle fonde avec le directeur de l'IRHT Jean Glénisson (1921-2010), lui aussi signataire du manifeste de Genicot, la section informatique de l'IRHT⁴⁶⁰ qui collaborera de près avec le CRAL jusque dans les années 1990.

À cette même période, certainement entre 1972 et 1974, Fossier apporte son aide à un projet de traitement informatique des registres des papes du XIV^e siècle, dirigé par l'archiviste-paléographe Michel Hayez (1932-) avec le concours de l'École Française de Rome et de l'IRHT⁴⁶¹. Dans ce cadre, Fossier entreprend d'organiser une rencontre entre cette équipe et celle dirigée par Cinzio Violante (1921-2001), professeur d'histoire médiévale à l'université de Pise, lui aussi signataire du manifeste de Genicot, qui s'est engagé dans un projet d'édition des chartes pisanes antérieures à 1200. L'objet initial de cette initiative est pour ces équipes de partager les difficultés qu'elles rencontrent dans

Allemand, Paul Tombeur, professeur au CETEDOC, Gérard Verbeke (1910-2001), professeur à l'université de Louvain, Cinzio Violante (1921-2001), professeur de paléographie et de diplomatique à l'université de Pise, Domenica Legge (1905-1986), du département de français de l'université d'Edinburgh, Jindrick Sebanek (1900-1977), professeur de philosophie à l'université d'Usti Nad Labem, Stephan Kuttner (1907-1996), de l'École de droit de l'université de Californie ou encore Peter Hayes Sawyer (1928-2018), de l'École d'Histoire de l'université de Leeds.

⁴⁵⁹ Marion Créhange et Lucie Fossier, « Essai d'exploitation sur ordinateur des sources diplomatiques médiévales », *Annales* 25, n° 1 (1970): 249-84.

⁴⁶⁰ Cette date est proposée par Louis Holtz dans un article de blog tiré d'une communication de 2007 : Louis Holtz, « L'institut de recherche et d'histoire des textes (IRHT). Premier laboratoire d'histoire au Centre National de la Recherche Scientifique », *Les Cahiers du Centre de Recherches Historiques. Archives*, n° 36 (30 juin 2005). Sur le contexte de la mise en place de services informatiques à l'IRHT nous recommandons encore un entretien de Jean Glénisson mené par Edmond-René Labande en 1971 : Jean Glénisson et Edmond-René Labande, « Formation et destin de l'Institut de Recherche et d'Histoire des Textes [Réponses de Jean Glénisson, directeur de l'I.R.H.T., aux questions posées par E.R. Labande au cours d'un entretien, le 11 juin 1971.] », *Cahiers de Civilisation Médiévale* 15, n° 57 (1972): 53-60.

⁴⁶¹ Voir Bruno Galland, « La publication des registres de lettres pontificales par l'École française de Rome. », *Bibliothèque de l'École des chartes* 154, n° 2 (1996): 625-34, <https://doi.org/10.3406/bec.1996.450837>.

leurs utilisations des ordinateurs pour l'édition de sources médiévales⁴⁶². En 1975, ce sera finalement un colloque de trois jours réunissant plus de soixante participants, dont Genicot et plusieurs signataires de son manifeste de 1973, qui se tiendra à l'École Française de Rome.

L'articulation entre ces trois initiatives constitue au début des années 1970 la matrice de l'organisation méta-collective qui se met en place entre les chercheurs utilisant des ordinateurs en histoire médiévale. Cependant, nous allons le voir, ces trois initiatives prennent forme dans des contextes différents, proposent des modes opératoires différents et poursuivent des objectifs différents. Plusieurs questions émergent de cette configuration.

Tout d'abord, quelles sont les raisons qui poussent ces acteurs à proposer des formes d'organisation à une autre échelle que celle de leur collectif de travail ? Et quelles sont les logiques d'organisation qui structurent ces initiatives ? Sont-elles les mêmes à l'échelle de leur champ de recherche et à l'échelle de la discipline ? Ensuite, comment ces initiatives s'articulent-elles les unes avec les autres ? Enfin, quelles sont les solutions proposées ou imaginées par les acteurs pour surmonter ces problèmes à ces différentes échelles ? Et quelles sont les réalisations concrètes sur lesquelles débouchent ces propositions ?

Pour répondre à ces questions, nous proposons dans ce chapitre une approche chronologique et comparative de l'évolution de deux de ces trois initiatives : celle qui prend forme à Louvain autour de Genicot, et celle qui prend forme autour de Fossier à l'IRHT. L'analyse du *CAMDAP* est ici laissée de côté en raison du fait que ses éditeurs se concentrent, pour des raisons que nous allons éclairer, essentiellement sur les collectifs qui travaillent sur le continent nord-américain à partir de 1975.

Pour valoriser l'intérêt de notre étude, il nous faut cependant revenir sur deux éléments qui fondent la valeur de ce chapitre pour notre propre champ de recherche. Le premier touche directement à l'issue de l'une de ces deux initiatives. En effet, le colloque de Rome conduit à la création d'un premier secrétariat de coordination, composé de six chercheurs français, qui lancent en 1977 une enquête sur les pratiques de leurs confrères

⁴⁶² Cet épisode nous est connu grâce à l'avant-propos des actes du colloque qui naît de cette initiative. Voir Lucie Fossier, André Vauchez, et Cinzio Violante, « Avant-propos », *Publications de l'École Française de Rome* 31, n° 1 (1977): 5-9. Nous reviendrons longuement sur les intentions initiales des organisateurs et sur le colloque lui-même au fil de ce chapitre.

utilisant des ordinateurs dans le cadre de leur travail. Les résultats de cette enquête, sur lesquels nous reviendrons dans la troisième partie de ce chapitre, les conduisent à mettre sur pied en 1979 une publication fondamentale pour comprendre les logiques de circulation des pratiques des ordinateurs en histoire médiévale en Europe : le bulletin de liaison *Le Médiéviste et l'ordinateur*⁴⁶³. L'étude de la genèse de cette publication nous paraît essentielle pour aborder l'étude de la fonction et des objectifs de cette publication, que nous analyserons dans le quatrième chapitre de cette thèse⁴⁶⁴.

Le deuxième élément qui fonde l'intérêt de ce chapitre concerne plus largement l'histoire des premiers moments des humanités numériques. Dans un article récent de 2019, Chris Allen Sula et Heather V. Hill déplorent les limites de l'histoire canonique des *digital humanities* telle qu'elle a été écrite dans le monde anglo-saxon depuis 2004⁴⁶⁵. Dans ce texte, les auteurs formulent une série de quatre critiques dont deux nous intéressons ici. Tout d'abord, ils affirment que ce récit canonique privilégie certains types d'usages des ordinateurs, que les observateurs ont désignés sous le terme de *text analysis*⁴⁶⁶. Ils insistent alors sur le fait que la conséquence principale de cette approche est la difficulté de saisir les relations entre ce type d'usages des ordinateurs et d'autres, qui devraient pourtant théoriquement être inclus dans l'histoire des *digital humanities*, comme l'analyse de réseau, l'histoire quantitative, etc.

Sur un autre plan, Sula et Hill critiquent le fait que cette histoire ne soit pas assez documentée, mais plutôt présumée et postulée (« *assumed and applied* »)⁴⁶⁷. Sans revenir sur les raisons qui expliqueraient cet état de fait, il apparaît évident que cette seconde critique est intimement liée à la première, puisque manque justement des études des documents qui permettraient de comprendre comment des acteurs aux pratiques différentes partagent des problèmes communs du fait de leur instrument de travail, l'ordinateur. Dans la conclusion de leur article, les deux chercheurs appellent donc aux

⁴⁶³ Cette collection est notamment disponible en ligne, à l'adresse suivante : <https://www.persee.fr/collection/medio>

⁴⁶⁴ Il faut par ailleurs informer ici le lecteur que les chapitres 5 et 6 de cette thèse s'appuient eux aussi essentiellement sur des documents extraits de ce périodique.

⁴⁶⁵ Chris Allen Sula et Heather V Hill, « The early history of digital humanities: An analysis of Computers and the Humanities (1966–2004) and Literary and Linguistic Computing (1986–2004) », *Digital Scholarship in the Humanities* 34, n° Supplement_1 (1 décembre 2019): 190-206.

⁴⁶⁶ Les deux auteurs citent, entre autres références, Susan Hockey, « The History of Humanities Computing », in *A Companion to Digital Humanities* (John Wiley & Sons, Ltd, 2004), 1-19 et Sandy Baldwin, « The Idiocy of the Digital Literary (and what does it have to do with digital humanities)? », *Digital Humanities Quarterly* 007, n° 1 (1 juillet 2013).

⁴⁶⁷ Sula et Hill, « The early history of digital humanities », 192.

développements de « *focused stories* »⁴⁶⁸, qui se concentreraient sur des manifestations disciplinaires ou régionales de l'articulation entre ces différents types d'usages.

Ce chapitre s'inscrit dans la continuité de cet appel, puisqu'il propose une analyse historique des échanges et des modes d'organisation qui se mettent en place entre des groupes d'acteurs qui développent des usages des ordinateurs différents. Cette étude, avec d'autres qui, nous espérons, suivront, doit permettre d'analyser en contexte les problématiques communes qui unissent ces acteurs, et les raisons qui les poussent à s'organiser méta-collectivement. Ce type de travail historique contribuera ainsi à montrer que les humanités numériques ne sont pas seulement un label inventé par les contemporains mais repose bien sur des problèmes partagés par des collectifs scientifiques dont il faut saisir les limites en contexte.

Pour aborder les questions qui nous intéressent ici, nous étudierons dans un premier temps les initiatives menées par Genicot et Fossier, qui prennent forme à partir de 1970. Nous analyserons dans un second temps les problématiques communes qui émergent au cours du colloque de Rome de 1975. Nous comparerons enfin les propositions pratiques que formulent Genicot et Fossier à l'issue du colloque. Cette comparaison nous permettra de comprendre les différentes visions qui s'opposent, tant pour ce qui est des ambitions que des objectifs assignés à ce cadre méta-collectif.

1. Les calculateurs électroniques, nouveau vecteur de collaboration à l'échelle de la discipline (1970-1975).

Les deux cas étudiés dans les deux premiers chapitres de cette thèse nous ont amenés à montrer les modalités de la coopération entre des médiévistes ayant recours aux ordinateurs pour leur recherche en histoire médiévale à partir de 1966. Ce travail nous a permis de mettre en lumière l'existence de collectifs de chercheurs, développant des cultures de travail qui pouvaient réunir des individus d'institutions différentes, voire de pays différents. Ces collectifs de chercheurs entretenaient par ailleurs d'autres types de relations avec des acteurs extérieurs, qui ne partageaient pas leur manière de travailler. Nous avons vu que ces relations pouvaient prendre plusieurs formes : conseils des

⁴⁶⁸ Sula et Hill, 202.

démographiques aux historiens de l'enquête du *catasto*, partenariat entre l'Atelier Vincent de Beauvais du CRAL et l'ERA 713 dans le cadre du projet d'enquête de Bernard Guenée (1927-2010), ou encore échange de textes entre l'équipe de Jean-Philippe Genet (1944-) et l'Oxford Computing Center.

Au-delà de ces formes observées par nous, de nombreux autres types d'interactions entre les nombreux collectifs engagés dans l'utilisation des ordinateurs ont pris forme dès la fin des années 1960. Deux d'entre elles, présentées brièvement en introduction, nous intéresseront ici puisqu'elles ont entraîné la mise en œuvre d'initiatives d'organisation à des échelles plus larges que la simple collaboration entre un ou deux de ces collectifs. Il s'agit 1) de l'initiative lancée par un groupe de l'université catholique de Louvain, dont Genicot est la figure de proue pour l'histoire médiévale et 2) du projet de colloque mis sur pied notamment par Fossier entre 1972 et 1974, qui avait pour objet de faire se rencontrer deux équipes travaillant sur des sources similaires.

Ce qu'il est important de mettre en avant ici est le fait que ces initiatives se sont formées dans des contextes spécifiques, dans lesquels les usages des ordinateurs et les fonctions qui étaient attribuées à ces instruments concordaient avec des objectifs de recherche et un programme historiographique. De fait, comme nous allons le voir, les ambitions attachées à la mise en œuvre d'une collaboration à l'échelle de la discipline sont donc différentes pour ces deux groupes d'acteurs.

1.1. Réorganiser toute la discipline : la proposition de Genicot.

Dans le texte de Genicot de 1973 mentionné en introduction, « Pour une organisation de la recherche en histoire médiévale », le médiéviste déplore que sa discipline « *souffre de gros défauts* »⁴⁶⁹, et soit dans « *beaucoup de secteur* »⁴⁷⁰ : 1) « *fragile* », car « *elle fonde ses enquêtes et ses conclusions sur des textes médiocres et sur des dépouillements insuffisants* »⁴⁷¹ ; 2) « *partielle* », car « *elle n'est trop souvent que*

⁴⁶⁹ Genicot, « Pour une organisation de la recherche en histoire médiévale », 692.

⁴⁷⁰ Genicot, 692.

⁴⁷¹ Genicot, 692.

qualitative »⁴⁷² ; et même 3) « *fausse à un certain degré* », puisqu' « *un aspect du passé est absent parce qu'on n'a pas recensé tous les manuscrits de toutes les œuvres* »⁴⁷³.

Pour remédier à cette situation, Genicot préconise la constitution de trois types « *d'instrument de travail* »⁴⁷⁴ : 1) des inventaires des sources disponibles, 2) des éditions de meilleure qualité ainsi que 3) des typologies des documents d'archives susceptibles de guider les médiévistes dans l'analyse critique et l'interprétation des documents. Pour préciser les relations qui existent entre ces instruments et les défauts que Genicot diagnostique, il nous faut revenir sur les travaux dans lesquels il est lui-même engagé en 1973.

1.1.1. Fonctions des ordinateurs dans l'élaboration de nouveaux instruments de travail.

En 1972, Genicot dresse le plan et publie l'introduction d'une collection aujourd'hui bien connue des médiévistes, intitulée *Typologie des sources du Moyen Âge occidental*⁴⁷⁵, qui a pour vocation d' « *établir la nature de chaque genre de sources et dégager ainsi les règles spéciales de critiques valables pour lui* »⁴⁷⁶. Cet instrument méthodologique doit donner aux médiévistes les clefs qui leur permettront de considérer dans leurs recherches les types de sources qui ne leur sont pas familières. Ils pourront ainsi, selon les mots de Genicot, « *entendre sur chaque point tous les témoignages* » et élaborer une « *histoire totale et rigoureuse* »⁴⁷⁷.

Plus important pour notre sujet, Genicot a par ailleurs entamé en 1962, en collaboration avec Tombeur et le CETEDOC, un inventaire et une édition sur cartes perforées des textes écrits en latin avant 1200 sur le territoire belge. L'objet de ce travail était de doter les historiens médiévistes :

⁴⁷² Genicot, 693.

⁴⁷³ Genicot, 694.

⁴⁷⁴ Genicot, 692.

⁴⁷⁵ Léopold Genicot, *Typologie des sources du Moyen Âge occidental*, éd. par Institut d'études médiévales (Turnhout, Belgique: Brepols, 1972).

⁴⁷⁶ Léopold Genicot, « La typologie des sources du Moyen Âge occidental », *Annales* 27, n° 6 (1972): 1258.

⁴⁷⁷ Genicot, 1257.

« d'un outil utile, voire indispensable, puisque beaucoup de leurs problèmes sont d'ordre sémantique et ne peuvent être résolus que par le repérage de tous les emplois de tel ou tel terme »⁴⁷⁸.

Dans son article de 1973, Genicot insiste sur la fonction des ordinateurs dans ce processus en ces termes :

« Va-t-on donc imposer à des chercheurs la tâche immense et fastidieuse de dépouiller tous les textes et d'en classer alphabétiquement tous les mots ? Ce serait gaspiller leur temps et leur compétence. Car des machines peuvent exécuter cette besogne. Les ordinateurs électroniques existent et les historiens doivent s'en servir. »⁴⁷⁹

Toujours dans son article de 1973, il précise l'intérêt de ce type de dépouillement pour les médiévistes :

« Ils [les médiévistes] lui [l'ordinateur] demanderont aussi d'analyser et caractériser la langue de chaque document, de chaque auteur, de chaque période ou contrée et ils pourront mieux juger ainsi de l'authenticité ou de la provenance des textes, mieux décider également de la valeur des copies en cas de perte de l'original [...] et publier par là des éditions plus sûres. »⁴⁸⁰

Ce type d'instrument lexicographique devait donc remédier à la « fragilité » de l'histoire médiévale diagnostiquée par Genicot, en offrant aux médiévistes des dépouillements plus satisfaisants et en permettant l'établissement d'éditions papier de meilleure qualité.

Plusieurs fonctions sont donc attribuées aux ordinateurs par Genicot dans le programme qu'il dessine dans son article de 1973. Premièrement, le recours aux ordinateurs doit permettre de trier et classer automatiquement et rapidement de grandes quantités de données textuelles dans le but de repérer les différents emplois des termes,

⁴⁷⁸ Genicot, « Index Scriptorum Operumque Latino-Belgicorum Medii Aevi. Nouveau répertoire des œuvres médiolatines belges publié sous la direction de L. Genicot et P. Tombeur. Première partie », 29.

⁴⁷⁹ Genicot, « Pour une organisation de la recherche en histoire médiévale », 694.

⁴⁸⁰ Genicot, 695.

libérant ainsi les historiens de ces tâches répétitives et chronophages. Deuxièmement, ce type d'opérations doit faciliter la critique des sources afin, notamment, d'améliorer leur édition. Genicot articule donc dans cet article trois éléments fondamentaux pour comprendre l'intérêt que certains médiévistes portent aux ordinateurs : 1) un diagnostic sur l'état de la discipline ; 2) les instruments de travail d'un nouveau type qu'il est nécessaire d'élaborer pour dépasser cet état de fait et 3) la fonction allouée aux ordinateurs dans la réalisation et l'exploitation de ces instruments de travail.

1.1.2. Une organisation collective nécessaire à l'échelle de la discipline.

Le médiéviste belge est cependant conscient de l'ampleur de ce programme, et notamment du coût des projets utilisant les ordinateurs (en temps et en subsides). Il identifie notamment deux écueils qui fondent selon lui la nécessité de mettre sur pied un « *secrétariat* »⁴⁸¹ coordonnant ce type de réalisation à l'échelle de la discipline. Le premier concerne ce qu'il appelle les « *doubles emplois* »⁴⁸² : en l'absence de circulation des informations, deux collectifs distincts pourraient en effet enregistrer en machine les mêmes textes, ou confectionner des programmes ayant les mêmes fonctionnalités. Ce type de situation nuirait au bon emploi des financements et du temps des chercheurs, puisqu'il conduirait à reproduire un travail déjà effectué par d'autres. Le second se rapporte à ce qu'il désigne sous le terme d' « *entreprises limitées* »⁴⁸³. Selon lui, il faut éviter au maximum que les historiens n'enregistrent qu'une seule partie des textes, ce qui limiterait les possibilités d'exploitation de ces nouveaux types d'instruments de travail par d'autres collectifs.

Il faut ici insister sur l'adhésion suscitée par cet appel au sein de la communauté des historiens médiévistes européens en 1973, en rappelant que ce projet de secrétariat est appuyé par vingt-six médiévistes de différents collectifs utilisant des ordinateurs en Europe et aux États-Unis. Parmi eux, on retrouve une grande majorité d'historiens à la tête d'institutions impliquées dans des projets à caractère informatique. C'est par exemple le cas de Philippe Wolff (1913-2001), directeur de la Recherche Coopérative sur

⁴⁸¹ Genicot, 696.

⁴⁸² Genicot, 695.

⁴⁸³ Genicot, 695.

Programme 181 (RCP) qui chapeaute l'enquête sur le *catasto* étudiée dans le premier chapitre ; Karl-Ferdinand Werner (1924-2008), qui envisage, au titre de directeur de l'Institut Historique Allemand, un projet de traitement par ordinateurs de fiches prosopographiques de la *Prosopographia regnorum orbis latini*⁴⁸⁴ ou Karl Hauck (1916-2007), fondateur en 1968 d'une équipe pluridisciplinaire, la *Sonderforschungsbereich 7* de l'Institut für Frühmittelalterforschung de l'université de Münster, au sein de laquelle sont menées de nombreuses enquêtes informatisées sur l'anthroponymie médiévale⁴⁸⁵.

La liste pourrait ainsi être allongée. Cependant, cette forte adhésion ne saurait masquer la spécificité du programme de Genicot. La comparaison avec une autre initiative de ce type, qui poursuit d'autres objectifs, travaille avec d'autres types de sources et propose un autre mode opératoire pour mettre en relation les collectifs utilisant des ordinateurs en histoire médiévale, nous permet d'éclairer la diversité des ambitions attachées aux moyens électroniques de calculs dans ce contexte.

1.2. Mettre en commun des difficultés : l'objet initial du colloque de Rome.

L'autre initiative qui nous intéressera ici prend forme à la marge d'une coopération entre deux institutions françaises : le Centre de Recherche et d'Applications Linguistiques (CRAL) et l'Institut de Recherche et d'Histoire des Textes (IRHT). Le CRAL a été créé par le médiéviste Jean Schneider (1903-2004) en 1966. Ce dernier occupe alors la position de doyen de la faculté de Lettres de Nancy depuis 1959⁴⁸⁶. Ce centre avait pour objet de permettre la collaboration de chercheurs avec l'Institut Universitaire de Calcul Automatique de Nancy, et hébergea dès sa création plusieurs équipes de médiévistes⁴⁸⁷. Certains d'entre eux, parmi lesquels Lucie Fossier,

⁴⁸⁴ Une présentation rapide en est notamment proposée dans Karl-Ferdinand Werner, « Problèmes de l'exploitation des documents textuels concernant les noms et les personnes du monde latin (IIIe-XIIe siècles) », *Publications de l'École Française de Rome* 31, n° 1 (1977): 205-12. Certains résultats sont discutés dans Karl-Ferdinand Werner, « Liens de parenté et noms de personne. Un problème historique et méthodologique », *Publications de l'École Française de Rome* 30, n° 1 (1977): 13-18.

⁴⁸⁵ Pour une présentation de l'unité au terme de ses activités voir « Mittelalterforschung in Münster in Der Nachfolge Des Sonderforschungsbereichs 7 », *Frühmittelalterliche Studien* 24, n° 1 (31 décembre 1990): 380-429.

⁴⁸⁶ Jean-Pierre Callu, « Ordre du jour et Allocution à la mémoire de M. Jean Schneider, membre de l'Académie », *Comptes rendus des séances de l'Académie des Inscriptions et Belles-Lettres* 148, n° 2 (2004): 797-98.

⁴⁸⁷ Voir Hélène Naïs, « Présentation des travaux du Centre de recherche et d'applications linguistiques (CRAL) », *Revue du LASLA*, n° 1 (1970): 71-79.

travaillaient notamment à l'indexation documentaire de sources du Moyen-Âge, solution intermédiaire entre l'enregistrement plein-texte et à l'édition d'une banque de données, sur laquelle nous reviendrons au 2.3.⁴⁸⁸.

Fossier était par ailleurs attachée à une autre institution centrale pour notre sujet : l'IRHT. Créée en 1937, l'IRHT est le plus vieux laboratoire en sciences humaines du CNRS. Sa fonction était, selon les mots de son fondateur Félix Grat (1898-1940), de créer une bibliothèque idéale rassemblant tous les manuscrits médiévaux et imprimés. Pour mener à bien cette mission, les membres de l'IRHT, pour la plupart des archivistes-paléographes, se proposent depuis 1937 de relever et d'identifier tous les textes conservés dans les dépôts d'archives publics en France, et de créer des instruments de travail pour les historiens : notices de sources, dépouillements de fonds d'archives, micro-films de manuscrits et éditions papier de sources anciennes⁴⁸⁹. Deux grandes collections sont ainsi publiées par l'Institut : « Documents, études et répertoires » depuis 1948, et « Sources pour l'histoire médiévale » depuis 1965.

À la fin des années 1960, c'est en raison de la volonté d'utiliser des ordinateurs pour l'indexation de corpus de manuscrits médiévaux (chartes, corpus de droit canon) que se développe la coopération entre le CRAL et l'IRHT. L'université de Nancy est à ce moment-là l'un des pôles de développement de l'informatique en France⁴⁹⁰, et l'IRHT cherche à intégrer cette nouvelle science auxiliaire à ses méthodes de travail. Fossier est au cœur de ce rapprochement, en tant que membre des deux institutions. En 1970, elle publie un article dans les *Annales* avec une informaticienne formée à Nancy, Marion Créhange (1937-), qui vise à présenter un essai d'exploitation d'un index de sources

⁴⁸⁸ Ce type de démarche, inspiré du travail mené par Jean-Claude Gardin (1925-2013) en archéologie, peut être découvert en contexte à l'aide du document suivant : Lucie Fossier et Gian Piero Zarri, *L'indexation automatique des sources documentaires anciennes*, Bibliographies, colloques, travaux préparatoires (Paris: Éditions du Centre national de la recherche scientifique, 1975). Pour une approche plus complète des travaux de Gardin voir la thèse récemment soutenue par Sébastien Plutniak, « L'opération archéologique. Sociologie historique d'une discipline aux prises avec l'automatique et les mathématiques. France, Espagne, Italie, 2e moitié du XXe siècle » (These de doctorat, Paris, EHESS, 2017). Il semble important d'informer ici le lecteur qu'un chapitre de cette thèse, consacré à ce type d'utilisation des ordinateurs à l'Institut de Recherche et d'Histoire des Textes (IRHT), a dû être abandonné en raison de l'impossibilité d'accéder aux archives du fait de la crise de la Covid-19.

⁴⁸⁹ Voir par exemple Holtz, « L'institut de recherche et d'histoire des textes (IRHT). Premier laboratoire d'histoire au Centre National de la Recherche Scientifique ».

⁴⁹⁰ Voir en particulier Marion Créhange et Marie-Christine Haton, « L'informatique universitaire à Nancy : un demi-siècle de développement », *Technique et science informatiques* 33, n° 1-2 (2014): 127-41; Michel Grossetti et Pierre-Eric Mounier-Kuhn, « Les débuts de l'informatique dans les universités », *Revue française de sociologie* XXXVI-n°2 (1995): pp.295-324. Une thèse sur ce sujet, menée par Hala Khassiba est par ailleurs en cours à l'université de Lorraine pour la période 1950-1980.

diplomatiques enregistrées sur cartes perforées⁴⁹¹. En 1971, la création de la section informatique de l'IRHT, dont Fossier prend en charge la direction, permet le développement de plusieurs projets de ce type, mais aussi la fabrication de programmes informatiques destinés à être partagés entre les collectifs des deux institutions.

1.2.1. Un projet de rencontre entre deux collectifs.

C'est dans ce contexte que Fossier se lance avec Violante, que nous avons déjà présenté, et André Vauchez (1938-), directeur des études médiévales à l'École Française de Rome, dans l'organisation d'un colloque ayant vocation à aménager une rencontre entre deux collectifs de recherche. Dans l'éditorial des actes du colloque, les trois éditeurs présentent ainsi l'objet initial du colloque :

« À propos de deux entreprises actuellement en cours - la publication des chartes pisanes antérieures à 1200 pour les historiens de l'Université de Pise et celle des registres des papes du XIV^e siècle pour l'École française de Rome - des médiévistes français et italiens ont été amenés à se poser le problème de la validité respective des méthodes traditionnelles et de l'informatique dans le domaine de l'exploitation des grandes séries de documents conservées dans les archives. Ayant fait, chacun de leur côté, l'expérience de la difficulté qu'il y a, à notre époque, à mener à bien, dans ce domaine, des entreprises d'édition qui soient à la fois conduites selon des critères scientifiques rigoureux et susceptibles d'être achevées dans des délais raisonnables, les signataires de ces lignes avaient, dès 1974, mis en commun leurs perplexités et leur désir de rechercher des solutions nouvelles. »⁴⁹²

Cette citation nous permet de repérer plusieurs différences entre cette initiative et celle de Genicot. La première est que cette dynamique d'organisation prend forme à une échelle locale. Elle ne vise pas, dans un premier temps, à créer un cadre général qui coordonnerait les activités des différents collectifs utilisant des ordinateurs à l'échelle de

⁴⁹¹ Créhange et Fossier, « Essai d'exploitation sur ordinateur des sources diplomatiques médiévales ».

⁴⁹² Fossier, Vauchez, et Violante, « Avant-propos », 5.

la discipline, mais à faire coopérer quelques-uns de ces collectifs, qui affrontent des difficultés communes.

Une autre différence majeure entre le programme de Genicot et cette initiative se manifeste dans cette citation. Elle concerne le type de textes auxquels doivent donner accès les instruments de travail que projettent d'élaborer ces acteurs. Dans ce second cas, c'est d'édition de sources qu'il s'agit, et non plus d'édition d'œuvres « littéraires ». Toujours dans l'avant-propos des actes du colloque, les trois éditeurs formulent cette distinction en ces termes :

« L'objet du présent colloque, c'était principalement le document historique, les sources archivistiques, dispersées, diverses dans leur forme comme dans leur contenu. Ce genre de texte, on s'en doute, ne peut être traité en machine à la façon d'une œuvre littéraire, d'abord parce qu'il se présente sous une autre forme, ensuite et surtout parce que le point de vue de l'historien qui utilisera les résultats du traitement diffère fondamentalement de celui d'un lexicologue ou d'un philologue. Si dans le cas d'un corpus littéraire, l'utilisateur demande à la machine de l'aider à dégager les caractéristiques de style, de vocabulaire, de grammaire ou de syntaxe présentées par une œuvre ou un auteur, la préoccupation essentielle de l'historien est de saisir le contenu, nous pourrions dire l'aspect proprement documentaire d'un texte de toutes façons dépourvu de valeur littéraire et le plus souvent anonyme. »⁴⁹³

Les instruments de travail que veulent élaborer les collectifs de Pise et de l'École Française de Rome ont donc pour finalité de donner accès à des ensembles documentaires tout à fait différents de ceux que visait Genicot dans son article de 1973. Mais c'est encore la finalité scientifique de ces instruments qui est envisagée différemment par ces acteurs : l'historien, au contraire du philologue ou du lexicologue, s'intéresserait au « contenu » plutôt qu'à la « langue » des textes.

⁴⁹³ Fossier, Vauchez, et Violante, 6.

1.2.2. D'autres fonctions pour d'autres types de sources.

Dans l'optique de donner accès au « contenu » des documents, les organisateurs du colloque de Rome mettent eux aussi en avant la nécessité de constituer de nouveaux instruments de travail. Ce sont ici les limites des conditions d'élaboration des éditions de sources « traditionnelles », qui justifient la constitution de tels instruments. En effet, la publication papier traditionnelle « implique[nt] de longs délais »⁴⁹⁴ de réalisation, souvent plusieurs dizaines d'années, ce qui empêche pratiquement l'accès à la majeure partie des collections de sources. Pour prendre seulement un exemple marquant, l'entreprise de publication de la collection des registres pontificaux du XIII^e siècle a été entamée en 1884 par l'École de Rome et a pris fin en 1960⁴⁹⁵. Dans ce contexte, les trois organisateurs du colloque de Rome insistent, dans l'avant-propos, sur les demandes des historiens :

*« Même s'il tire profit, un jour, d'une édition parfaite, il [l'historien] veut dans l'immédiat pouvoir user des sources déjà dépouillées et demande qu'elles soient à sa disposition avant même leur mise en forme définitive. »*⁴⁹⁶

Comme pour le programme de Genicot, on remarque que ces instruments de travail ne visent pas à remplacer les éditions traditionnelles, mais à fournir aux historiens un nouveau type de ressources pour mener certains types de recherches.

Pour poursuivre cet objectif, le recours aux ordinateurs apparaît comme une solution nouvelle. Toujours dans l'avant-propos des actes, les trois organisateurs présentent la fonction de ce nouvel instrument dans cette dynamique en ces termes :

*« À ce stade, l'informatique peut, sans aucun doute, intervenir de façon heureuse, en autorisant une utilisation rapide de données innombrables - et aussi d'ailleurs en contribuant à accélérer les procédures habituelles d'édition, en particulier par l'élaboration automatique d'index de noms de personnes et de lieux ainsi que d'index-matière. »*⁴⁹⁷

⁴⁹⁴ Fossier, Vauchez, et Violante, 6.

⁴⁹⁵ Galland, « La publication des registres de lettres pontificales par l'École française de Rome. »

⁴⁹⁶ Fossier, Vauchez, et Violante, « Avant-propos », 6.

⁴⁹⁷ Fossier, Vauchez, et Violante, 7.

Le recours aux ordinateurs doit donc non seulement permettre la réalisation d'instruments de travail qui autorisent « *une utilisation rapide de données* », comme les bases documentaires sur lesquels travaillent Fossier au CRAL et à l'IRHT, mais également contribuer à accélérer l'élaboration des index des éditions traditionnelles.

Ces éléments nous révèlent les rapports qu'établit ce second groupe d'acteurs entre : 1) un diagnostic sur la discipline ; 2) les instruments de travail d'un nouveau type qu'il est nécessaire d'élaborer et 3) la fonction allouée aux ordinateurs dans la réalisation et l'exploitation de ces instruments de travail. Comme pour le cas de Genicot, on constate que les moyens électroniques sont présentés par ce deuxième groupe d'acteurs comme des outils qui visent à faciliter, à accélérer des tâches d'érudition plutôt que de recherche. On remarque cependant que ces ambitions attachées aux utilisations des ordinateurs ne concernent ni le même type de sources, ni le même type d'usages de ces sources. Les projets concernés par ces deux programmes ne sont donc, *a priori*, pas les mêmes. L'initiative de Genicot concerne plutôt des entreprises d'étude de la « *langue* » de corpus d'œuvres littéraires, au contraire de l'initiative initiale des organisateurs du colloque de Rome qui s'intéresse aux projets portant sur le « *contenu* » de grands ensembles de documents d'archives.

Cependant, nous l'avons dit, l'initiative de Fossier, Violante et Vauchez présentée ici ne constitue que la matrice initiale du colloque de Rome. Pour saisir plus exhaustivement la teneur des problématiques attachées à ce type d'usages des ordinateurs, les organisateurs du colloque décident finalement d'« *élargir le champ d'investigation* »⁴⁹⁸, en réunissant non pas deux mais une dizaine de collectifs utilisant des moyens électroniques dans le cadre de leur travail.

1.3. Une première rencontre méta-collective : le colloque de Rome de 1975.

Avant de plonger dans l'analyse des difficultés et des problématiques partagées par les acteurs, il est nécessaire de présenter succinctement les collectifs présents à Rome, dans le but de montrer que cette rencontre de 1975 déborde du projet initial présenté par

⁴⁹⁸ Fossier, Vauchez, et Violante, 7.

Fossier, Violante et Vauchez dans l'avant-propos. En effet, si de nombreuses communications concernent la réalisation d'instruments de travail permettant de donner accès de façon anticipée au « contenu » de grandes séries de documents, on retrouve, parmi les quarante-six articles publiés dans les actes, plusieurs textes qui concernent des projets d'une autre nature.

1.3.1. Instruments de travail génériques.

Un premier ensemble d'articles se rapporte directement aux usages des ordinateurs définis par les éditeurs dans l'avant-propos. Les actes du colloque s'organisent ainsi dans leur première partie autour des différents genres d'ensembles documentaires qui intéressent les médiévistes. Autour des registres pontificaux par exemple, neuf articles sont publiés. Ils s'articulent tous autour de la problématique suivante : comment donner accès, plus rapidement qu'en éditant traditionnellement sur papier, à certains aspects des quelque 970 000 documents des archives du Vatican⁴⁹⁹ ? Le même type de défi est partagé par les éditeurs et les utilisateurs des collections de chartes, comme par les chercheurs qui travaillent au contact des documents notariés.

En plus des représentants des deux collectifs à l'initiative de cette rencontre, on retrouve plusieurs autres groupes d'acteurs intéressés par ce type d'usage des ordinateurs. Certains, comme Fossier ou les équipes du CRAL, ont déjà élaboré des outils informatiques permettant la production de différents types de répertoires, d'inventaires et surtout d'index onomastiques ou d'index-matières. D'autres, comme Robert-Henri Bautier (1922-2010), professeur de diplomatique à l'École Nationale des Chartes depuis 1961 et éditeur de la collection des « Sources de l'histoire économique et sociale du Moyen Age »⁵⁰⁰, expérimentent la lenteur des processus de l'édition traditionnelle et se pose la question de l'intérêt de l'utilisation des moyens électroniques. Le colloque de Rome est ainsi le moment d'un échange entre des utilisateurs pionniers de ces nouvelles

⁴⁹⁹ Giulio Battelli, « Una proposta per un indice dei registri pontifici », *Publications de l'École Française de Rome* 31, n° 1 (1977): 23.

⁵⁰⁰ Voir par exemple Olivier Guyotjeannin, « Les sources de histoire économique et sociale du Moyen Age par Robert-Henri Bautier et Janine Soknay, deuxième série. Les Etats de la maison de Bourgogne, vol I, Archives centrales de Etat bourguignon (1384-1500), Archives des principautés territoriales Les principautés du sud Les principautés du nord supplément) comtés de Hollande et Zélande et duché de Gueidre par Michel van Gent mise à jour du fasc. additions et corrections », *Bibliothèque de l'École des chartes* 160, n° 1 (2002): 343-44.

méthodes et des acteurs utilisant des méthodes plus traditionnelles de l'archivistique, mais réfléchissant à la nécessité de les faire évoluer.

L'une des particularités de cet ensemble d'articles est d'être très majoritairement le fruit d'archivistes de formation, issus d'institutions qui ont pour fonction la production d'instruments d'érudition : archives nationales françaises ou italiennes, archives départementales, archives du Vatican, École Nationale des Chartes, IRHT. Cependant, cette problématique est aussi partagée par les historiens, et en particulier par deux d'entre eux qui proposent des interventions dans lesquelles ils soulignent la nécessité de ces instruments pour leur domaine de recherche. Par exemple, l'historien polonais Jerzy Kłoczowski (1924-2017) insiste dans sa contribution sur l'intérêt qu'il peut y avoir à disposer d'un index à défaut d'avoir accès aux sources, en particulier pour les historiens mettant en œuvre des méthodes quantitatives⁵⁰¹. Michel Parisse (1936-2020), quant à lui, présente l'une des utilisations possibles d'un index onomastique enregistré sur support magnétique, en montrant l'intérêt de cette nouvelle forme de consultation automatisée⁵⁰².

L'autre particularité de ce premier ensemble d'articles est de s'inscrire dans un certain type d'usages des ordinateurs, qui visent à la constitution d'instruments de recherche documentaire complémentaire des éditions de sources sur support papier plus traditionnel. Ce type d'instrument n'est pas attaché à un type de problématique historique précise mais a pour finalité d'être le plus générique possible et de permettre à de nombreux chercheurs de s'en emparer.

Au sein de ce premier ensemble d'articles, il est cependant possible de déceler d'autres intérêts que ceux initialement visés par les organisateurs. On retrouve par exemple dans les actes du colloque plusieurs textes qui présentent des projets qui s'intéressent à la « *langue* » de ces ensembles documentaires plutôt qu'à leur « *contenu* ». C'est le cas de deux communications du philologue néerlandais Félicien de Tolleneare (1912-), qui s'intéresse aux chartes néerlandaises non pour les informations qu'elles

⁵⁰¹ Jerzy Kłoczowski, « Les pays de l'Est de la chrétienté occidentale dans les Archives du Vatican (période médiévale) », *Publications de l'École Française de Rome* 31, n° 1 (1977): 37-41.

⁵⁰² Michel Parisse, « Traitement des documents diplomatiques : exploitation de l' « Index Verborum » », *Publications de l'École Française de Rome* 31, n° 1 (1977): 335-45.

contiennent, mais pour étudier le moyen-néerlandais⁵⁰³, ou encore du travail de Maria Giovanna Arcamone (1936-), philologue à l'université de Pise⁵⁰⁴.

Genicot participe lui aussi au colloque de Rome, dont il assure même la conclusion, nous y reviendrons. Il y présente aussi un projet d'élaboration d'un instrument de travail portant sur une collection de chartes, qui a pour finalité de rendre possible aussi bien l'étude du « contenu » que de la « langue » des sources⁵⁰⁵. Les instruments de travail qu'élaborent ces acteurs prennent ici la forme de dépouillement de tous les mots, mais aussi de concordances, dont nous avons évoqué l'utilité pour les historiens dans le chapitre précédent. Ce type de projet atteste du dépassement possible de la distinction opérée par les organisateurs des actes entre « langue » et « contenu », et donc, de l'intérêt d'analyser cette rencontre de Rome pour l'observateur que nous sommes.

Un autre élément marque la diversité des intérêts des acteurs de ce premier ensemble d'articles. Il s'agit des champs de recherche dans lesquels s'inscrivent les différents intervenants. En plus des philologues déjà mentionnés, plusieurs historiens du droit sont par exemple engagés dans la constitution d'instrument de travail à l'aide de moyens électroniques. En raison de leurs intérêts scientifiques, ils ne s'intéressent pas aux mêmes aspects des documents que leurs collègues, ce qui implique la nécessité d'élaborer des instruments de travail adaptés. C'est l'ambition que poursuit le travail mené par les médiévistes Josette Metman (1918-1986) et Bernadette Auzary (1947-) à l'aide de leur équipe à l'université de Bourgogne⁵⁰⁶, ou encore celui des considérations sur les archives judiciaires développées par le juriste italien Antonio Padoa Schioppa (1937-), qui insiste dans sa contribution sur l'importance de la langue des actes (latin ou langues vernaculaires) pour l'histoire du droit⁵⁰⁷.

⁵⁰³ Félicien de Tollenaere, « Les chartes néerlandaises en langue vulgaire », *Publications de l'École Française de Rome* 31, n° 1 (1977): 105-6.

⁵⁰⁴ Maria Giovanna Arcamone, « Per lo spoglio elettronico dell'onomastica italiana altomedievale », *Publications de l'École Française de Rome* 31, n° 1 (1977): 193-97.

⁵⁰⁵ Léopold Genicot, « Le traitement électronique des textes diplomatiques belges antérieurs à 1200 », *Publications de l'École Française de Rome* 31, n° 1 (1977): 97-104.

⁵⁰⁶ Josette Metman et Bernadette Auzary, « Les archives judiciaires et leur traitement informatique », *Publications de l'École Française de Rome* 31, n° 1 (1977): 361-77.

⁵⁰⁷ Antonio Padoa Schioppa, « Fonti giuridiche e informatica », *Publications de l'École Française de Rome* 31, n° 1 (1977): 121-23.

1.3.2. Instruments de travail spécifiques.

Un second ensemble d'articles concerne plus spécifiquement des projets d'historiens, que l'on peut caractériser par les institutions au sein desquelles ils travaillent : universités, centres de recherches du type du CNRS ou de l'Institut Historique Allemand. Les projets présentés dans ce deuxième ensemble de textes se caractérisent par une utilisation des ordinateurs qui ne vise pas à l'établissement d'instruments de travail génériques, mais, par contraste, à l'élaboration de banques de données extraites des sources en fonction d'une problématique historique prédéfinie.

L'exemple le plus pertinent concerne à ce titre les articles qui traitent d'enquêtes prosopographiques, que nous mentionnerons rapidement ici avant d'y revenir plus longuement dans la deuxième partie de ce chapitre. Françoise Autrand (1932-) et Genet, tous les deux membres de l'ERA 713 de l'université Paris 1, présentent leurs travaux de recherche, en insistant sur l'intérêt des ordinateurs pour l'histoire sociale et la constitution de banques de données regroupant des informations biographiques⁵⁰⁸. Il faut ajouter à ces deux enquêtes un projet présenté par le médiéviste Werner, directeur de l'Institut Historique Allemand, qui s'inscrit dans la lignée de ces approches prosopographiques, et sur lequel nous reviendrons également⁵⁰⁹.

L'enquête du *catasto*, présentée par Christiane Klapisch-Zuber (1936-) dans un article des actes du colloque, et dont nous avons proposé l'étude dans le chapitre 1 de cette thèse, s'inscrit, elle aussi, dans ce second groupe d'articles⁵¹⁰, tout comme celui de Monique Zerner (1935-), qui prépare une thèse sur la base d'un corpus de cadastre du Comtat Venaissin, dirigée par Georges Duby à l'université d'Aix-Marseille⁵¹¹. Deux autres interventions concernent encore l'intérêt de ce type d'usages des ordinateurs pour l'histoire économique. C'est le cas du travail présenté par Jacques Lefort (1939-2014) sur la fiscalité byzantine, mené à l'École Pratique des Hautes Études⁵¹² et de l'article de

⁵⁰⁸ Françoise Autrand, « Le personnel du Parlement de Paris », *Publications de l'École Française de Rome* 31, n° 1 (1977): 239-43; Jean-Philippe Genet, « Histoire sociale et ordinateur », *Publications de l'École Française de Rome* 31, n° 1 (1977): 231-37.

⁵⁰⁹ Werner, « Problèmes de l'exploitation des documents textuels concernant les noms et les personnes du monde latin (IIIe-XIIe siècles) ».

⁵¹⁰ Christiane Klapisch-Zuber, « Exploitation démographique et anthroponymique du *catasto* florentin de 1427 », *Publications de l'École Française de Rome* 31, n° 1 (1977): 215-23.

⁵¹¹ Monique Zerner, « Un type de document répétitif fournissant des données quantitatives », *Publications de l'École Française de Rome* 31, n° 1 (1977): 157-64.

⁵¹² Jacques Lefort, « Analyse automatique des documents fiscaux byzantins », *Publications de l'École Française de Rome* 31, n° 1 (1977): 277-89.

Michel Mollat du Jourdin (1911-1996), directeur d'étude à la IV^e section de la même institution, sur les comptes portuaires⁵¹³.

Ce type de matériau ainsi constitué, que Genet qualifie dans l'une de ses contributions de « *métasource* »⁵¹⁴, ne peut être assimilé, selon nous, au type d'objet initialement visé par les éditeurs du colloque, pour au moins deux raisons. Premièrement, ces projets ne visent pas à donner accès plus rapidement à d'autres chercheurs le contenu d'une grande collection de documents, puisqu'ils peuvent porter sur des ensembles de sources de dimensions plus petites. Deuxièmement, les instruments élaborés dans ces contextes ont vocation à permettre la poursuite d'un unique projet scientifique, sans pour autant que le matériau d'analyse produit ait vocation à être partagé en dehors du collectif de travail. Les possibilités de circulation de ces instruments de travail sont dès lors tout à fait différentes de celles des instruments de recherches documentaires que cherchent à produire les collectifs du premier ensemble d'article décrit plus haut.

Nous verrons dans les deux prochaines parties de ce chapitre que cette problématique est au cœur des discussions du colloque, puisqu'elle engage des questions portant sur l'intérêt de ce type d'instruments de travail pour la discipline, sur les modalités de l'enregistrement des données et sur la standardisation de ces méthodes.

Un dernier article atteste de l'élargissement du thème du colloque, puisqu'il porte non sur des sources dites « documentaires », mais sur des sources « littéraires ». Il s'agit d'une intervention de Genet qui porte sur le module d'analyse automatique des contextes du programme ALINE, étudié dans la troisième partie du chapitre précédent⁵¹⁵. En introduction, Genet y défend l'idée que n'importe quel document peut être approché pour sa « *langue* » ou pour son « *contenu* », et met en avant le fait que seule la problématique poursuivie par les médiévistes conditionne les formes des instruments de travail produits par eux. Cet article révèle là-encore que la distinction entre étude du contenu de grandes séries de sources et étude de la langue de corpus d'œuvres littéraires, qui apparaissait centrale pour distinguer l'initiative de Genicot de celle de Fossier, est elle-même remise en cause par les acteurs pendant le colloque.

⁵¹³ Michel Mollat du Jourdin, « Comptes portuaires et informatique », *Publications de l'École Française de Rome* 31, n° 1 (1977): 149-56.

⁵¹⁴ Genet, « Histoire sociale et ordinateur », 232.

⁵¹⁵ Jean-Philippe Genet, « Ordinateur, lexique, contexte », *Publications de l'École Française de Rome* 31, n° 1 (1977): 297-317.

Ces différents exemples nous montrent qu'au contraire de l'intention initiale des organisateurs, les échanges qui prennent forme au moment du colloque de Rome de 1975 débordent largement des discussions prévues au départ, qui devaient porter sur les difficultés posées par un seul type d'usages des ordinateurs en histoire médiévale. À la lumière de cette conclusion, les moyens électroniques de calculs apparaissent comme un vecteur de déclenchement de discussions entre les médiévistes à propos de l'organisation de leur discipline. Pour l'observateur, les actes du colloque de Rome se présente ainsi comme une formidable source pour étudier et analyser dans le détail les problématiques communes qui se posent à ces différents collectifs, en raison du fait qu'ils utilisent, ou projettent d'utiliser, des ordinateurs.

2. Problématiques communes : les échanges du colloque de Rome.

Le colloque de Rome est, comme nous venons le voir, le moment d'une rencontre entre de nombreux collectifs ne partageant ni les mêmes objectifs scientifiques, ni les mêmes fonctions au sein de la discipline historique. Pour éclairer ce qui les rassemble en dépit de cette diversité, il nous faut repérer les problématiques communes dont débattent les intervenants présents à Rome, non seulement pour 1) saisir les difficultés qu'ils partagent, et donc, ce qui les réunit, mais aussi pour 2) relever les multiples ambitions attachées à l'utilisation des ordinateurs par ces divers acteurs. En effet, ce sont ces deux éléments qui poussent ces collectifs à s'organiser à l'échelle méta-collective à l'issue du colloque, selon des modalités sur lesquelles nous reviendrons plus en détail dans la troisième partie de ce chapitre.

La table des matières des actes du colloque de Rome s'organise en trois parties, qui correspondent aux trois journées de cette rencontre en mai 1975 : 1) « Les historiens face aux grandes séries de documents » ; 2) « L'informatique et la problématique historique » et 3) « Traitement informatique de documents historiques ». Cette organisation, sur laquelle nous nous appuyons pour notre propre analyse, nous permet de saisir les axes de débats dégagés par les éditeurs des actes du colloque. Les 51 textes qui composent ces actes nous permettront de saisir la teneur des échanges, des propositions et des discussions sur chacun de ces points. Nous verrons cependant que l'imbrication entre ces trois problématiques conduit de nombreux auteurs à aborder

plusieurs d'entre elles dans leurs articles. Pour cette raison, nous nous autoriserons à inclure, au cours de notre analyse, des éléments d'articles publiés en dehors des parties *a priori* concernées.

2.1. Archives en série, collections homogènes et documents répétitifs.

Dans l'avant-propos des actes, les éditeurs affirment que :

*« l'historien d'aujourd'hui a un grand besoin d'une documentation sans cesse renouvelée, il la lui faut aussi large et aussi exhaustive que possible parce que ses préoccupations (étude des sociétés, des mentalités, prosopographie) lui font dédaigner le document rare et l'incitent à recourir aux sources sérielles. »*⁵¹⁶

Sans revenir ici sur les « *préoccupations* » des historiens visées ici par les éditeurs, que nous analyserons plus longuement au 2.2., cette citation nous renseigne sur l'une des volontés des acteurs dans ce contexte : donner accès aux historiens à une documentation « *aussi large et aussi exhaustive que possible* ». Pour ce faire, la principale difficulté à laquelle sont confrontés les acteurs tient à la dimension des collections d'archives qui les intéressent. Pour prendre d'autres exemples que les 970 000 actes des archives du Vatican mentionnés plus haut, on citera par exemple les « *dizaines de milliers de chartes* » évoquées par Genicot dans sa présentation d'un projet d'édition sur cartes perforées des documents diplomatiques belges antérieurs à 1200⁵¹⁷, ou encore l'échantillon d'actes notariés en cours de traitement par les Archives Nationales françaises qui comporte 70 000 documents⁵¹⁸.

L'autre difficulté centrale est corrélée à la dimension des collections, mais concerne la richesse des documents historiques qui les composent, sur la base desquels il est possible d'extraire une infinité d'informations (lexicologiques, codicologiques, biographiques, onomastiques etc.). Or, nous l'avons vu dans les deux premiers chapitres,

⁵¹⁶ Fossier, Vauchez, et Violante, « Avant-propos », 6.

⁵¹⁷ Genicot, « Le traitement électronique des textes diplomatiques belges antérieurs à 1200 », 98.

⁵¹⁸ Ivan Cloulas, « L'inventaire automatisé des actes notariés : principes d'analyse et résultats d'expérimentation aux Archives Nationales de Paris », *Publications de l'École Française de Rome* 31, n° 1 (1977): 127.

il est impossible, pour des raisons matérielles comme intellectuelles, d'enregistrer sur cartes perforées autre chose qu'un modèle, qu'une représentation du texte du document. Dans ce contexte, comment établir des instruments de travail donnant accès de manière « *aussi exhaustive et aussi large que possible* » à des collections de documents de ces dimensions ?

2.1.1. Un modèle de critique des ensembles documentaires emprunté aux historiens modernistes.

Pour répondre à cette question, l'une des problématiques dont débattent les collectifs présents à Rome concerne les caractéristiques des différents genres d'ensembles documentaires auxquels ils s'intéressent. Cette problématique est centrale pour les acteurs, puisque ces caractéristiques conditionnent à la fois les logiques de sélection 1) des parties de la documentation qui pouvaient intéresser les historiens et 2) des informations contenues dans ces documents qui pouvaient alimenter leurs études. Cette question est abordée au cours de la première journée du colloque, qui correspond, nous l'avons dit, à la première partie des actes⁵¹⁹. Cette partie regroupe presque exclusivement des articles de collectifs qui visent à mettre à la disposition des chercheurs des instruments de travail génériques (index, dépouillement, inventaire), tels que nous les avons évoqués plus haut.

Il faut d'abord insister ici sur le fait que les genres d'ensembles documentaires sur lesquels travaillent ces collectifs sont de natures variées. La première partie des actes du colloque se compose ainsi de cinq sous-parties, chacune consacrée à l'un des genres suivants : 1) « les registres pontificaux », 2) « les chartes », 3) « les archives judiciaires et sources juridiques », 4) « les documents notariés » et 5) ce que les éditeurs ont rassemblé sous le nom de « documents répétitifs et quantitatifs », parmi lesquels on retrouve entre autres les comptes portuaires, les listes de baptêmes, les cadastres et les censiers⁵²⁰.

Pour mener l'analyse des caractéristiques de ces ensembles éclectiques, il est possible de repérer que les acteurs s'appuient sur un modèle de critique commun,

⁵¹⁹ Lucie Fossier, « Informatique et documents médiévaux », *Annales* 31, n° 6 (1976): 1131.

⁵²⁰ Pour une définition de ces différents ensembles documentaires voir par exemple Olivier Guyotjeannin, *Les sources de l'histoire médiévale* (Paris, France: Librairie générale française, 1998).

développé par des historiens modernistes français, spécialistes de l'histoire sociale et impliqués dans la mise en œuvre d'approches quantitatives de leurs corpus⁵²¹. Parmi eux, on peut notamment citer les deux historiens modernistes François Furet (1927-1997) et Adeline Daumard (1924-2003), qui sont parmi les premiers à avoir publié en France un article à propos de l'intérêt des moyens automatiques pour la poursuite de ce programme de recherche⁵²². En reprenant ce modèle, les médiévistes recyclaient donc pour leurs propres sources les critères développés par des historiens pour d'autres périodes historiques. Ils recyclaient encore une approche des sources attachée à ces critères, pour laquelle les modalités de constitution des corpus dépendaient de la volonté des historiens d'y appliquer des procédés mathématiques et statistiques.

Dans un autre article célèbre de 1971, Furet, alors directeur d'étude à l'École des Hautes Études en Sciences Sociales (EHESS), présentait la spécificité des relations entre les sources et les corpus qui intéressent l'histoire quantitative en ces termes :

« Dans la mesure où l'histoire quantitative suppose l'existence et l'élaboration de longues séries de données homogènes et comparables, le premier problème qui se pose en termes nouveaux est celui des sources. D'une manière générale, les archives européennes ont été constituées et classées au XIX^e siècle selon des procédures et des critères qui reflètent les préoccupations idéologiques et méthodologiques de l'histoire à cette époque [...]. Au contraire, les données de l'histoire quantitative ne renvoient pas à un insaisissable découpage externe du « fait », mais à des critères de cohérence interne : le fait n'est plus l'événement sélectionné parce qu'il scande les temps forts d'une histoire dont le « sens » a été préalablement défini, mais un phénomène choisi et éventuellement construit en fonction de son caractère répétitif, donc comparable à travers une unité-temps. »⁵²³

⁵²¹ Nous ne reviendrons ici ni sur les multiples visages de l'histoire quantitative, ni sur les discussions dont elle a été l'objet chez les observateurs, mais nous nous bornerons à renvoyer à quelques-uns des articles désormais classiques qui traitent de ces questions du point de vue des acteurs : Pierre Chaunu, « Histoire quantitative ou histoire sérielle », *Cahiers Vilfredo Pareto* 2, n° 3 (1964): 165-76; François Furet, « Histoire quantitative et construction du fait historique », *Annales* 26, n° 1 (1971): 63-75; Bernard Lepetit, « L'histoire quantitative : deux ou trois choses que je sais d'elle », *Histoire & Mesure* 4, n° 3 (1989): 191-99.

⁵²² Voir Adeline Daumard et François Furet, « Méthodes de l'Histoire sociale : les Archives notariales et la Mécanographie », *Annales* 14, n° 4 (1959): 676-93.

⁵²³ Furet, « Histoire quantitative et construction du fait historique », 66.

Dans cette citation, Furet met en lumière, entre autres choses, plusieurs éléments fondamentaux pour saisir la spécificité des logiques de constitution des corpus des historiens quantitativistes. Le premier concerne le rapport entre les logiques de constitution des archives européennes depuis le XIX^e siècle et les logiques de constitution de ce type de corpus. Cette tension s'articule par ailleurs avec le second élément mis en lumière par Furet, qui concerne les caractéristiques de ces corpus : 1) leurs caractères sériels ; 2) l'homogénéité des documents qui les composent et 3) le caractère répétitif, ou la « *cohérence interne* » des éléments qui les constituent.

C'est justement sur la base de ces trois éléments que les intervenants du colloque de Rome examinent les genres d'ensembles documentaires sur lesquels ils travaillent. Cependant, la nature de leur documentation diffère de celles utilisées par les modernistes. Sur ce sujet, il s'agissait donc à Rome de débattre de la pertinence des critères dégagés par les historiens modernistes pour les archives médiévales.

2.1.2. Application aux sources médiévales.

Dans un article publié en 1976 dans la revue les *Annales*, Fossier propose un compte-rendu du colloque de Rome. Elle y établit une distinction entre deux types de collections : les documents conservés dans les archives sous formes de séries (registres pontificaux, chartes, actes notariés) et les « *documents réellement sériels* »⁵²⁴ comme les comptes portuaires, les cadastres, les censiers ou les registres de baptêmes. La différence établie ici par Fossier renvoie en réalité à deux sens différents du terme série, et nous permet de nous interroger avec les intervenants du colloque sur le caractère sériel des archives médiévales.

Le premier sens est attaché au contexte de l'archivistique, et désigne une forme de subdivision d'un classement plus général des fonds d'archives. Il renvoie directement aux logiques d'archivages des collections de sources médiévales mis en œuvre depuis le début du XIX^e siècle en France et mentionnées par Furet⁵²⁵. Les registres pontificaux ont par exemple été classés en série par pontificat, ou par origine des actes (Vatican, Avignon,

⁵²⁴ Fossier, « Informatique et documents médiévaux », 1133.

⁵²⁵ Voir par exemple pour le cas des archives départementales : Julie Lauvernier, « Mettre en ordre les archives des départements : genèse et élaboration du cadre de classement des Archives départementales », *Gazette des archives* 229, n° 1 (2013): 23-40.

Latran). C'est sur la base de cette logique de classement que ces documents ont été mis à la disposition des historiens jusque-là, sous forme de « *fichiers de noms* », de « *publication in extenso* », de « *dépouillement* » ou d'« *analyses de l'ensemble des documents relatifs à certains pontificats* »⁵²⁶.

Le second renvoie plus directement au terme tel qu'il est utilisé par Furet dans son article de 1971, à savoir, comme désignant un corpus constitué d'une suite continue et limitée de « *données* » de même nature. Cette tension entre les logiques d'archivage des sources et la constitution de corpus de données est au cœur des discussions du colloque, puisque les intervenants débattent justement des logiques de sélection des documents et des éléments de ces documents qui seront mis à la disposition des historiens via de nouveaux instruments de travail.

Dans l'introduction de leur article, les deux historiennes du droit de l'université de Bourgogne, Metman et Auzary, présentent ce problème pour le cas des archives judiciaires :

« Qu'entend-on par Archives judiciaires ? Une définition large peut en être donnée : ensemble des documents conservés dans des greffes, c'est-à-dire minutes, registres et sacs de procédure. Cette simple énumération montre la richesse et la complexité d'une telle documentation. Si les minutes, normalement enregistrées, les pièces de procédure, parfaitement hétérogènes (il s'y trouve des chartes, des actes notariés, des mémoires, des articles. . .), n'ont pas à être considérées ici, les registres constituent-ils une source homogène ? Pas davantage : rien de commun entre les procès-verbaux, les sentences, les appointements de procédure. Les registres de sentences, du moins, correspondent-ils à la notion d'archives sérielles, si tentante pour l'historien informaticien ? Loin de là : aussi divers dans le fond qu'apparemment stéréotypés dans la forme, les jugements n'obéissent à aucune règle dans l'exposé ni le dispositif »⁵²⁷.

⁵²⁶ Fossier, « Informatique et documents médiévaux », 1132.

⁵²⁷ Metman et Auzary, « Les archives judiciaires et leur traitement informatique », 361.

Le raisonnement développé ici par nos deux auteures nous éclaire sur la complexité des relations entre logiques d'archivage, typologie des sources et caractères sériele et homogène des collections d'archives. Au sein de l'ensemble « *Archives judiciaires* », les deux historiennes mettent en avant non seulement la diversité des types de dossiers existants (minutes, registres, sac de procédures) mais aussi l'hétérogénéité des documents contenus dans ces unités archivistiques (chartes, actes notariés, sentences, procès-verbaux, etc.).

Par ailleurs, elles insistent sur le fait que les « *registres de sentences* », forme d'unité archivistique qui les intéressent en particulier, ne sauraient être considérés comme des sources sérielles en raison de l'absence 1) d'une structure commune à l'ensemble de ces jugements et 2) de régularité dans les informations que ces documents contiennent (« *le fond* »). Cette citation extraite de l'article de Metman et Auzary nous renseigne sur l'existence de différents niveaux d'analyse des collections de sources, pour lesquels les auteurs s'attachent à différentes caractéristiques. L'analyse de l'homogénéité des collections repose ainsi sur l'examen des types de sources contenus dans les dossiers d'archives. L'analyse du caractère sériel des archives est quant à elle corrélée à l'analyse du caractère répétitif de la structure et du contenu des documents.

Au contraire des ensembles d'archives utilisés par les modernistes, les « *registres de sentences* » ne remplissent cependant aucune de ces deux conditions. Dès lors, les deux médiévistes formulent le problème auquel elles font face en ces termes :

« *Comment, dans ces conditions, peut-on espérer traiter exhaustivement cette riche matière, comment la faire entrer dans un cadre systématique propice à une automatisation ?* »⁵²⁸

Cette question soulève l'un des éléments centraux pour comprendre l'articulation entre l'analyse des caractéristiques des genres d'ensembles et la volonté de traiter exhaustivement ces riches collections d'archives. Il s'agit de la nécessité, pour créer ces instruments de travail, de créer un « *cadre systématique propice à une automatisation* ».

⁵²⁸ Metman et Auzary, 361.

2.1.3. Élaboration d' « un cadre systématique propice à une automatisation ».

La sous-partie dédiée aux chartes nous permet de repérer la teneur des discussions sur ce type de « *cadre* » pour le cas des ensembles d'archives hétérogènes. Dans son article, Bautier propose par exemple un modèle de description « *classique* » de la « *forme des chartes* »⁵²⁹, basé sur les aspects externes et internes de ce type de document. Comme il le précise, ce modèle reprend en fait les critères de description dégagés dès le XVIII^e siècle par Jean Mabillon et les Bénédictins. Dans la contribution de Bautier, la finalité de cette description est cependant d'une nouvelle nature. En effet, ce modèle a ici pour objet de relever les « *éléments distinctifs* » des chartes, dans le but que les « *spécialistes chargés de leur exploitation informatique* » puissent les « *traiter à bon escient* »⁵³⁰. Dans le contexte du colloque, ce modèle de description sert donc de base pour échanger sur le type d'informations que ces spécialistes doivent inclure dans leurs instruments de travail. Est-il possible d'enregistrer tous ces « *éléments distinctifs* » en ordinateur ? Si non, lesquels faut-il privilégier ? Nous reviendrons sur ces questions au 2.2.

À cette description classique des chartes, orientée vers de nouvelles fins, Bautier adjoint par ailleurs dans sa communication une liste de « *critères de classement des chartes* », dans laquelle il détermine dix grandes articulations de cet « *ensemble fortement hétérogène* »⁵³¹. Ce second aspect de son article atteste d'un autre intérêt que le repérage d'éléments descriptifs, puisqu'il s'agit de proposer de nouveaux critères de classement des sources, basés non plus sur leurs caractères formels, mais sur les informations qu'elles contiennent. Parmi ces critères, on retrouve par exemple « *la forme de l'engagement des parties contractantes* »⁵³², « *les personnes qui ont assumé la confection de l'acte écrit* »⁵³³, « *les personnes ayant concouru à l'élaboration de l'acte* »⁵³⁴, etc.

À travers l'énonciation de ces critères, Bautier propose en fait une autre composante constitutive d'un « *cadre* » du type de celui auquel Metman et Auzary faisaient allusion. Elle doit permettre aux archivistes engagés dans la création

⁵²⁹ Robert-Henri Bautier, « Caractères spécifiques des chartes médiévales », *Publications de l'École Française de Rome* 31, n° 1 (1977): 88.

⁵³⁰ Bautier, 83.

⁵³¹ Bautier, 83.

⁵³² Bautier, 86.

⁵³³ Bautier, 84.

⁵³⁴ Bautier, 83.

d'instruments de recherche d'enregistrer exhaustivement non l'ensemble du contenu des chartes, mais seulement certains de leurs éléments. Pour le cas des chartes, l'enjeu est notamment de multiplier les types d'index existants, ce qui suppose de dégager, sur la base des éléments répétitifs de ces collections de sources, de nouvelles logiques de classement des documents. En plus des traditionnelles tables (index de noms de personnes et de lieux), dont bénéficient certaines collections d'archives, Bautier propose dans son article d'utiliser les fonctions de classement des ordinateurs pour fournir aux chercheurs de nouveaux instruments de recherche que l'on pourrait présenter ainsi : index des actes en fonction de la forme de l'engagement des parties contractantes, index des actes en fonction des personnes qui en ont assumé la confection, etc.

Les discussions sur l'élaboration de ce cadre se posent dans des termes similaires pour les ensembles documentaires « *véritablement sériels* » mentionnés par Fossier. L'un des exemples les plus marquants est à ce titre celui des comptes portuaires, que l'historien médiéviste Mollat du Jourdin qualifie dans son article d'ensemble documentaire « *sériel par définition* »⁵³⁵ et, concomitamment, de source parmi « *la plus apte au calcul électronique* »⁵³⁶. C'est d'ailleurs à partir de ce type d'archives qu'ont été menées, pour la période moderne, certaines des premières applications des méthodologies informatiques en histoire aussi bien en France qu'aux États-Unis⁵³⁷. Cependant, Mollat nuance cette affirmation pour les comptes portuaires du Moyen-Âge, et affirme que « *la continuité et l'homogénéité font défaut très souvent* », en raison des « *lacunes* » de ces sources⁵³⁸, mais aussi « *des mutations dans le contenu de l'information* »⁵³⁹.

Selon lui, ces limites rendent particulièrement complexe l'établissement d'un cadre de description commun pour toutes les archives de ce type, en particulier du fait de l'hétérogénéité des informations contenues dans ce genre de documents. Il reste cependant possible d'établir un « *type de fiche signalétique des navires* »⁵⁴⁰, qui regrouperaient les « *informations constantes et comparables* »⁵⁴¹, c'est-à-dire, en

⁵³⁵ Mollat du Jourdin, « Comptes portuaires et informatique », 150.

⁵³⁶ Mollat du Jourdin, 149.

⁵³⁷ Pour la France voir Jean Delumeau, « Méthode mécanographique et trafic maritime : les terre-neuviens malouins à la fin du XVIIe siècle », *Annales* 16, n° 4 (1961) : 665-85. Pour les États-Unis voir Bernard Bailyn et Lotte Bailyn, *Massachusetts shipping, 1697-1714* (États-Unis d'Amérique: Belknap Press of Harvard University Press, 1959).

⁵³⁸ Mollat du Jourdin, « Comptes portuaires et informatique », 150.

⁵³⁹ Mollat du Jourdin, 150.

⁵⁴⁰ Mollat du Jourdin, 156.

⁵⁴¹ Mollat du Jourdin, 155.

l'espèce, la mention de la présence des navires dans les ports. Il serait alors possible de constituer des « dictionnaires » des navires, qui « *rendrait des services considérables* »⁵⁴², notamment pour repérer le passage des navires dans les différents ports du monde, étudiés jusque-là séparément par les historiens depuis le début des années 1960. Mollat du Jourdin s'inscrit donc lui aussi dans un débat similaire à celui qui porte sur les chartes, en proposant la constitution d'instruments communs qui reposent sur les régularités repérées entre ces différentes sources.

Cependant, à la différence de Bautier, ces remarques ne portent pas sur les documents eux-mêmes, mais sur les objets de la réalité visés par ces documents (les navires). L'instrument de travail ainsi constitué est détaché de la logique de conservation des sources, puisqu'il doit au contraire permettre de repérer dans différents fonds d'archives la présence des mêmes navires. Les dictionnaires imaginés par Mollat du Jourdin fonctionneraient ainsi comme une forme d'index des navires, non pour une collection d'archives mais pour un ensemble disparate de sources, conservées en différents endroits du monde.

Ces discussions sur les caractéristiques des genres d'ensembles nous renseignent sur une première problématique qui anime ce groupe de médiévistes. Elle repose sur la volonté de repérer les informations communes que partagent les différents types d'ensembles de sources, ou en d'autres termes, les éléments répétitifs des documents, dans le but de proposer de nouveaux instruments permettant la recherche de ces informations au sein de ces ensembles. Comme nous l'avons vu plus haut, pour poursuivre cet objectif, les collectifs présents à Rome font encore des propositions visant à harmoniser les normes de description des sources. Ils élaborent encore de nouvelles logiques de classement de ces documents qui ne reposent plus seulement sur la typologie des sources mais sur leur contenu.

Ce type de discussion qui apparaît dans le contexte du colloque de Rome constitue l'une des manifestations des ambitions de réorganisation de la discipline engendrées par la volonté d'utiliser des ordinateurs. Pour éclairer les ambitions corrélées à ces discussions dans ce contexte, il est cependant nécessaire de s'intéresser de plus près aux finalités des instruments de travail qu'élaborent ou que projettent d'établir les acteurs, pour comprendre les usages qu'ils en attendent.

⁵⁴² Mollat du Jourdin, 156.

2.2. Instruments de travail et problématiques historiques.

Pour comprendre comment ce premier axe de discussion s'articule avec les « préoccupations »⁵⁴³ des historiens, il faut regarder la deuxième partie des actes du colloque de Rome, intitulée « Les demandes des historiens à l'informatique ». Les neuf articles qui composent cette partie sont organisés en quatre sous-parties, qui concerne chacune un domaine de l'histoire médiévale : 1) « Diplomatique », 2) « Onomastique », 3) « Histoire démographique » et 4) « Prosopographie et histoire sociale ». Y sont présentés deux types d'instruments de travail en cours d'élaboration ou déjà existants : 1) des instruments de recherche génériques, destinés à permettre à différents spécialistes de poursuivre un large éventail de problématiques historiques, du type de ceux proposés pour les chartes et 2) des instruments de travail spécifiques, constitués dans le but de poursuivre des problématiques spécialisées, du type de ceux visés par Mollat du Jourdin.

2.2.1. Diplomatique.

Les deux articles de la sous-partie consacrée à la diplomatie portent sur les attentes des historiens qui travaillent au contact des actes écrits, et particulièrement des chartes. Dans un article intitulée « Les demandes des historiens à l'informatique : la forme diplomatique et le contenu juridique des actes », Bautier dégage différents types d'éléments qu'il peut être pertinent d'enregistrer en machine : 1) les caractères associés à la tradition (chancellerie, période de rédaction, auteur, destinataire des actes, scribe en charge de la rédaction) ; 2) les caractères permettant la datation des actes (témoins des documents, lieu de rédaction) ; 3) les noms de lieu mentionnés dans les actes (lieux-dits, paroisses, communes, circonscriptions) ; 4) les noms de personnes mentionnés dans les actes et les professions, qualités, titres ou liens de parenté attachés à ces noms ; 5) les mots et concepts de nature juridique et 6) le contenu des formules des actes.

Bautier est cependant conscient de l'impossibilité d'enregistrer en plein texte (*full text*), c'est-à-dire sous la forme de chaînes de caractères, l'ensemble du matériau contenu dans toutes les chartes :

⁵⁴³ Fossier, Vauchez, et Violante, « Avant-propos », 6.

« *Le traitement des centaines de milliers (ou de millions) de chartes conservées dans les chartriers et les cartulaires, les millions d'actes notariés méridionaux - plus de 4.000.000 d'actes antérieurs à 1500 dans les seuls départements français de Provence, du Comtat et du Dauphiné - exigent le recours à des méthodes qui relèvent de l'informatique, mais qui, excluant le full text, donnent la priorité à l'indexage.* »⁵⁴⁴

L'indexage, ou indexation documentaire, solution technique sur laquelle nous reviendrons au **2.3.**, permettrait selon lui de constituer des instruments de travail à même d'étudier certains des éléments qu'il avait distingués, et que nous avons présentés ci-dessus. Cette question quant à la méthode d'enregistrement se pose cependant de manière plus délicate pour l'enregistrement des formules des actes.

Dans son article de 1976 précédemment cité, Fossier évoque ce problème. Elle affirme que Bautier considère, au moins pour les collections antérieures au XI^e siècle, les formules des chartes comme « *une gangue souvent artificielle et sclérosée* »⁵⁴⁵. Pour ce dernier :

« *La formulation doit, en effet, être connue de l'historien pour être par lui éliminée et lui permettre d'aller en toute sécurité au fait même de l'acte* »⁵⁴⁶.

Pour lui, il n'est donc pas nécessaire de les enregistrer en machine, mais seulement de donner une indication sur la présence de formulations « *normales* »⁵⁴⁷ ou « *anormales* »⁵⁴⁸, puisque c'est au contraire « *la substance* », c'est-à-dire les cinq autres types d'éléments présentés ci-dessus, qui constituerait « *le noyau véritablement utile* »⁵⁴⁹ de ces documents.

⁵⁴⁴ Robert-Henri Bautier, « Les demandes des historiens à l'informatique », *Publications de l'École Française de Rome* 31, n° 1 (1977): 180.

⁵⁴⁵ Fossier, « Informatique et documents médiévaux », 1134.

⁵⁴⁶ Bautier, « Les demandes des historiens à l'informatique », 182.

⁵⁴⁷ Bautier, 182.

⁵⁴⁸ Bautier, 183.

⁵⁴⁹ Fossier, « Informatique et documents médiévaux », 1134.

Fossier, qui travaille justement à l'indexation documentaire d'un ensemble de chartes au CRAL de Nancy, s'oppose à cette déclaration de Bautier en ces termes dans son article de 1976 :

*« Étudier les formules qui ne varient pas avec l'objet de l'acte, mais seulement avec la chancellerie, l'époque, etc., c'est tenter de percer, au travers d'elles, l'évolution des institutions et du droit ; il est donc nécessaire après les avoir recensées et enregistrées en machine, de procéder à une étude comparative permettant de dégager une certaine évolution de ces formules et d'en tirer les conclusions. »*⁵⁵⁰

Ce que souligne Fossier dans ce passage, c'est que l'élément des sources que Bautier propose de laisser de côté ici, les formules, peut constituer pour certains historiens une source d'information tout à fait indispensable.

Ce type d'échange, qui prend forme au colloque de Rome, soulève l'un des problèmes centraux qui se pose aux acteurs qui cherchent à élaborer des instruments de travail génériques. En effet, nous l'avons vu dans les deux premiers chapitres, les modalités de l'enregistrement sur cartes perforées supposent l'élaboration d'une représentation, d'un modèle des sources, plutôt qu'une reproduction des sources elles-mêmes. Cette contrainte entraîne de fait des limites dans les possibles réutilisations de ces instruments. Les conditions matérielles de la recherche (dimension des collections d'archives, moyens humains, subsides), supposent encore des choix quant aux éléments contenus dans les documents qu'il est possible d'enregistrer. Dès lors, la difficulté qu'affrontent les acteurs est d'envisager tous les types de problématiques qui peuvent être abordées sur la base des sources, dans le but d'enregistrer tous les éléments qui en permettront l'étude.

En pratique, ce problème insoluble conduit Bautier à envisager la constitution d'instruments de travail permettant de traiter deux types de problématique. Le premier anime les diplomatistes depuis les travaux de Mabillon au XVII^e siècle, et touche à la critique des documents. Il s'agit de 1) l'étude de la tradition, que Guyotjeannin définit comme l'étude des « *anneaux de la chaîne qui mène de l'original [...] à l'état dans lequel*

⁵⁵⁰ Fossier, 1134.

le document nous est parvenu », mais aussi 2) de la critique des faux et 3) de la datation des actes⁵⁵¹.

L'autre volet de problématiques qu'il est possible de dégager sur la base des éléments qu'ils proposent d'enregistrer en machine rejoint des préoccupations déjà anciennes de l'histoire sociale médiévale, mais d'une actualité brûlante pour la discipline. Il s'agit de 4) l'étude des noms de lieux, ou toponymes, qui remonte à la fin du XIX^e siècle⁵⁵², mais qui constitue au milieu des années 1970 l'un des axes de recherche permettant de mener des enquêtes sur l'occupation humaine des sols et l'essor des peuplements⁵⁵³, en s'appuyant notamment sur l'étude des cadastres et des censiers⁵⁵⁴ ; 5) des enquêtes basées sur les noms de personnes contenus dans ces actes, informations qui intéressent les historiens depuis, *a minima*, une cinquantaine d'année⁵⁵⁵, mais que le développement des approches prosopographiques⁵⁵⁶ et généalogiques⁵⁵⁷ a remis au cœur de la discipline depuis le début des années 1970 et 6) la recherche sur le vocabulaire juridique des actes qui intéressent les historiens depuis au moins le début du XX^e siècle⁵⁵⁸, mais qui constitue l'un des terrains de débat fertile pour l'histoire de l'évolution des institutions féodales notamment.

⁵⁵¹ Olivier Guyotjeannin, « La diplomatique médiévale et l'élargissement de son champ », *Gazette des archives* 172, n° 1 (1996): 12-18.

⁵⁵² Voir par exemple Elisabeth Zadora-Rio, « Archéologie et toponymie : le divorce », *Les petits cahiers d'Anatole*, n° 8 (2001). Pour un exemple des travaux de ce type à la fin du XIX^e voir Henri d'Arbois de Jubainville et Georges Dottin, *Recherches sur l'origine de la propriété foncière et des noms de lieux habités en France: période celtique et période romaine* (Paris, France: Ernest Thorin, éditeur, 1890).

⁵⁵³ Voir, entre autres : Charles Higounet et Charles Préfacier Samaran, *Paysages et villages neufs du Moyen âge: recueil d'articles* (Bordeaux, France: Fédération historique du Sud-Ouest, 1975); Robert Fossier, « Peuplement de la France du Nord entre le Xe et le XVI^e siècles », *Annales de démographie historique*, 1979, 59-99.

⁵⁵⁴ Cette approche est notamment discutée au cours du colloque de Rome dans deux communications : Robert Fossier, « Censiers et terriers », *Publications de l'École Française de Rome* 31, n° 1 (1977): 169-74 et Zerner, « Un type de document répétitif fournissant des données quantitatives ».

⁵⁵⁵ Voir par exemple Marc Bloch, « Mr Karl Michaelsson, Études sur les noms de personne français d'après les rôles de taille parisiens (rôles de 1292, 1296-1300, 1313) », *Annales* 4, n° 13 (1932): 67-69.

⁵⁵⁶ Voir par exemple le numéro de 1970 des *Annales* dans lequel trois articles sont consacrés à l'essor de ces approches : « Annales. Économies, Sociétés, Civilisations. 25^e année, N. 5, 1970. » 25, n° 5 (1970).

⁵⁵⁷ Voir en particulier les actes du colloque de 1974, sur lesquels nous reviendrons plus bas : « Famille et parenté dans l'Occident médiéval. Actes du colloque de Paris (6-8 juin 1974) » 30, n° 1 (1977).

⁵⁵⁸ Voir par exemple Léo Verriest, « Les faits et la terminologie en matière de condition juridique des personnes au Moyen âge : serfs, nobles, villains, sainteurs », *Revue du Nord* 25, n° 98 (1939): 101-27.

2.2.2. Onomastique, histoire sociale et prosopographie.

La situation est quelque peu différente pour les discussions de la sous-partie liée à l'onomastique, puisque les instruments de travail envisagés dans ce cadre ne doivent plus rendre compte d'une multitude d'éléments, mais seulement des noms de personnes ou de lieux. L'archiviste Marianne Mulon (1927-2011), directrice du Centre de toponymie des Archives Nationales françaises, a, par exemple, entrepris le regroupement au niveau national de tous les toponymes attestés aussi bien dans les sources manuscrites qu'imprimées, ensemble qui pourrait dans l'idéal contenir près de 20 millions de toponymes anciens⁵⁵⁹. La finalité de cet instrument est de permettre aux chercheurs de retrouver, dans l'immense collection de sources des archives nationales, les noms des lieux qui les intéressent et d'identifier les documents qui les mentionnent. Cet instrument, semblable à ce que l'on appelle aujourd'hui un moteur de recherche, se présente en l'espèce plutôt comme un nouveau type d'index, puisque chaque requête entraîne l'impression d'un nouveau listing contenant les références des documents concernés.

Un instrument de travail similaire est par ailleurs en projet à l'Institut Historique Allemand pour les noms de personnes du haut Moyen-Âge. Dans sa contribution aux actes du colloque, le médiéviste Karl-Ferdinand Werner, directeur de l'Institut, présente ce projet d'informatisation qui repose sur le travail déjà effectué dans le cadre de la *Prosopographia Regnorum Orbis Latini*, qui a permis de relever près de 120 000 mentions de noms de personnes dans un ensemble hétérogène d'archives en Allemagne (diplôme royaux, actes privés, lois, capitulaires, etc.).

Dans ce cadre, ce collectif projette d'élaborer, sur la base des quelque « 300 000 fiches mises en ordre alphabétiques »⁵⁶⁰ des mentions de noms de personnes, une série d'instruments de travail qui porteraient non plus sur les noms, mais bien sur les personnes identifiées dans ces sources. Dans son article, Werner indique quelques-uns des instruments qu'il envisage en lien avec cette première étape du travail : des « *instruments de travail chronologiques, comme, par exemple, des listes de fonctionnaires dans l'empire franc pour une période données* », des « *tableaux généalogiques* » et des outils

⁵⁵⁹ Marianne Mulon, « Une réalisation expérimentale du fichier automatisé des formes anciennes. Le répertoire de l'Ain », *Publications de l'École Française de Rome* 31, n° 1 (1977): 199-204.

⁵⁶⁰ Werner, « Problèmes de l'exploitation des documents textuels concernant les noms et les personnes du monde latin (IIIe-XIIe siècles) », 212.

de « « *Prosopographie* » proprement dite, donnant les détails connus de tous les personnages d'une certaine importance »⁵⁶¹.

Cet intérêt pour l'approche des toponymes et des noms de personnes rejoint ainsi la cinquième problématique dégagée ci-dessus pour la diplomatique, et nous éclaire sur l'existence d'une relation privilégiée entre les moyens informatiques et l'étude des noms. Par ailleurs, la possibilité offerte par les moyens électroniques de constituer des instruments à même d'assister l'historien dans l'identification des noms de personnes ouvrait la voie au développement de différents types d'enquêtes historiques. Dans son article, Werner insistait sur ce fait en ces termes, en mettant au cœur de ces développements les approches prosopographiques :

*« Sur la base des données historiques concernant des personnes données, il est possible de reconstituer partiellement les liens sociaux de la famille dans le sens étroit et plus large, de la couche sociale, de l'appartenance régionale, etc. La prosopographie s'avère donc comme une science auxiliaire de l'histoire au carrefour de la généalogie et de l'histoire sociale, des structures et des événements »*⁵⁶².

Le premier type d'enquêtes mentionnées ici concerne la question de l'histoire de la famille et de la parenté, qui était justement au cœur du colloque de Paris de 1974 cité plus haut. Cet événement, qui avait été organisé pour faire se rencontrer historiens et ethnologues, mettait notamment l'accent sur l'étude « *du problème capital des structures familiales* »⁵⁶³, en s'inscrivant dans la continuité des recherches menées au sein du séminaire de Duby au Collège de France sur « *Famille et Sexualité au Moyen Age* »⁵⁶⁴. Parmi les participants, on y retrouvait notamment Werner, mais aussi plusieurs autres intervenants au colloque de Rome : Violante, Vauchez, Klapisch-Zuber ou le médiéviste italien Michele Luzzati (1939-2014), professeur à l'université de Pise.

⁵⁶¹ Werner, 209.

⁵⁶² Werner, 208.

⁵⁶³ Jacques Le Goff, « Avant-propos », *Publications de l'École Française de Rome* 30, n° 1 (1977): 5.

⁵⁶⁴ Voir par exemple Georges Duby, « Présentation de l'enquête sur « Famille et sexualité au Moyen Age » », *Publications de l'École Française de Rome* 30, n° 1 (1977): 9-11.

Le second type d'enquête mentionnée par Werner dans cette citation concerne les questions propres à l'histoire sociale. Glénisson, qui ouvre la session sur la prosopographie à Rome en 1975, insiste justement dans son article sur le lien entre cette approche historique et les moyens informatiques. Selon lui, il s'est produit :

*« une sorte de rencontre providentielle, d'harmonie pré-établie entre les entreprises prosopographiques telles que l'histoire que nous écrivons aujourd'hui – à dominante économique et sociale, - les conçoit, et l'informatique. »*⁵⁶⁵

Dans la sous-partie consacrée à ce type d'approche, deux historiens, que nous avons déjà croisés dans le chapitre 2, présentent leurs travaux. Il s'agit de Genet et Autrand, qui mènent tous les deux des doctorats d'État au sein de l'ERA 713 sous la direction de Bernard Guenée (1927-2010). Glénisson inscrit leurs travaux dans un programme de recherche différent de celui de Duby, en insistant plutôt sur la filiation avec le travail du moderniste français Pierre Goubert (1915-2012), initié dans sa célèbre thèse sur Beauvais et le Beauvaisis, publiée en 1960⁵⁶⁶. Ces exemples nous renseignent sur le fait que les problématiques partagées par les participants au colloque de Rome s'inscrivent dans la continuité de travaux plus anciens, et que les moyens électroniques sont au début des années 1970 une ressource nouvelle à même de permettre l'amélioration des instruments de travail disponibles pour poursuivre ces problématiques.

Pour ce type de projets tournés vers l'histoire sociale, les problèmes sont cependant tout à fait différents de ceux qui sont soulevés dans la sous-partie sur la diplomatique, puisque l'historien fabrique lui-même un nouveau type d'archive qui est directement lié aux problématiques de recherche qui l'intéressent. Ces instruments présentent donc des limites différentes de celles discutées dans les deux autres sous-parties étudiées plus haut, et supposent donc d'autres types de discussion entre les acteurs.

⁵⁶⁵ Jean Glénisson, « Prosopographie et informatique », *Publications de l'École Française de Rome* 31, n° 1 (1977): 227.

⁵⁶⁶ Pierre Goubert et Guy Arbellot, « Beauvais et le Beauvaisis de 1600 à 1730 : contribution à l'histoire sociale de la France du XVIIe siècle » (Thèse d'État, France, Université de Paris (1896-1968). Faculté des lettres, 1958).

Dans l'article qui ouvre cette sous-partie, Glénisson insiste, entre autres, sur deux problèmes que partagent les historiens engagés dans ce type de projet. Le premier concerne la question de savoir comment « *toutes les données accumulées en mémoire [...] subsistent après l'achèvement du travail individuel* »⁵⁶⁷. Cette préoccupation rejoint l'une des ambitions développées dans le cadre de l'enquête du *catasto*, en lien avec laquelle David Herlihy (1930-1991) envisageait, dès 1966, la création d'une bibliothèque de base de données en histoire médiévale. Pour ce faire, il proposait notamment de standardiser les questionnaires qui portaient sur les sources de ce type, dans le but de permettre la comparaison entre les données extraites de lois fiscales de différentes périodes et de différentes régions. Si Glénisson ne mentionne pas dans son article de coordination de ce type, il affirme cependant la possibilité d'aboutir à des « *banques générales de données* »⁵⁶⁸ qui semblent être du même acabit que celles imaginées par Herlihy en 1966.

Le second problème soulevé par Glénisson ne concerne pas la question de la pérennisation des banques de données mais celui du « *choix des programmes* » qui en permettent l'exploitation. Il le formule en ces termes :

*« Il est bien évident que nous ne sommes pas les seuls à avoir lancé des entreprises de prosopographie à destination d'histoire sociale. Ne faudrait-il pas procéder à un recensement des programmes en Europe, comme en Amérique, et favoriser des échanges qui éviteraient les redites coûteuses en temps comme en argent ? »*⁵⁶⁹

L'intérêt de l'organisation collective proposée ici par Glénisson s'ajoute donc à ceux formulés dans la sous-partie sur la diplomatique et sur l'onomastique. Il ne s'agit plus seulement de standardiser les instruments de travail produits dans ces contextes, mais également de mettre en commun les programmes informatiques qui en permettraient l'exploitation, à l'image de l'exemple de ALINE que nous avons analysé dans le chapitre précédent.

⁵⁶⁷ Jean Glénisson, « Prosopographie et informatique », 228.

⁵⁶⁸ Glénisson, 229.

⁵⁶⁹ Glénisson, 229.

2.2.3. Étude du vocabulaire juridique et institutionnel.

Un dernier ensemble de problématiques historiques est au cœur des discussions du colloque de Rome. Il rejoint directement la sixième problématique soulevée plus haut pour la diplomatique, à savoir, l'étude du vocabulaire juridique et institutionnel. Il est cependant possible de l'étendre à d'autres types d'approches présentées au colloque de Rome, comme celle du projet ALINE, que Genet présente également à Rome⁵⁷⁰, ou des travaux sur les chartes des philologues de Tolleneare et Arcamone mentionnés au 1.3, qui visent à étudier la « *langue* » des documents. En dépit de leur différence apparente, selon les catégories des acteurs (« *langue* » vs « *contenu* »), un point commun permet toutefois de rapprocher ces derniers types de projets aux autres. Il s'agit du type d'instruments de travail produits dans ces contextes : index de mots, index de fréquences et concordances. Pour en comprendre précisément les usages, nous partirons de la contribution de Michel Parisse, directeur d'une des équipes du CRAL, intitulé « Traitement de documents diplomatiques : l'exploitation de l'« index verborum » », qui présente un aperçu de l'utilisation qui peut être faite de ce type d'outil⁵⁷¹.

L'instrument présenté par Parisse porte sur un ensemble de chartes conservées dans les archives départementales de Lorraine. Dans le cadre de ce projet, trois directions d'études sont poursuivies : une étude diplomatique, qui porte sur l'étude des « *règles du discours diplomatique* », une étude onomastique, « *qui permet de dégager des textes les noms de personnes et de lieux* » et une étude linguistique, qui vise à exploiter « *la totalité des mots* »⁵⁷². Dans son article Parisse présente certaines questions qu'il est possible de se poser sur la base de l'un des instruments de travail produit dans le cadre de cette troisième direction d'étude, l'*index verborum*, qui contient l'ensemble des termes d'un corpus de 215 documents, soit près de 90 000 formes, caractérisés par leurs fonctions grammaticales (noms, adjectifs, pronoms, etc..) et leur genre.

Pour exposer l'intérêt de cet instrument, il prend plusieurs exemples de recherches rendues possible par cet instrument. Le chercheur intéressé par la vie rurale pourrait par exemple repérer les noms des animaux domestiques mentionnés dans ces sources, et relever par-là les espèces qui existent dans cette région au Moyen-Âge (chevaux, porcs

⁵⁷⁰ Genet, « Ordinateur, lexique, contexte ».

⁵⁷¹ Parisse, « Traitement des documents diplomatiques ».

⁵⁷² Parisse, 336.

et moutons plutôt que bovins). Ce premier résultat lui permettrait notamment de calculer et de comparer la fréquence de ces différents termes dans le corpus. L'index l'aiderait encore à repérer les cotes des documents dans lesquelles ces termes apparaissent, et donc, à sélectionner dans la masse d'actes de ce corpus les documents qui pourraient l'intéresser plus particulièrement. Le chercheur aurait par ailleurs l'occasion de connaître, sur la base de l'index, la chronologie de l'emploi de ces termes, leurs géographies et même le type d'actes dans lequel ils apparaissent (donations, diplômes royaux, etc.).

Pour d'autres termes, au contenu sémantique plus mouvant, comme *altare* (autel), Parisse montre qu'il est possible d'étudier « *le changement de signification d'un vocable* »⁵⁷³, en s'appuyant sur les concordances produites par la machine. Pour le terme *altare*, Parisse remarque notamment qu'apparaît en 1019 une nouvelle expression, *trado altare* (transmettre l'autel), qui atteste de l'évolution du sens du mot : celui-ci ne désigne alors plus seulement l'autel d'un saint, mais revêt une « *signification juridique, occasion de perception de droits, objet d'exercice d'une autorité* »⁵⁷⁴.

Ce type d'outil apparaît dans ce contexte comme un formidable instrument de travail, en particulier pour l'étude sémantique des termes juridiques, institutionnels et religieux. Ce type d'études agitent justement la communauté des médiévistes, notamment parce qu'elles s'articulent avec deux grandes problématiques de l'histoire médiévale : l'Antiquité tardive, qui questionne la spécificité des sociétés européennes du bas Moyen-Âge par rapport à celle de la période précédente⁵⁷⁵ et la mutation de l'an mil, qui touche au rythme du développement des institutions féodales⁵⁷⁶. Comme pour le cas de la diplomatique ou de l'histoire sociale, l'instrument de travail présenté par Parisse se présente donc comme une nouvelle ressource pour poursuivre des questions anciennes, selon des méthodes déjà bien connues des médiévistes.

L'analyse des discussions qui prennent forme autour des « *préoccupations* » des historiens nous renseigne ainsi sur les différents domaines de l'histoire médiévale dans lesquels s'inscrivent les projets informatiques des intervenants, qui, nous l'avons vu, sont

⁵⁷³ Parisse, 338.

⁵⁷⁴ Parisse, 338.

⁵⁷⁵ Voir par exemple Henri-Irénée Marrou, *La civilisation de l'antiquité tardive* (Roma, Italie: Accademia nazionale dei Lincei, 1968) ou Peter Brown, *The world of late Antiquity: from Marcus Aurelius to Muhammad* (London, Royaume-Uni de Grande-Bretagne et d'Irlande du Nord: Thames and Hudson, 1971).

⁵⁷⁶ Voir par exemple sur cette question la synthèse suivante : Jean-Pierre Poly et Éric Bournazel, *La mutation féodale : Xe-XIIIe siècles* (Paris, France: Presses universitaires de France, 1980).

nombreux. Parmi ceux qui manquent, malgré l'intérêt que leurs portent les médiévistes dans les années 1970, on peut cependant citer l'histoire de l'art, mais aussi plus largement l'histoire dite culturelle, qui englobe aussi bien ce premier domaine que d'autres portant sur les sciences ou de la vie matérielle. Si l'absence de l'histoire de l'art s'explique en partie par le fait que l'informatisation des images se présente comme une démarche techniquement délicate, dans laquelle aucun médiéviste ne s'engage avant les années 1980⁵⁷⁷, l'absence des deux autres domaines nous pousse à nous interroger sur les raisons qui expliquent cette configuration. Est-ce lié au fait que ces domaines reposent sur des approches moins quantitatives de leurs sources ? Ou sur les frontières disciplinaires qui existent alors entre histoire des sciences, anthropologie et histoire médiévale ? Ou est-ce le résultat d'une déformation documentaire, liée à l'absence des chercheurs intéressés par l'utilisation des ordinateurs dans ces domaines parmi les intervenants du colloque de Rome ?

Cette analyse nous conduit par ailleurs à mettre en lumière l'importance des questions techniques qui demeurent l'un des enjeux centraux pour ces collectifs. En effet, une fois mis en avant le type d'éléments à extraire des sources pour les différents domaines historiques ou les différentes problématiques de recherche, il est nécessaire de définir les méthodes d'enregistrement de ces éléments en machine. En fonction des corpus et des intérêts des historiens, faut-il les enregistrer en plein texte, c'est-à-dire, comme une chaîne de caractère ? Faut-il indexer ces informations, en créant des listes de descripteurs qui regrouperaient certains termes ? Ou créer un questionnaire et extraire seulement certaines informations ?

2.3. Méthodes d'enregistrement des données en machine.

Nous avons examiné, dans les deux dernières sous-parties de ce chapitre, deux axes de discussions développés au colloque de Rome. Au fil de cette analyse, nous avons cependant aperçu que ces discussions étaient corrélées à des questions techniques, celles qui portent sur les méthodes d'enregistrement des données. Cette dimension se trouvaient

⁵⁷⁷ Pour donner un exemple, la première mention d'un projet d'élaboration d'un moteur de recherche d'image en histoire médiévale en France date, à notre connaissance, de la deuxième moitié des années 1980 : Marion Créhange, « Par et pour la recherche d'images : EXPRIM », *Le Médiéviste et l'ordinateur* 19, n° 1 (1988): 15-21.

être à la croisée des questions portant sur les caractéristiques des corpus et des problématiques qui intéressaient ces médiévistes. La troisième partie des actes du colloque est justement dédiée aux considérations de ce type. Sobrement intitulée « Traitement informatique de documents historiques », elle est composée de dix-sept articles, regroupés en quatre sections.

Dans son article de 1976, Fossier présente les différents « *traitement en machine* » de ces documents en ces termes :

*« Un document historique peut être soumis à divers traitement en machine : le texte même peut être entré directement dans l'ordinateur soit intégralement, soit sous forme d'extraits, si par exemple on veut éliminer de l'exploitations des parties du texte jugées sans intérêt : c'est ce que l'on appelle le traitement en full-text. On peut aussi analyser le contenu du document et l'enregistrer sous forme de « descripteurs » normalisés ; on peut enfin – en particulier pour les traitements statistiques – établir a priori un questionnaire et c'est dans la mesure où elles pourront fournir une réponse que les informations figureront en machine, en regard des questions, sous une forme plus ou moins codée. »*⁵⁷⁸

Cette citation nous éclaire sur la définition d'une typologie des méthodes d'enregistrement des données, qui se répartissent en trois types, que nous qualifierons en des termes empruntés aux acteurs : 1) l'enregistrement *full-text*, dont nous avons analysé les modalités, sur la base d'un exemple, dans la deuxième partie du chapitre 2 ; 2) l'enregistrement sous forme de descripteurs normalisés, méthode désignée sous le terme d'indexation documentaire, sur laquelle nous reviendrons plus précisément au 2.3.2. et 3) l'enregistrement statistique, que nous avons étudié à travers les pratiques d'un collectif dans la deuxième partie du chapitre 1.

2.3.1. Limites des enregistrements *full-text* et statistique.

Pour mettre en lien les remarques tirées des exemples développés plus haut avec cet axe de discussion, il est possible de remarquer les relations privilégiées

⁵⁷⁸ Fossier, « Informatique et documents médiévaux », 1137.

qu'entretiennent certaines problématiques et certains types de corpus avec certaines de ces méthodes d'enregistrement. Comme nous l'avons vu, l'enregistrement *full-text* va par exemple de pair avec un intérêt pour le vocabulaire des documents. Le cas le plus frappant concerne l'étude de corpus de textes politiques anglais du bas Moyen-Âge développée par Genet, que nous avons analysée en détail au chapitre 2. Pour rappel, cette recherche visait à caractériser statistiquement le discours politique d'un ensemble de textes anglais du bas Moyen-Âge, en se fondant sur une approche lexicométrique. L'exemple de l'*index verborum*, présenté par Parisse, va lui aussi dans le sens de cette remarque. En effet, pour ces deux projets, l'enregistrement intégral est inévitable, puisqu'ils visent à étudier les occurrences des mots, leurs fréquences et leurs contextes.

La pertinence de cette méthode apparaît cependant tout aussi forte pour d'autres types de problématiques. Comme nous l'avons vu au 2.2., les spécialistes du CRAL, de l'IRHT comme du CETEDOC affirment par exemple la nécessité d'enregistrer intégralement les formules des chartes plutôt que de les caractériser à l'aide d'un descripteur, comme le proposait Bautier. Pour rappel, Fossier justifie ce choix méthodologique par le fait que cette méthode autorise la poursuite d'un plus grand nombre de problématiques. Cet exemple nous montre que l'enregistrement intégral est non seulement adapté à certaines problématiques, mais qu'il permet encore d'éviter de restreindre les possibilités d'exploitation, élément fondamental, en particulier pour les instruments de travail génériques qui visent à profiter à de nombreuses spécialités de la recherche en histoire médiévale.

Pour la grande majorité des intervenants du colloque, ce type d'enregistrement constitue donc la solution idéale pour l'enregistrement, puisqu'elle permet d'enregistrer les documents sous la forme la plus exhaustive possible, en conservant un maximum d'éléments et d'informations des sources. De ce fait, elle s'adapte aussi facilement à tout type d'ensemble documentaire, peu importe leurs caractéristiques (homogénéité, répétitivité), puisqu'aucun codage n'est indispensable en amont de l'enregistrement. Par ailleurs, ce type d'enregistrement est fortement attaché à certaines problématiques discutées plus haut, et particulièrement à l'étude du vocabulaire juridique et institutionnel et à l'histoire des mentalités. L'enregistrement *full-text* permet par ailleurs la production d'instruments de travail capitaux pour la poursuite de ce type de problématiques, et en

particulier des index de mots, des index de fréquences et des concordances, qui permettent l'étude quantitative et qualitative du vocabulaire médiéval.

Cependant, cette méthode pose un problème fondamental pour ces différents collectifs : le temps et les moyens que nécessite l'enregistrement intégral des textes. Cette question est notamment traitée dans deux articles publiés dans la première sous-partie de la troisième partie des actes du colloque, intitulée « Traitement intégral ou indexation ? ». Pour exposer ce problème de manière concrète aux participants du colloque, le médiéviste italien Silio Pietro Paolo Scalfati (1942-), qui travaille au sein du collectif de Violante sur des actes notariés pisans, propose dans son article le résultat d'une expérience visant à comparer deux essais de traitement automatique. Le premier porte sur 20 documents, enregistrés intégralement, et le second sur 108 documents, dont il n'a choisi d'enregistrer que certains éléments, comme les noms des personnes, des lieux et des organismes religieux et institutionnels, selon la méthode de l'enregistrement statistique. Le résultat de cette expérience méthodologique est clair : le deuxième procédé est « *bien plus rapide et économique* »⁵⁷⁹, et permet donc de traiter plus de documents.

Dans le second article de cette sous-partie, Elio Califano (1921-1993), président du *Gruppo di studio per l'informatica degli archivi di stato italiani*, soulève ce problème en d'autres termes, en insistant sur le fait que les dimensions des collections qu'il s'agit de rendre accessibles conditionne le choix de la méthode d'enregistrement⁵⁸⁰. Faut-il permettre aux chercheurs de faire des recherches sur une documentation plus volumineuse en disposant d'un niveau d'information (« *livello di informazione* »⁵⁸¹) moins approfondi ? Ou au contraire offrir un niveau d'information plus détaillé sur une documentation moins importante ? Car en effet les deux méthodes ne présentent pas les mêmes limites. Nous l'avons dit, l'enregistrement intégral permet aux chercheurs de bénéficier d'informations très détaillées, au plus proche des sources, mais demandent plus de temps et plus de moyens. Au contraire, l'enregistrement statistique, plus rapide et plus économique, suppose une déformation complète du contenu de la documentation, au prisme du questionnaire de recherche et de la problématique poursuivie par les historiens.

⁵⁷⁹ Silio Pietro Paolo Scalfati, « Problèmes de méthodes, à propos des actes notariés de Pise », *Publications de l'École Française de Rome* 31, n° 1 (1977): 251.

⁵⁸⁰ Elio Califano, « Registrazione diretta e integrale dei documenti », *Publications de l'École Française de Rome* 31, n° 1 (1977): 253-56.

⁵⁸¹ Califano, 254.

Les banques d'informations prosopographiques constituées par Genet et Autrand font état des limites de ce type d'instruments : difficilement réutilisables par d'autres historiens, elles ont vocation à permettre la réalisation d'une enquête donnée par un collectif de chercheur. Cet usage spécifique est renforcé, dans ces deux cas, par le fait que ces deux médiévistes se basent sur des ensembles documentaires extrêmement hétérogènes, ce qui accentue encore le fait que la cohérence de ces ensembles de données repose sur des problématiques de recherche.

Le problème apparaît cependant similaire pour les projets utilisant des méthodes d'enregistrement statistique sur la base d'une documentation homogène. L'exemple de l'enquête du *catasto*, étudiée au chapitre 1, constitue à ce titre un autre exemple des limites de la réutilisation de ces instruments en dehors de leur contexte d'élaboration. Pour rappel, la grande majorité des réutilisations de l'« édition » du *catasto* avait pris la forme d'interrogations ponctuelles, plutôt que d'enquêtes reprenant le jeu de données élaboré.

2.3.2. Une solution intermédiaire : l'indexation documentaire.

Un troisième type de méthodes d'enregistrement des données est encore présenté au colloque de Rome : l'indexation documentaire, aussi appelée par les acteurs du colloque « indexage », ou « analyse de contenu ». Cette solution, notamment évoquée par Bautier, constitue une forme de solution intermédiaire entre l'enregistrement statistique et l'enregistrement intégral. Une section de la troisième partie des actes du colloque, intitulée « Analyse de contenu » y est consacrée. Dans le premier article, intitulée « Les langages documentaires », Fossier présente la nécessité de cette méthode en ces termes :

« On a vu au cours des séances précédentes quel était pour la recherche historique l'intérêt d'un traitement de texte intégral, et combien l'établissement d'index de fréquence et de concordances pouvait faire progresser notre connaissance du vocabulaire et des institutions qu'il recouvre. Mais on en aperçoit aussi bien évidemment les limites quand il s'agit de l'exploitation d'un ensemble considérable et hétérogène de documents historiques qu'il ne peut être question d'enregistrer in-extenso en machine. [...] Reste donc à exploiter l'énorme masse des documents non dépouillés. La seule solution, pour l'instant,

si l'on tient à travailler sur le texte-même et la langue-même du document, est de le limiter à l'essentiel et de n'enregistrer en machine que des extractions de textes, termes significatifs ou ensemble articulé de termes. »⁵⁸²

Le lien fait ici par Fossier entre le caractère hétérogène de la documentation, les dimensions du corpus et le type de traitement informatique nécessaire nous renseigne sur la spécificité de cette troisième méthode d'enregistrement des données.

L'analyse de contenu peut être pratiquée de plusieurs manières, que Fossier détaille dans son article. Cette méthode peut tout d'abord se baser sur un enregistrement *full-text* des documents. Mais plutôt que d'offrir au chercheur la possibilité de chercher dans le corpus des formes des mots, ou des lemmes, comme nous l'avons vu pour les projets de Genet ou Parisse, cette méthode distingue les termes des documents des concepts qu'ils représentent. Pour permettre l'accès à l'ensemble des termes qui renvoie à un concept, cette méthode propose d'indexer les termes du corpus à l'aide d'un langage documentaire, c'est-à-dire, d'un ensemble organisé de descripteurs, ou mot-clefs. La méthode peut aussi directement procéder d'un enregistrement de ces descripteurs, sans enregistrement intégral.

Dans les deux cas, l'instrument de travail doit donc permettre aux utilisateurs de lancer une recherche sur la base d'un thésaurus détaillant les logiques de ce langage documentaire, dans le but de trouver dans les documents un ensemble de termes qui s'y rapporte. Pour illustrer son propos, Fossier prend pour exemple le descripteur « moyen de transport », qui permettra de retrouver les mentions de termes comme bateau, véhicule, bête de somme, chariot, etc.

Cette méthode d'indexation s'inscrit donc dans une voie médiane entre enregistrement *full-text* et traitement statistique. Elle doit permettre de créer des instruments de travail pour les historiens intéressés par le vocabulaire comme par le contenu, tout en contournant les limites matérielles posées par l'enregistrement intégral. Pour ce faire, la principale difficulté réside dans la nécessité d'établir le langage documentaire, qui doit non seulement s'adapter 1) aux termes contenus dans les sources, mais aussi 2) aux problématiques qui intéressent les historiens. Cette opération

⁵⁸² Lucie Fossier, « Les langages documentaires », *Publications de l'École Française de Rome* 31, n° 1 (1977): 351.

d'indexation est ainsi au cœur de cette méthodologie, et suppose à la fois, une connaissance sommaire mais exhaustive des documents du corpus, et des échanges avec les futurs utilisateurs pour établir les thésaurus.

Ce dernier point relève directement des questions qui animent ce chapitre, puisque ces thésaurus constituent des outils de travail qui permettraient d'harmoniser les méthodes d'indexation documentaire. Leur circulation d'un collectif à l'autre pourrait encore éviter à certains un travail long et pénible et générer des économies en temps comme en subsides. Dans l'article de Bautier précédemment évoqué, « Les demandes de l'historien à l'informatique », ce dernier propose par exemple d'appliquer le langage documentaire développé par l'historienne du droit Metman aux contenus des chartes :

« Je pense que le thesaurus à la préparation duquel elle a donné tous ses soins pourrait servir de base pour répondre à toutes les demandes que les historiens du droit, des institutions et de l'économie pourraient adresser aux chartes. »⁵⁸³

On découvre ici, en rapport avec ces méthodes d'indexation documentaire, un nouveau vecteur potentiel pour l'organisation méta-collective entre les médiévistes : l'échange de thésaurus documentaire.

Ces quelques exemples montrent que le choix de la méthode d'enregistrement repose sur des critères intellectuels, comme les caractéristiques des genres d'ensembles ou les problématiques visées par les corpus, mais que ce sont aussi des critères matériels, de coût et de temps, qui conditionnent le choix des méthodes par les acteurs. Les différentes configurations dans lesquelles se trouvent les collectifs de recherche les poussent donc à rechercher la meilleure solution pour poursuivre leurs travaux de recherches. Pour ce faire, les collectifs bien avancés dans l'élaboration de leurs instruments de travail font figure de modèles autant que de ressources pratiques pour les collectifs qui s'engagent dans ce type d'entreprise.

À cette logique collaborative, qui se développe au fil des échanges, se superpose une autre dynamique, qui vise à coordonner les manières de travailler au contact des ordinateurs. Elle prend cependant des formes différentes en fonction des configurations de travail dans lesquelles s'inscrivent ces collectifs. Comme nous l'avons vu, Mollat de

⁵⁸³ Bautier, « Les demandes des historiens à l'informatique », 186.

Jourdin propose par exemple, pour les comptes portuaires, la constitution de dictionnaires des navires, qui doivent permettre de repérer le passage des navires dans différents ports dans le monde. L'échelle de l'organisation est ici mondiale, quoique restreinte aux spécialistes de l'histoire maritime.

Glénisson propose quant à lui une coordination du même type mais pour l'échange de programmes informatique de traitement de banque de données prosopographiques, qui pourraient prendre forme à plusieurs échelles, mais surtout pour les médiévistes intéressés par l'histoire sociale. Dans un autre ordre d'idée, l'échange de thesaurus s'inscrit aussi dans un type de configuration particulière, qui concerne surtout les collectifs engagés dans des entreprises d'indexation documentaire de chartes ou de documents à teneur juridique.

Le colloque de Rome est ainsi le moment d'une mise en commun de ces différentes et nombreuses idées locales qui visent à constituer des passerelles entre ces collectifs, en coordonnant ou en harmonisant les méthodes de travail assisté par ordinateur, en favorisant l'échange d'instruments de travail, par exemple, les programmes informatiques, et en faisant circuler des *documents de travail* (thesaurus, dictionnaires des navires). Au carrefour de ces questions se pose le problème de la réalisation de ces idées, qui doit passer pour les intervenants par l'établissement d'un cadre général à une nouvelle échelle.

3. Deux propositions d'organisation méta-collective à l'échelle européenne.

Nous avons analysé plus haut les problématiques communes auxquelles font face les collectifs présents à Rome du fait de leur volonté d'utiliser des moyens électroniques pour traiter de grands ensembles de documents historiques. Nous avons encore aperçu pourquoi il était nécessaire pour eux d'échanger et de débattre à propos des solutions choisies par certains de ces collectifs plus avancés dans leurs entreprises. Nous avons encore pu souligner, au fil de l'analyse, l'existence de propositions visant à organiser, harmoniser, coordonner les usages des moyens électroniques. Au-delà de ces propositions locales, quelques acteurs insistent à Rome sur la nécessité de mettre en place des cadres généraux d'information et de coordination pour l'ensemble des utilisateurs d'ordinateurs en histoire médiévale.

L'une de ces propositions est formulée en conclusion du colloque par Genicot, qui réitère dans ce texte sa proposition de 1973 : mettre en place un organisme qui aurait vocation à informer et à coordonner les activités de ces collectifs. Quelques mois seulement après le colloque de Rome, il publie un nouvel article dans la revue *Francia*, dans lequel il propose en annexe un « modèle de description de toute réalisation achevée, en cours ou envisagée »⁵⁸⁴. Ce document, sur lequel nous reviendrons, a vocation à circuler au sein de la discipline dans le but de permettre un recensement normalisé des projets utilisant des ordinateurs. Les trois articles mentionnés ici nous permettront d'examiner les modalités de cette proposition et la logique de l'organisation méta-collective imaginée par Genicot dans ce contexte.

Simultanément, une seconde résolution du même type se met en place à l'initiative de Fossier et du service informatique de l'IRHT, auxquels les participants du colloque ont proposé de confier « la centralisation et la diffusion des informations sur le traitement électronique des sources médiévales »⁵⁸⁵. Dans ce cadre, un premier groupe composé de cinq chercheurs élabore, dès 1975, un projet d'enquête qui vise à recenser les collectifs utilisant des ordinateurs en Europe. Les origines et les aboutissements de cette proposition nous sont connus par l'intermédiaire de trois textes de Fossier : 1) un court article de 1975, dans lequel elle présente un bilan des décisions prises à l'issue du colloque de Rome⁵⁸⁶ ; 2) un article publié en 1976, lui aussi dans *Francia*, en réponse à l'article de Genicot de 1975 précédemment cité⁵⁸⁷ et 3) l'éditorial du numéro bilan des dix premières années du *Médiéviste et l'ordinateur* publié en 1989⁵⁸⁸. En plus de ces trois textes, nous nous appuierons sur une boîte d'archives conservée au centre de l'IRHT à Orléans, qui contient des documents qui nous renseignent sur les résultats de l'enquête menée par Fossier et la section informatique de l'IRHT entre 1977 et 1979⁵⁸⁹.

⁵⁸⁴ Léopold Genicot, « Pour une organisation de la recherche en histoire médiévale, II. L'information sur les recours aux ordinateurs », *Francia* 3 (1975): 644.

⁵⁸⁵ Genicot, 644.

⁵⁸⁶ Lucie Fossier, « L'utilisation de l'informatique pour l'exploitation des documents textuels médiévaux », *Mélanges de l'école française de Rome* 87, n° 2 (1975): 617-18.

⁵⁸⁷ Lucie Fossier, « A propos d'un formulaire d'enquête relatif à "L'information sur les recours aux ordinateurs" », *Francia* 4 (1976): 719-21.

⁵⁸⁸ Lucie Fossier, « Éditorial », *Le Médiéviste et l'ordinateur* 20, n° 1 (1988): 1-3.

⁵⁸⁹ J'aimerais ici remercier Caroline Bourlet et toute l'équipe du centre de l'IRHT à Orléans pour leur accueil pendant les quelques séjours que j'ai passés en ces lieux.

3.1. Le questionnaire du groupe « Ordinateurs et Recherches Médiévales ».

Dans la conclusion du colloque de 1975, Genicot soulève plusieurs problèmes communs abordés par les intervenants au cours de ces trois jours de discussions. Le premier qui nous intéressera ici concerne la finalité des instruments de travail que ces méthodes permettent d'élaborer. Il le formule en ces termes :

« va-t-on mettre l'ordinateur au service de la recherche « ouverte » ou de recherches « fermées » ? L'affecter, en d'autres termes, à l'élaboration d'instruments de travail ou à la solution de questions limitées ? »⁵⁹⁰.

Cette question, qui comme nous l'avons montré est au cœur des discussions du colloque, rejoint la distinction que nous avons établie entre instrument de travail générique et instrument de travail spécifique. Elle rejoint également la problématique du choix de la méthode d'enregistrement.

3.1.1. Finalité des instruments de travail et modalité d'enregistrement des données.

Dans la conclusion du colloque, Genicot classe chacune de ces méthodes sur un axe qui va de « recherche « ouverte » » à « recherches « fermées » ». Le traitement *full-text* est à une extrémité de cet axe, puisqu'il permet de :

« préparer et faciliter toutes les recherches indistinctement [...]. Y compris celles auxquelles nous ne pensons pas aujourd'hui et que les progrès de la problématique susciteront un jour »⁵⁹¹.

On retrouve l'indexation documentaire au milieu de cet axe, puisque tout en cherchant à produire des instruments de travail générique, cette méthode d'enregistrement limite les réutilisations possibles. La méthode d'enregistrement statistique est quant à elle

⁵⁹⁰ « Conclusion générale », *Publications de l'École Française de Rome* 31, n° 1 (1977): 425.

⁵⁹¹ « Conclusion générale », 425.

considérée comme une forme de recherche fermée, puisqu'une fois résolu le problème pour lequel cet instrument a été conçu, « *la bande ou le disque enregistré n'offre plus d'intérêt* »⁵⁹².

Comme nous l'avons vu, l'enregistrement *full-text* pose cependant des difficultés matérielles (temps, coût) qui limitent son utilisation systématique. Pour surmonter ce problème, Genicot propose plusieurs solutions à ses collègues. Tout d'abord, conjuguer les méthodes, en enregistrant par exemple intégralement certains échantillons des collections d'archives tout en utilisant une méthode d'indexation pour le reste de la collection. Ensuite, suivre certains « *impératifs* »⁵⁹³ qui doivent permettre, peu importe la méthode retenue, de constituer des instruments de travail réutilisables par d'autres chercheurs : 1) « *opérer sur un ensemble ou sur une tranche d'un ensemble, aux arêtes nettes* »⁵⁹⁴ ; 2) « *s'inspirer au maximum de l'emploi que les érudits feront aujourd'hui et demain des bandes et disques, c'est-à-dire pratiquement des questions qu'ils leur poseront* »⁵⁹⁵ ; 3) « *avertir l'utilisateur du procédé suivi et des limites qui en découlent* »⁵⁹⁶ et 4) « *réduire au minimum toute manipulation, intervention ou interprétation préalable à l'enregistrement* »⁵⁹⁷.

Comme on peut le remarquer à la lumière de ces quelques règles, ce souci d'harmoniser les méthodes d'enregistrement se conjugue avec la question de la mise à disposition de ces instruments de travail. Dans la conclusion des actes, Genicot insiste justement sur ce problème en ces termes :

*« Faut-il dresser des concordances, de divers types, et les distribuer sous forme de livre, ainsi que l'a fait le Centre de traitement électronique des documents, le Cetedoc de Louvain pour Latran IV ou Vatican II ? Ou en rester au listing sorti de l'ordinateur et offrir à qui le désire d'en prendre copie à ses frais ? Ou même se borner à reporter texte ou analyse sur bandes et disques et rendre ceux-ci accessibles aux intéressés ? »*⁵⁹⁸

⁵⁹² « Conclusion générale », 426.

⁵⁹³ « Conclusion générale », 428.

⁵⁹⁴ « Conclusion générale », 428.

⁵⁹⁵ « Conclusion générale », 428.

⁵⁹⁶ « Conclusion générale », 428.

⁵⁹⁷ « Conclusion générale », 429.

⁵⁹⁸ « Conclusion générale », 430.

Ces différents modes de publication des instruments de travail sont eux aussi au centre des discussions sur le partage et la mise à disposition des produits de l'utilisation de l'informatique. Si Genicot considère que la publication papier est « *la meilleure* » solution, il souligne cependant que son coût élevé empêche en pratique son application systématique. En l'absence de crédits, il peut ainsi être suffisant, selon lui, de « *faire savoir qu'un ensemble a été entré en machine suivant telle modalité et est consultable dans telles conditions* »⁵⁹⁹.

Cette dernière proposition de Genicot dessine les contours d'une nouvelle forme de publications des instruments de travail en histoire médiévale, qui, ne pouvant passer par le médium de l'édition académique traditionnelle (le livre), se doit d'être communiquée sous d'autres formes. C'est justement à ce type de questions que renvoie la dernière partie de la conclusion de Genicot, dans laquelle il dégage un véritable programme d'organisation pour les collectifs impliqués dans l'utilisation des ordinateurs en histoire médiévale :

*« Il faudrait d'abord décrire attentivement les genres d'ensembles [...] dans leurs caractères externes et surtout dans leur contenu et leur texture [...] pour décider de la manière de les traiter, juger de la nécessité et de la possibilité de les enregistrer ou non [...]. On ferait ensuite un second pas : élaborer de façon précise, avec le concours de techniciens, le mode de traitement, le « programme », le mieux adapté à chaque genre. Pour obtenir ou permettre que toute entreprise se coule désormais dans un moule uniforme et approprié. Il s'agirait enfin de faire connaître les projets achevés, en cours ou envisagés. Et cela sous une forme identique, claire, complète ; beaucoup des informations aujourd'hui diffusées sont inutilisables parce qu'elles manquent de rigueur ; certaines omettent même de dire le type d'ordinateur employé »*⁶⁰⁰.

On remarque que le programme de Genicot reprend certains des grands axes de discussions du colloque que nous avons analysé dans la partie 2 de ce chapitre :

⁵⁹⁹ « Conclusion générale », 430.

⁶⁰⁰ « Conclusion générale », 430.

description des genres d'ensembles documentaires, définition des méthodes d'enregistrement les plus adaptées, discussion sur les programmes informatiques applicables à chaque ensemble. Cependant, la dernière partie de ce passage propose une première définition des exigences à suivre pour permettre le partage de ces nouveaux instruments de travail de façon « *identique, claire, complète* », ce qui passe par la définition d'une nouvelle forme de publication.

3.1.2. Un « *modèle de description* » pour les projets mobilisant des ordinateurs.

En 1975, peu de temps après le colloque de Rome, Genicot publie dans *Francia* le deuxième volet de son article « Pour une réorganisation de la recherche en histoire médiévale », dans lequel il propose en annexe un « *modèle de description* »⁶⁰¹ qui avait pour objet de standardiser les éléments de présentation des réalisations des collectifs utilisant des ordinateurs en histoire médiévale. Ce document est adapté de la proposition faite par Tombeur dans le cadre de la commission ORM créée à Madrid en 1972, dont il a déjà été question en introduction. Il comporte deux parties, l'une pour les projets en cours ou déjà achevés, l'autre pour les projets en cours de définition. Dans son article, Genicot précise que ce modèle

« ne porte pas seulement sur les aspects techniques de l'entreprise mais aussi et davantage sur les objectifs, son contenu, ses limites scientifiques »⁶⁰².

Ces informations doivent permettre, comme nous l'avons dit, de rendre ces instruments de travail exploitables par d'autres collectifs. Nous nous intéressons ici uniquement à la première partie de ce « *modèle de description* », qui est bien plus développée que la seconde. Ce document invite les chercheurs à présenter leurs travaux en deux temps : par une description générale puis par une description technique.

La partie description générale est composée de cinq sous-parties, destinées à renseigner sur les différents aspects du projet. La première porte sur la « *documentation*

⁶⁰¹ Genicot, « Pour une organisation de la recherche en histoire médiévale, II. L'information sur les recours aux ordinateurs », 644.

⁶⁰² Genicot, 644.

étudiée », et vise à éclairer certaines caractéristiques du corpus (langue, époque), mais aussi les buts poursuivis et l'état d'avancement des travaux. À cela s'ajoute encore un élément fondamental pour permettre la mise à disposition des instruments de travail : l'adresse de l'institution d'accueil et les noms des chercheurs impliqués dans ces projets. La deuxième sous-partie de cette description générale invite les chercheurs à détailler la méthode d'enregistrement choisie, en insistant pour chacune d'entre elle sur la nécessité de décrire les modalités de cet enregistrement - en précisant par exemple la liste des descripteurs pour l'indexation. La troisième sous-partie concerne les méthodes de référencement mise en œuvre par les collectifs, c'est-à-dire les procédés mis en place pour localiser un objet (document, noms, etc.) dans l'ensemble de données enregistrées. La quatrième sous-partie vise à renseigner sur le type d'analyse des documents mis en œuvre par ces collectifs : linguistique, lexicographique, sociologique, philosophique, etc. La cinquième et dernière sous-partie de cette description générale invite quant à elle les chercheurs à communiquer leur bibliographie, ainsi que la documentation préexistante accompagnant ces projets.

La partie « description technique » comporte quant à elle quatre sous-parties. La première est destinée à renseigner les futurs utilisateurs sur le matériel employé pour les opérations d'enregistrement et de traitement. Quel est le type d'ordinateurs utilisé ? Quelle est la capacité de la mémoire centrale ? Ces renseignements portent aussi sur la partie *software*, avec au cœur de la question la disponibilité des programmes. Quel est le langage employé ? En combinaison avec quel système d'exploitation ? La deuxième sous-partie porte plus précisément sur les techniques d'entrées des données. Quels sont les supports de stockage employés ? Les codes utilisés (alphanumérique, numérique, caractères spéciaux) ? La troisième sous-partie incite les chercheurs à décrire succinctement les fonctions des programmes informatiques élaborés ou mis en œuvre dans le cadre de ces entreprises. C'est cependant la dernière sous-partie qui doit permettre de donner les informations sur la matérialité de la sortie des données. Quels sont les supports employés (cartes, bandes magnétiques, disques magnétiques, listings, etc.) ? Les acteurs interrogés doivent encore préciser si ces documents de sortie sont disponibles pour d'autres collectifs, et si oui, sous quelles formes.

Cette longue liste d'informations reflète les éléments indispensables dégagés par Genicot et la commission ORM pour améliorer la circulation des instruments de travail

informatique. Nous remarquons que ces informations concernent aussi bien des questions intellectuelles, qui portent sur la finalité des instruments de travail, le type de sources ou de collections concernées, ainsi que des questions éminemment matérielles, comme le type d'ordinateurs utilisé ou le mode de stockage des données, qui conditionnent, comme nous l'avons vu dans les deux premiers chapitres, la portabilité de ces instruments de travail d'un centre de calcul à un autre.

Ce modèle nous renseigne encore sur l'invention d'un mode de publication original qui constitue, à notre connaissance, une véritable nouveauté dans le champ de l'histoire médiévale. À défaut de pouvoir imprimer sur papier et publier sous forme de livre ces index, ces dictionnaires, ces bases de données, ... les acteurs proposent la mise en œuvre d'un nouveau type de document descriptif, qui doit permettre une circulation plus efficace des informations, et donc, une réutilisation plus large de ces outils. L'ambition de Genicot dans ce cadre n'est pas seulement de promouvoir ou de favoriser les travaux utilisant des ordinateurs en histoire médiévale, mais bien de faire progresser la discipline toute entière. La coordination des activités des collectifs utilisant des ordinateurs n'est selon lui qu'une opportunité, qu'un vecteur qui doit permettre la réorganisation des relations de travail entre les médiévistes à une autre échelle. Dans la conclusion de son article de 1973, il affirme :

« Dans toutes les sciences, l'organisation et la collaboration sont aujourd'hui une des conditions majeures du progrès. L'histoire n'échappe pas à cette loi »⁶⁰³.

3.2. L'initiative de Lucie Fossier et du service informatique de l'IRHT.

Nous l'avons dit, l'une des décisions prises au colloque de Rome par les participants est de confier la coordination et la diffusion des informations relatives à l'utilisation des ordinateurs en histoire médiévale à la section informatique de l'IRHT, dirigée par Fossier. Dans un article de 1975, Fossier nous renseigne sur les raisons qui fondent le choix de cette institution. Tout d'abord, elle possède une infrastructure dédiée à l'informatique. Ensuite, elle est déjà engagée dans de nombreux projets impliquant

⁶⁰³ Genicot, « Pour une organisation de la recherche en histoire médiévale », 697.

l'utilisation des moyens informatiques pour l'exploitation de grandes séries de documents en France. Enfin, on peut encore ajouter que Fossier est la correspondante française de la newsletter *CAMDAP*, dont il a été question en introduction. Dans ce cadre, elle est chargée depuis 1972, de faire le lien entre les éditeurs de ce périodique et les chercheurs utilisant les ordinateurs en Europe⁶⁰⁴. Ces trois raisons font de Fossier et de son équipe à l'IRHT un collectif tout désigné pour coordonner la centralisation et la diffusion des informations relatives aux utilisations des ordinateurs.

3.2.1. Une autre enquête et un autre questionnaire.

À la suite du colloque, la section informatique de l'IRHT entreprend ainsi de lancer une enquête à l'échelle internationale qui doit permettre non seulement de 1) « recenser au moyen d'un questionnaire »⁶⁰⁵ les projets existants, mais aussi 2) d'« établir une coopération [...] entre des laboratoires spécialisés [...] afin de procéder à des échanges de programmes »⁶⁰⁶ et 3) de « susciter [...] la création de banques de données »⁶⁰⁷. Dès 1975, ce sont donc deux modèles d'enquêtes qui se développent à Paris et à Louvain. Dans cette situation, le « modèle de description » publié par Genicot en 1975 se présente comme une main tendue vers la réalisation commune d'un questionnaire.

En 1976, Fossier répond à la proposition de Genicot dans un article publié dans *Francia*, intitulé « À propos d'un formulaire d'enquête relatif à « L'information sur les recours aux ordinateurs »⁶⁰⁸. Dans ce texte, elle insiste sur le fait que ce plan-type ne prend pas assez en compte les problématiques qui se posent pour le traitement des documents médiévaux. Elle suggère plusieurs modifications de ce modèle, basées sur les réflexions du projet d'enquête entrepris depuis l'issue du colloque de Rome. Fossier propose par exemple d'ajouter une distinction entre source primaire et source secondaire, puisque ces deux types de documentation n'impliquent pas le même type de difficultés

⁶⁰⁴ Voir René Pellen, « « CAMDAP » (« Computer and Medieval Data Processing »), revue s'intéressant à tout ce qui concerne le traitement par ordinateur des documents médiévaux. Vol. I, n° 1, février 1971 à vol. VIII, n° 1, 1978. », *Cahiers de Civilisation Médiévale* 22, n° 86 (1979): 215-16.

⁶⁰⁵ Fossier, « L'utilisation de l'informatique pour l'exploitation des documents textuels médiévaux », 618.

⁶⁰⁶ Fossier, 618.

⁶⁰⁷ Fossier, 618.

⁶⁰⁸ Fossier, « A propos d'un formulaire d'enquête relatif à "L'information sur les recours aux ordinateurs" ».

vis-à-vis de la coordination entre les collectifs. En effet, si Genicot soulignait dans son article de 1973 le problème des « double-emploi », qui consisterait en l'enregistrement de corpus identiques par des équipes différentes, Fossier explique dans son article que ce problème ne se pose pas pour

*« les documents d'archives, qui, du fait de leur unicité, de leur caractère national, voire local, et le plus souvent de leur non publication, risquent peu d'être simultanément exploités en plusieurs endroits. »*⁶⁰⁹.

Fossier propose encore dans son article de rééquilibrer les parties du plan-type de Genicot qui portent sur les méthodes d'enregistrement. Elle avance notamment la nécessité de prévoir une section permettant aux interrogés de détailler les éléments qu'ils ont choisis d'extraire des sources, en plus de la section prévue pour indiquer les méthodes d'enregistrement. Cette modification vise à distinguer les éléments enregistrés de la méthode d'enregistrement, puisque, comme nous l'avons vu pour les formules des chartes par exemple, seuls certains des éléments des documents peuvent être enregistrés intégralement.

Fossier suggère aussi de rééquilibrer la section sur les méthodes d'analyses, qui oriente presque exclusivement vers des possibilités liées au traitement de corpus d'œuvres littéraires. Dans le même ordre d'idée, elle invite à étendre la liste des documents de sortie mentionnés par Genicot, dont la proposition se concentre avant tout sur les aspects matériels, et notamment le support de stockage. Fossier propose au contraire d'inclure le type de produit de sortie (index, concordance) mais aussi les autres formes d'utilisation possible de ces instruments de travail : recherche rétrospective ou exploitation en temps réel.

En plus de ces précisions, qui vont dans le sens d'une extension de ce plan-type à l'ensemble des projets informatiques, Fossier remet aussi en cause la finalité de ce document, notamment pour la question de la coordination des méthodes d'analyses des sources dites « documentaires ». En effet, selon elle,

⁶⁰⁹ Fossier, 721.

*« la diversité des traitements à leur appliquer [aux documents d'archives] rend difficile des échanges de programmes et de codages. »*⁶¹⁰

Elle conclut alors :

*« Le traitement informatique, pour les historiens médiévistes, à cause de l'extrême diversité du matériau concerné, restera toujours affaire individuelle »*⁶¹¹.

Cette considération attachée aux traitements informatiques des corpus de source conduit Fossier à redéfinir la finalité de la coordination pour ce type d'entreprise. Pour elle, un cadre général doit bien être mis en place, mais

*« moins pour susciter une entente entre laboratoires déjà constitués que pour tenter de venir en aide aux hésitants, aux isolés dépourvus de ressources, pour établir des contacts que l'on voudrait fructueux entre individus »*⁶¹².

En ce sens, elle propose donc une vision différente de l'organisation méta-collective telle qu'elle est défendue par Genicot, pour qui l'intérêt d'un cadre général vise plutôt, nous l'avons vu, à favoriser la circulation des instruments de travail entre les collectifs utilisant déjà des ordinateurs, et à harmoniser les cultures de travail locales développées par ces collectifs. Au contraire, Fossier propose un modèle d'organisation plus horizontal, qui offrirait aux chercheurs désireux de s'engager dans cette voie : les informations, les ressources et même l'aide leur permettant de le faire. Pour les collectifs déjà engagés dans ce type de projet, cette organisation favoriserait par ailleurs la mise en relation entre individus partageant des problèmes communs.

Dans l'année qui suit, l'enquête lancée par la section informatique de l'IRHT donne d'ailleurs ses premiers résultats, et confirme la pertinence du modèle imaginé par Fossier. Sur la base des suggestions formulées dans l'article de *Francia*, un questionnaire avait en effet été mis en circulation par un premier embryon de secrétariat, constitué

⁶¹⁰ Fossier, 721.

⁶¹¹ Fossier, 721.

⁶¹² Fossier, 721.

autour de Fossier, et composé de cinq chercheurs parisiens présents à Rome⁶¹³. Cette initiative vise à recenser précisément les projets en cours en France et en Italie. Malheureusement, l'enquête n'aboutit qu'à de maigres résultats : sur les vingt et une réponses qu'ils reçoivent, seulement quatre d'entre elles font état d'un travail terminé, neuf présentent des études qui ne sont pas encore achevées et les huit restantes ne sont pas complètes⁶¹⁴.

3.2.2. Bilan de l'enquête de l'IRHT et changement de stratégie.

Les archives relatives à cette enquête, conservées au centre de l'IRHT à Orléans, nous permettent de saisir précisément la provenance de ces réponses. On remarque notamment que la plupart des réponses complètes sont le fait de chercheurs déjà très proches des cinq membres du secrétariat. C'est par exemple le cas de l'équipe de recherche diplomatique du CRAL, que Fossier avait créée au début des années 1970, ou encore de certains collègues de Genet au sein de l'ERA 713. Le questionnaire n'a donc visiblement que trop peu touché des collectifs extérieurs au réseau qui s'était constitué en amont du colloque de Rome.

Ces archives nous renseignent aussi sur l'existence de chercheurs plus isolés, qui ne sont, pour certains, pas en mesure de répondre de manière complète au questionnaire de l'enquête. Un cas en particulier montre notamment le type de difficultés que rencontrent ces chercheurs. En 1977, Georges Jehel prépare un doctorat de troisième cycle qui vise à exploiter des actes notariés conservés aux archives d'État de Gênes. Pour ce faire, il ambitionne d'utiliser des moyens informatiques, mais il se confronte à des difficultés pratiques. Sa réponse au questionnaire s'apparente plus un appel à l'aide qu'à un état des lieux de sa recherche en cours. Dans la rubrique « observation » du questionnaire, il indique :

⁶¹³ Les membres de ce premier groupe nous sont, pour certains d'entre eux, désormais bien connus comme Klapisch-Zuber et Genet, ou bien ont été présentés rapidement dans ce chapitre, comme Metman. Les deux autres membres de ce groupe n'ont été mentionnés que trop brièvement, en dépit de leur participation au colloque de Rome. Il s'agit de Jacques Lefort, médiéviste spécialiste de l'empire byzantin, qui travaille au sein de la IV^e section de l'École Pratique des Hautes Études, et de Gian Piero Zarri, jeune informaticien, co-auteur avec Fossier d'un manuel sur l'indexation documentaire précédemment cité, et qui prépare en 1979 une thèse visant à appliquer les techniques de l'intelligence artificielle au traitement de données biographiques.

⁶¹⁴ Le résultat de cette enquête nous est connu grâce aux archives que nous avons conservées au centre de l'IRHT à Orléans.

« Il m'a été impossible de bénéficier de quelque aide que ce soit. En particulier, une démarche auprès du centre de recherche en sciences humaines du boulevard Raspail est restée sans réponse et sans suite. J'ai entrepris mes dépouillements à partir de vagues renseignements et j'ai relevé à ce jour environ 1 500 articles. Je considère l'analyse systématique de l'ensemble des minutiers à laquelle je suis disposé à collaborer dans une perspective exhaustive comme présentant un intérêt décisif. ».

Jehel note ici deux types de difficultés auxquelles il fait face dans son entreprise. La première concerne l'accès aux ordinateurs et aux spécialistes de l'informatique. Le centre de recherche qu'il mentionne ici est le Laboratoire d'Informatique pour les Sciences de l'Homme (LISH) du CNRS, qui avait été créé en 1975 afin d'offrir aux chercheurs en sciences humaines du matériel (ordinateurs et périphériques) et des spécialistes de l'informatique pour assister ce type de projet. Le LISH ne semblait cependant pouvoir répondre aux nombreuses sollicitations dans les premières années de sa création, certainement parce qu'il était très fréquenté, et particulièrement prisé par les chercheurs de l'EHESS⁶¹⁵. La seconde difficulté est attachée au manque de renseignements disponibles sur les méthodes de l'informatique, qui en dépit des nombreuses publications (newsletter, bulletin, actes de colloque etc...), laissait à chaque chercheur le soin de constituer lui-même sa bibliographie.

Confrontée au désarroi de Jehel, Fossier lui propose donc son aide. Durant les deux années qui suivent, elle s'occupera notamment de corriger le code qu'il a établi pour l'enregistrement des données des actes notariés, lui conseillera des programmes permettant de mettre en place les analyses statistiques qu'il désire mener et organisera pour son compte des rendez-vous avec d'autres historiens coutumiers du traitement informatique des actes notariés.

Ce type d'interaction, qui naît de cette enquête, conduira finalement les membres du premier secrétariat de coordination à réévaluer la fonction d'une coordination et d'une circulation des informations au sein de la discipline. C'est justement de ce constat que va

⁶¹⁵ Ces informations nous ont été données par l'un des anciens ingénieurs en charge de l'assistance informatique au LISH, Philippe Cibois, dans un entretien que nous avons mené en 2018.

naître, sous l'égide de ce groupe de chercheurs, le bulletin de liaison *Le Médiéviste et l'ordinateur* en 1979. Dans le vingtième numéro, publié en 1989, Fossier revient justement sur l'importance de cette enquête dans la définition du projet du bulletin, en la résumant en ces termes :

*« Ce fut un fiasco. Tout à fait explicable, car outre l'aspect fastidieux et parfois traumatisant d'un questionnaire, bien des médiévistes ne pouvaient être à même de formuler des réponses précises. En fait, c'était l'inverse qu'il, leur fallait : non pas donner, mais recevoir l'information. »*⁶¹⁶

C'est de ce constat que naîtra le bulletin de liaison *Le Médiéviste et l'ordinateur*, avec pour objectif d'aller

*« au-devant du public des médiévistes, en leur offrant, [...], une information qui les touchât de près »*⁶¹⁷.

En 1979, ce sont finalement six médiévistes et un informaticien, tous présents à Rome, qui signent l'éditorial du premier numéro, posant la première pierre d'un réseau d'information qui sera au centre du développement des méthodologies informatiques dans la discipline en Europe.

Conclusion.

Le changement d'échelle d'analyse que nous avons opéré dans ce troisième chapitre par rapport aux deux précédents nous a permis de dégager plus largement la diversité des manières par lesquelles les médiévistes s'étaient emparés des moyens de calculs électroniques à partir du début des années 1970, et d'observer une nouvelle série de conséquences de ce phénomène sur les organisations collectives et les pratiques des sources des médiévistes.

⁶¹⁶ Fossier, « Éditorial », 2.

⁶¹⁷ Fossier, 2.

Quant aux questions liées aux organisations collectives, ce chapitre nous a permis de repérer certaines des logiques de rassemblement qui avaient conduit les acteurs à formuler, dans ce contexte, des initiatives en direction d'une organisation méta-collective de la discipline. À ce titre, nous avons mis en lumière les spécificités des deux initiatives visant à une organisation méta-collective des activités des collectifs utilisant des ordinateurs en histoire médiévale. Nous avons montré que le programme proposé en 1973 par Genicot et l'initiative lancée par Fossier se distinguaient sur plusieurs points : 1) les types de corpus considérés étaient différents (grandes collections de documents *versus* œuvres littéraires), tout comme ; 2) les objets de ces recherches (« *contenu* » *versus* « *langue* ») et 3) les méthodes d'enregistrement utilisées (indexation documentaire *versus* enregistrement *full-text*). Toutefois, ces deux initiatives, qui visaient à rassembler, *a priori*, des collectifs aux usages des ordinateurs différentes, engagés dans des enquêtes poursuivant des problématiques distinctes, conduisirent à des mobilisations plus larges que celles imaginées par ces chercheurs initialement.

On retrouve, par exemple, dans les signataires de l'article de Genicot de 1973, plusieurs chercheurs impliqués dans des projets d'étude du « *contenu* » de grandes collections de documents, qui mettaient en œuvre des méthodes d'indexation de leur corpus, comme Werner ou Glénisson. Le colloque de Rome de 1975 constitue un autre exemple de cette mobilisation au-delà des enjeux initialement définis par ces deux groupes d'acteurs. En effet, si cette rencontre trouvait sa cohérence dans le fait que la grande majorité des collectifs invités étudiaient de grands ensembles de documents, nous avons vu que ce type de corpus était également approché par certains médiévistes pour traiter des problématiques liées à la « *langue* », et qu'ils pouvaient encore être l'objet d'enregistrements dans le but d'en assurer une exploitation statistique.

Dès lors, si l'on considère que les travaux des collectifs présents au colloque de Rome se rapportent, pour certains d'entre eux, aux usages des ordinateurs qualifiés par les historiens des *digital humanities* sous le terme de *text-analysis*, et pour d'autres, aux types d'usages visés par Sula et Hill dans leur article de 2019 (que certains observateurs qualifient de *data-processing*)⁶¹⁸, la rencontre de Rome apparaît comme un formidable

⁶¹⁸ Voir par exemple Lou Burnard, « Du literary and linguistic computing aux digital humanities : retour sur 40 ans de relations entre sciences humaines et informatique », in *Read/Write Book 2 : Une introduction aux humanités numériques*, éd. par Pierre Mounier, Read/Write Book (Marseille: OpenEdition Press, 2012), 45-58.

évènement pour saisir ce que partageait ces utilisateurs des ordinateurs, en dépit de la diversité de leurs approches.

Dans cette optique, le premier élément fondamental pour l'histoire des humanités numériques concerne le fait que l'ordinateur, en tant que nouvel instrument pour la discipline, devient un vecteur de rassemblement pour les médiévistes. Si l'on se base sur l'exemple du colloque de Rome, nous avons montré que cette logique de rassemblement a pour conséquence le rapprochement entre des chercheurs aux sujets variés (périodes et aires géographiques), aux problématiques historiographiques différentes (histoire sociale, histoire économique, histoire des mentalités, histoire politique) et aux approches multiples (prosopographie, onomastique, toponymie, philologie).

Nous avons montré que ce rassemblement entre des acteurs aux préoccupations scientifiques si différentes tenait avant tout à deux éléments, qui animaient les instigateurs de ces premières initiatives. Genicot et Fossier partageaient en effet 1) des convictions, tirées de leurs propres expériences, sur la nécessité d'utiliser les ordinateurs pour élaborer de nouveaux *instruments de travail* pour la recherche (index automatisés, éditions, base de données, programmes informatiques) et 2) des observations sur les difficultés matérielles (coût, temps) et techniques (options d'enregistrement des données, modes de circulation des instruments de travail) auxquelles ils se confrontaient dans la mise en œuvre de ces nouvelles méthodes.

La mise en œuvre de ces premières initiatives d'organisations collectives reposait ainsi sur des ambitions similaires, qui s'articulaient autour de l'idée que le recours aux ordinateurs devait aider les médiévistes dans l'étude de corpus de grandes dimensions, et permettre d'aller plus loin dans la poursuite de certains programmes historiographiques dominants des années 1970 (études des sociétés, histoire des mentalités, histoire sociale). Pour certains, les moyens électroniques de calculs devaient encore servir à faire progresser la critique des sources, en offrant aux médiévistes les moyens d'étudier de façon exhaustive certaines collections. Dans ces deux directions, l'ordinateur devait permettre de gagner du temps dans la réalisation d'opérations répétitives et chronophages, et faciliter ainsi aussi bien les opérations d'édition des sources que la mise en œuvre d'analyse à l'aide de méthodes statistiques. Au cœur de cette ambition portée par les acteurs du colloque de Rome se trouvait l'idée de produire et de mettre en commun de

nouveaux instruments de travail, qui auraient pour vertu d'offrir aux historiens des données à même d'être traitées quantitativement.

Derrière cette ambition se cachait cependant un autre élément qui justifiait la nécessité d'une collaboration entre ces chercheurs. Il s'agissait des difficultés techniques et matérielles qu'ils partageaient dans la réalisation de ce type de projets. Elles tenaient tout particulièrement aux limites financières d'une approche qui coûtait cher, du fait du temps nécessaire pour la mettre en œuvre, des compétences qu'elle requérait et du coût des équipements qu'il était nécessaire de mobiliser. Des difficultés techniques venaient s'ajouter à ces problèmes matériels et, en particulier, le fait que l'utilisation des cartes perforées IBM à 80 colonnes imposait une certaine forme de modélisation des sources, qui contraignaient les médiévistes à ne retenir que certains éléments des documents, au détriment d'autres, et qui limitaient ainsi les possibles utilisations et réutilisations qui pouvaient être faites de ces nouveaux types d'instruments de travail.

Pour dépasser cet état de fait, il leur apparaissait nécessaire de confronter leurs méthodes et de tenter de définir des standards à même d'homogénéiser les pratiques d'enregistrement des sources. La question générale que posaient les acteurs était alors de savoir comment enregistrer les sources de la façon la plus universelle possible, à la lumière des problématiques qui animaient les études des médiévistes. Cette interrogation était, de fait, l'enjeu principal du colloque de Rome, et se corrélait à des difficultés que nous avons abordées tout au long de ce chapitre.

L'analyse que nous avons menée dans ce chapitre révèle par ailleurs un autre ensemble de conséquences de l'utilisation des ordinateurs à une échelle plus large que celle que nous avons abordée dans la première partie de cette thèse. Au colloque de Rome, nous avons notamment mis en évidence qu'avaient pris forme plusieurs ensembles de discussions qui concernaient ce que nous avons qualifié dans les deux premiers chapitres de *pratiques des sources*.

L'un des axes de discussion sur ce sujet avait trait aux règles de l'édition des sources et, plus généralement, aux règles de la production d'archives, dont les évolutions étaient déjà présentées par Furet dans son article de 1971 :

« Toute la conception même de l'archivistique s'en trouve radicalement transformée au moment même où ses possibilités techniques sont multipliées par le traitement électronique de l'information. Cette rencontre d'une révolution méthodologique et d'une révolution technique, qui ne sont d'ailleurs pas étrangères l'une à l'autre, permet d'envisager la constitution d'archives nouvelles, conservées sur bandes perforées, qui renvoient non seulement à un nouveau système de classement, mais surtout à une critique documentaire différente de celle du XIX^e siècle. »⁶¹⁹

L'étude des discussions du colloque de Rome permet d'affirmer que l'une des conséquences majeures de l'utilisation des ordinateurs qui prend forme, en lien avec cette « révolution méthodologique » et cette « révolution technique », concerne justement le renouvellement de ces réflexions.

En effet, la volonté affichée par de nombreux acteurs présents à Rome de réaliser des instruments de travail génériques, qui auraient vocation à être utilisés pour poursuivre de nombreux types de problématiques historiques, conduit les intervenants à engager des discussions sur les limites des logiques d'archivage de leurs sources. L'une des oppositions qui est notamment discutée est celle postulée entre sources « documentaires » et sources « littéraires », que certains intervenants, et plus particulièrement Genet, remettent explicitement en cause⁶²⁰. Les échanges sur les caractères des collections d'archives visées (homogénéité, sérialité) s'inscrivent elles aussi dans le giron de ces discussions.

Pour dépasser ce type de limites, et permettre de mettre en œuvre des analyses automatisées des collections d'archives existantes, certains proposent de standardiser les formes de description des documents en fonction de leurs genres (chartes, actes notariés, archives judiciaires, etc.) et imaginent de nouvelles logiques de classement qui permettraient de donner accès à ces sources d'une nouvelle manière.

Dans cette direction, le colloque de Rome apparaît comme un moment fondateur, puisque les acteurs s'y proposent de mettre en commun les *documents de travail* qu'ils ont élaborées à l'échelle de leurs collectifs (thésaurus, programmes, fiches signalétiques,

⁶¹⁹ Furet, « Histoire quantitative et construction du fait historique », 67.

⁶²⁰ Genet, « Ordinateur, lexique, contexte ».

etc.). La réorganisation de la discipline autour des utilisations des moyens électroniques de calcul, que Genicot appelait de ces vœux en 1973, passe ainsi par des discussions qui portent aussi bien sur les sources elles-mêmes que sur ces documents de travail élaborés localement pour assurer l'enregistrement et la collecte des données. L'autre intérêt majeur de ce phénomène pour l'histoire de l'histoire est qu'il nous offre l'occasion d'observer comment se mettent en place des modalités de partage de ces produits des cultures scientifiques locales en même temps que s'exercent sur eux des injonctions d'harmonisation et de standardisation.

Une autre conséquence de ces échanges apparaît fondamentale pour notre sujet. Elle repose sur les difficultés que rencontrent les acteurs pour faire circuler les instruments de travail qu'ils élaborent sur supports informatiques (bande magnétique, carte perforée). Nous avons documenté ces difficultés, sur la base d'exemples précis, dans les deux premiers chapitres de cette thèse. Si Genicot propose la publication sur papier de ce type d'instruments, on comprend, notamment avec l'exemple d'utilisation de l'« *index verborum* » élaboré au CRAL que nous avons présenté, mais encore plus clairement avec le cas des programmes informatiques, que cette forme de publication limite fortement l'intérêt que peuvent présenter ces nouveaux instruments de recherches, que l'on a pu comparer à des moteurs de recherches. Pour favoriser la circulation de ce type d'instruments de travail en dehors des collectifs qui les ont produits, nous avons montré que certains acteurs imaginent de nouvelles formes de publication, comme la réalisation de « *modèle de description* » ou de plan-type, qui constituent là encore, une forme d'écrits nouvelle au sein de la discipline.

Nous avons encore mis en lumière qu'à ces difficultés liées à la circulation des instruments de travail réalisés par ces différents collectifs venaient s'ajouter, dans ce contexte, de grandes difficultés pour certains acteurs qui désiraient employer des ordinateurs pour mener leurs recherches. Le cas de Jehel, sur lequel nous avons insisté dans la troisième partie de ce chapitre, est à ce titre un exemple frappant de la situation dans laquelle se trouve le médiéviste en quête d'assistance et d'information sur ce type de méthodologie. Le témoignage, qu'il donne dans sa réponse au questionnaire élaboré par le secrétariat constitué à l'issue du colloque de Rome, conduit Fossier à réaliser qu'il est nécessaire, pour assurer le développement de l'utilisation des ordinateurs en histoire

médiévale, de centraliser les informations et d'offrir ainsi à la « *tribu* » des médiévistes⁶²¹ une publication à même d'aider ou de convaincre les « *hésitants* »⁶²².

Pour favoriser la circulation de ces nombreux nouveaux types d'objets scientifiques (instruments de travail, documents de travail, plan-type), favoriser la standardisation de ces nouvelles pratiques de recherches et aider les médiévistes néophytes à s'emparer de ces nouvelles méthodes, le petit groupe qui se forme à l'issue du colloque de Rome lance, en 1979, une publication originale, *Le Médiéviste et l'ordinateur*, sur laquelle nous nous appuyerons dans le chapitre suivant pour étudier les axes de recherche que nous avons dégagés ici.

⁶²¹ Cette expression est reprise de l'article déjà cité dans l'introduction générale de cette thèse : Olivier Dumoulin, « La tribu des médiévistes », *Genèses. Sciences sociales et histoire* 21, n° 1 (1995): 120-33.

⁶²² Ce terme est celui employé par Fossier dans son article précédemment cité : Fossier, « A propos d'un formulaire d'enquête relatif à "L'information sur les recours aux ordinateurs" », 721.

Chapitre 4 : Objectifs et stratégies de publication du bulletin de liaison *Le Médiéviste et l'ordinateur* (1979-1989).

Le colloque de Rome de 1975 aboutit, nous l'avons vu dans le chapitre précédent, à la constitution d'un premier noyau de six chercheurs regroupés autour de la section informatique de l'IRHT. On y retrouve 1) Lucie Fossier (1924-), directrice de cette section ; 2) Jean-Philippe Genet (1944-), médiéviste au sein de l'Équipe de Recherche Autonome 713 (ERA 713), dont nous avons présenté le travail au chapitre 2 ; 3) Christiane Klapisch-Zuber (1936-), maître-assistante au Centre de Recherches Historiques (CRH) de l'École des Hautes Études en Sciences Sociales, dont nous avons étudié le projet d'enquête du *catasto* au chapitre 1 ; 4) Jacques Lefort (1939-2014), spécialiste de l'histoire de l'empire byzantin à la IV^e section de l'École Pratique des Hautes Études (EPHE) ; 5) Josette Metman (1918-1986), historienne du droit médiéval au Centre d'Histoire Juridique de Paris, qui mène entre autres des projets en collaboration avec des chercheurs à l'université de Bourgogne et 6) Gian-Piero Zarri, informaticien rattaché au Centre National de la Recherche Scientifique (CNRS). Il faut ajouter à ces six membres une septième collaboratrice, Caroline Bourlet, ingénieure de recherche à l'IRHT et véritable bras droit de Fossier au sein de la section informatique, qui rejoint ce groupe à une date inconnue et signera avec eux le premier éditorial du *Médiéviste et l'ordinateur*.

Dès 1976, ce groupe avait lancé une enquête qui visait à recenser les projets des collectifs utilisant des ordinateurs en histoire médiévale, ainsi que les simples vellétés d'utilisation des moyens informatiques. Comme nous l'avons vu au chapitre précédent, le faible nombre de réponses reçues dans ce cadre avait conduit ces acteurs à repenser leur approche, en projetant non plus de recueillir des informations, mais plutôt d'en diffuser. En 1979, dans l'éditorial du premier numéro du *Médiéviste et l'ordinateur*, les sept éditeurs reviennent sur la genèse de cette publication en rapport avec les deux initiatives que sont le colloque de Rome et l'enquête de 1976. Ils insistent notamment sur

la continuité qui existe entre la rencontre de Rome et cette publication, rappelant que les débats du colloque avaient conduit à la conclusion suivante :

*« le manque d'information et de coordination était à l'origine d'une mauvaise ou insuffisante utilisation de l'ordinateur pour l'exploitation des sources du Moyen-Âge »*⁶²³.

Pour remédier à cette situation, il propose alors de créer, sur la base d'un premier groupe de 25 chercheurs, identifiés grâce à l'enquête, un

*« réseau d'informations qui peut-être les aidera à mieux orienter leur effort et, qui sait, à persuader avec plus d'assurance les réticents lorsqu'il leur semblera que l'utilisation de l'informatique peut porter ses fruits. »*⁶²⁴

Pour aller dans ce sens, ils insistent la nécessité de poursuivre leurs activités dans

*« trois directions : se rendre sur place pour voir ce qui existe et quels sont les moyens informatiques dont disposent les auteurs de projets ; organiser des rencontres entre ces derniers ; enfin, diffuser périodiquement un bulletin de liaison »*⁶²⁵.

Ils précisent par ailleurs que

*« l'organisation de rencontres ne portera ses fruits que si un premier bulletin de liaison a déjà circulé et provoqué des réactions »*⁶²⁶.

Le développement de ce réseau passe ainsi par un ensemble de stratégies dont le bulletin de liaison est la pierre angulaire.

⁶²³ Caroline Bourlet et al., « Éditorial », *Le médiéviste et l'ordinateur* 1, n° 1 (1979): 1. La version numérisée de ce bulletin est accessible depuis cette adresse : https://www.persee.fr/issue/medio_0223-3843_1979_num_1_1

⁶²⁴ Bourlet et al., 1.

⁶²⁵ Bourlet et al., 2.

⁶²⁶ Bourlet et al., 2.

Pour notre étude, qui vise, rappelons-le, à analyser les modalités de l'organisation méta-collective des collectifs impliqués dans l'utilisation des ordinateurs, *Le Médiéviste et l'ordinateur* offre de fait un matériau formidable. Plus largement, l'intérêt de ce type de sources pour l'histoire des sciences a notamment été mis en avant dans un article de 2012 par l'historien des sciences Christopher M. Kelty de l'université de Californie à Los Angeles⁶²⁷. Dans ce texte, qui se concentre sur les *newsletters* publiées en biologie, Kelty insiste sur le rôle central de ces périodiques dans la construction des communautés scientifiques depuis le début du XX^e siècle. La définition qu'il donne des *newsletters* nous permet d'inclure notre étude du *Médiéviste et l'ordinateur* dans la continuité de ses conclusions, puisqu'il affirme :

« *newsletters are tools of coordination and collaboration that emerge whenever a scientific or technical problem overruns the bounds of a single laboratory or office.* »⁶²⁸

Selon lui, trois principales raisons fondent l'intérêt de ce type de sources, dont deux nous intéresserons particulièrement ici : 1) la possibilité qu'elles offrent d'une étude de la constitution d'une communauté scientifique (« *scientific community* ») simultanément de la création de concepts, de techniques ou de théories et 2) l'accès à une échelle méso des pratiques scientifiques en train de se faire (« *science in action* »), entre la documentation privée des scientifiques et les formes de publications traditionnelles du monde académique que sont les revues.

Pour ces deux raisons, l'étude de ce type de sources se présente selon nous comme une voie particulièrement innovante pour approcher les organisations collectives en histoire des sciences humaines. En effet, ce type de publication offre à l'observateur une prise directe sur les modalités du partage des pratiques scientifiques des chercheurs, qui ne sont que trop rarement documentées autrement que sur la base d'études des publications académiques traditionnelles ou des supports d'enseignements (manuels, séminaire). De plus, ce type de sources permet de s'intéresser aux collectifs scientifiques

⁶²⁷ Christopher M Kelty, « This Is Not an Article: Model Organism Newsletters and the Question of 'Open Science' », *BioSocieties* 7, n° 2 (1 juin 2012): 140-168. Je remercie ici Mathias Grote (Humboldt-Universität zu Berlin) de m'avoir indiqué l'existence de cette ressource précieuse.

⁶²⁸ Kelty, 141.

en tant qu'ils se constituent indépendamment des logiques hiérarchiques et institutionnelles propres aux disciplines ou aux « écoles »⁶²⁹. L'histoire des humanités numériques, qui se caractérise par une forte interdisciplinarité et l'absence d'institution dominante, peut à ce titre bénéficier tout particulièrement de ce type d'approche, inédite à notre connaissance en histoire des sciences humaines.

Ces considérations nous paraissent d'autant plus adaptées à notre sujet que de nombreux périodiques similaires se développent dans le contexte qui nous intéresse. On peut notamment citer l'exemple du bulletin d'information *Computer and Medieval Data Processing*, mentionné dans le chapitre précédent, qui avait été créé en 1971 sous l'impulsion de la Medieval Academy of America, ou encore *Calculi*, édité et distribué par le Dartmouth College depuis 1967 pour rendre compte des travaux utilisant des ordinateurs en philologie⁶³⁰. Pour la France, le premier périodique de ce type est une publication intitulée les *Cahiers du CRAL*⁶³¹, créée en 1967 dans le but de publier les travaux menés au sein de ce centre de recherche que nous avons présenté dans le chapitre précédent.

Le Médiéviste et l'ordinateur se présente ainsi comme un cas d'étude idéal pour entamer l'étude du vaste corpus que nous dessinons ici. Deux raisons justifient cette affirmation. Tout d'abord, il s'agit, à notre connaissance, de la première publication spécialement dédiée aux relations entre la discipline historique et l'informatique en Europe. De plus, comme nous le verrons en conclusion, plusieurs autres *newsletters* ou bulletin d'information se développent en suivant son modèle pendant les années 1980, comme le périodique *Mémoire Vive*⁶³², créé par des historiens modernistes en 1989.

Pour étudier la spécificité de ce type de publications par rapport à d'autres, comme les revues académiques, il nous apparaît particulièrement pertinent de nous intéresser à ce que certains linguistes appellent la *situation de communication*. Ce concept, qui vise à préciser la notion plus flottante de contexte, désigne l'ensemble des conditions qui

⁶²⁹ Sur l'utilisation de la notion d'école en histoire des sciences humaines, voir notamment l'article introductif d'un numéro de la *Revue d'histoire des sciences humaines* dédié à ces questions : Olivier Orain, « Les Écoles en sciences de l'homme : usages indigènes et catégories analytiques », *Revue d'histoire des sciences humaines*, n° 32 (1 juillet 2018) : 7-38.

⁶³⁰ Dartmouth college, Department of Classics, *Calculi* (Hanover, N.H, Etats-Unis d'Amérique: Dept. of Classics, Dartmouth College, 1967).

⁶³¹ Centre de recherches et d'applications linguistiques, *Cahiers du C.R.A.L.* (Nancy, France: Université de Nancy 2, Centre de recherches et d'applications linguistiques, 1967).

⁶³² Association française pour l'histoire et l'informatique, *Mémoire vive* (Paris, France: Association française histoire et informatique, 1989).

président à l'acte de langage. Plusieurs grilles d'analyses ont été proposées par les linguistes pour en saisir pleinement toutes les dimensions, par exemple le modèle SPEAKING du sociolinguiste américain Dell Hymes (1927-2009)⁶³³.

Dans un article de 2004, le linguiste français Dominique Maingueneau résumait les différents éléments proposés jusque-là en sept points⁶³⁴, que nous adapterons à notre objet d'étude, le bulletin de liaison, en ces termes : 1) Quelle est sa finalité ? 2) Qui parle à qui ? 3) Quel est le mode d'existence matérielle de ce bulletin ? 4) Comment s'inscrit-il dans une temporalité ? 5) Quelles sont les conditions de production des textes qui le composent ? 6) Quel est le « canevas » dans lequel s'inscrivent les discours publiés dans le bulletin ? 7) Comment les différents répertoires de la langue sont-ils utilisés ?

Deux différences notables doivent cependant être soulignées par rapport à l'utilisation de cette grille telle que la présente Maingueneau. La première concerne l'objet visé par notre étude, qui n'est pas un texte, mais un ensemble de textes, qui prend la forme d'un *genre* de publication, le bulletin de liaison. Nous nous intéresserons cependant plus précisément à certains des articles qui composent le bulletin dans le chapitre suivant. La seconde différence consiste dans la finalité de notre approche. En effet, notre étude ne propose pas, par contraste avec les linguistes, d'utiliser cette grille d'analyse pour mener une étude sémantique du discours des acteurs, mais s'intéresse à la situation de communication comme indice des relations entre les éditeurs du bulletin et ses lecteurs, dans le but d'examiner les conditions qui président à la circulation des savoirs dans ce contexte.

Cette analyse de la situation de communication nous permettra par ailleurs d'étudier plus précisément les stratégies mises en place par les éditeurs du *Médiéviste et l'ordinateur* pour développer ce réseau au-delà du premier noyau de 25 membres évoqué plus haut. Pour en saisir pleinement la teneur, il faut insister à nouveau sur le fait que, contrairement aux bulletins d'information présentés ci-dessus (*CAMDAP*, *Calculi*), *Le Médiéviste et l'ordinateur* se définit, selon les termes des éditeurs, comme un bulletin de liaison. Cette distinction doit, selon nous, être prise au sérieux pour déceler les objectifs poursuivis par les éditeurs, qui tiennent autant, comme nous le montrerons, à la circulation

⁶³³ Dell Hymes, *Foundations in Sociolinguistics: An Ethnographic Approach* (University of Pennsylvania Press, 1974).

⁶³⁴ Dominique Maingueneau, « La situation d'énonciation entre langue et discours », in *Dix ans de S.D.U.* (Craiova: Editura Universitarai Craiova, 2004), 197-210.

des savoirs attachés aux utilisations des ordinateurs en histoire médiévale qu'à la construction d'une *culture épistémique* commune relevant de ces pratiques⁶³⁵.

Plusieurs questions clefs orienteront cette deuxième facette de notre analyse. Comment ce périodique met-il en relation les acteurs concernés par ces pratiques, qu'ils soient utilisateurs de ces nouvelles méthodes ou seulement intéressés par celles-ci en tant que lecteurs ? Qu'est-ce que les auteurs partagent dans les pages du bulletin ? Et comment les éditeurs orientent-ils les auteurs pour que leurs contributions s'accordent avec les objectifs poursuivis par le bulletin ? Enfin, comment cette publication s'intègre-t-elle dans la stratégie générale mise en place par les éditeurs pour promouvoir les usages des ordinateurs en histoire médiévale ?

Pour répondre à ces deux ensembles de questions, nous nous appuyerons dans un premier temps sur les archives du secrétariat de rédaction du *Médiéviste et l'ordinateur*, conservées au centre de l'IRHT à Orléans⁶³⁶. Ces documents nous permettront d'examiner comment les activités des éditeurs, depuis le choix des articles jusqu'à la distribution des bulletins de liaison aux abonnés, s'articulent avec le développement d'un réseau de chercheurs. Cette analyse nous conduira par ailleurs à dégager quelques caractéristiques générales propres à ce *genre* de publication, qui recourent les points 1), 3), 4) et 5) de la grille d'analyse de la situation de communication présenté ci-dessus.

Nous examinerons ensuite la fonction de l'organisation thématique du bulletin, mise en œuvre par les éditeurs dans l'intention de structurer les contenus des numéros. L'analyse que nous développerons s'appuiera en partie sur une procédure d'analyse automatique du vocabulaire de notre corpus rendu possible par la plateforme numérique CorText⁶³⁷, que nous présenterons dans la deuxième partie de ce chapitre. Cette approche nous permettra de comparer le « canevas » thématique élaboré par les éditeurs avec les

⁶³⁵ Ce concept de culture épistémique, que nous définirons en contexte dans la troisième partie de ce chapitre, est emprunté à K. Knorr-Cetina, *Epistemic cultures: how the sciences make knowledge* (Cambridge, Mass: Harvard University Press, 1999).

⁶³⁶ Il nous faut préciser ici que, comme dans le cas de certaines sources des chapitres précédents, aucun inventaire ni classement n'a été dressé par l'IRHT pour les fonds qui nous intéressent ici. Il nous est donc impossible de rendre compte aux lecteurs d'une cote d'archives pour les documents cités. Selon la solution adoptée précédemment, nous indiquerons dans le corps du texte les informations permettant de les identifier (type de document, auteurs, date, titre, auteur, destinataire, etc.). Je profite encore de cette note pour remercier Caroline Bourlet, qui m'a permis d'accéder à ces archives, Sandra Labouré, qui m'a accueilli sur les lieux, et tous les membres du centre de l'IRHT à Orléans que j'ai croisés au cours de mes séjours sur place.

⁶³⁷ Pour une présentation complète de la plateforme numérique CorText, dont nous préciserons les modalités de notre utilisation dans la deuxième partie de ce chapitre, voir : <https://docs.cortext.net/>

sujets effectivement abordés dans les pages du bulletin. Nous nous concentrerons uniquement sur les dix premières années d'activité des éditeurs, c'est-à-dire sur les vingt premiers numéros du bulletin. Le choix de ce corpus est motivé par la présence continue de Fossier dans le comité éditorial sur la période 1979-1989, et par la permanence de l'organisation interne du bulletin pendant cette période. Il est par ailleurs de dimension tout à fait respectable - 227 articles et près d'un million de mots - pour en permettre une analyse à l'aide d'outils d'analyse automatique de textes.

Nous nous intéresserons dans un troisième temps à la composition du lectorat du bulletin, sur la base des archives conservées à l'IRHT. Nous nous concentrerons particulièrement sur les stratégies mises en œuvre par les éditeurs dans le but de faire circuler des contenus techniques aux lecteurs et sur les modalités du partage des pratiques de la recherche assistée par ordinateurs, qui passent notamment par la reproduction de documents de travail dans les pages du bulletin. Cette analyse recoupera à la fois le point 2) et le point 7) de la grille d'analyse de la situation de communication présenté ci-dessus. Nous nous appuierons aussi sur d'autres travaux menés par des linguistes, et en particulier sur des études récentes menées en analyse du discours sur l'utilisation des questions directes.

1. L'organisation du *Médiéviste et l'ordinateur*.

Le bulletin de liaison *Le Médiéviste et l'ordinateur* peut, selon nous, être considéré comme la matrice d'une communauté discursive (*discourse community*), au sens que lui a donné le socio-linguiste anglais John Swales (1938-)⁶³⁸. Dans un article de 1987, Swales différencie la communauté linguistique (*speech community*) de la communauté discursive, en insistant notamment sur le fait que ce second concept se distingue du premier car il permet de définir l'appartenance de ces membres sur la base de discours écrits (*writing*), et non seulement de discours oraux (*speech*). De ce premier élément de définition découle un autre intérêt de ce concept : une communauté discursive

⁶³⁸ Ce concept a été massivement investi par les socio-linguistes depuis le début des années 1980. En 1987, Swales en propose une définition qui repose sur six critères, sur lesquels nous reviendrons au fil de l'analyse. Cette approche nous apparaît particulièrement pertinente, notamment en raison du fait qu'elle a été élaborée à propos de communautés discursives scientifiques. Voir John Swales, *Approaching the Concept of Discourse Community*, 1987.

réunit des membres indépendamment de leur proximité géographique. Pour reprendre les mots de Swales :

« *In a speech community, the community creates the discourse; in a discourse community, the discourse creates the community* »⁶³⁹.

Appliquées à notre cas, ces considérations nous conduisent à nous demander comment les éditeurs du *Médiéviste et l'ordinateur* ont organisé la production et la circulation de ces discours. Pour répondre à cette question, il nous apparaît nécessaire de nous intéresser à trois aspects de cette publication : 1) les relations qu'entretiennent ses caractères matériels et son organisation interne sous forme de rubriques avec les objectifs poursuivis par les éditeurs ou, en d'autres termes, avec sa finalité – que le linguiste Dominique Maingueneau définit comme la volonté du discours de modifier la situation dont il participe⁶⁴⁰ ; 2) les activités extra-éditoriales de ses éditeurs et 3) les logiques qui conduisent à la publication des articles dans le bulletin. L'examen de ces trois points nous permettra d'analyser, dans les deuxième et troisième parties de ce chapitre, les stratégies mises en place par les éditeurs pour développer une culture épistémique commune au sein de cette communauté.

1.1. Qu'est-ce que *Le Médiéviste et l'ordinateur* ?

Comme nous l'avons mentionné en introduction, *Le Médiéviste et l'ordinateur* se caractérise avant tout par les objectifs que poursuivent ses éditeurs au sein de la discipline qu'ils investissent. En effet, le bulletin de liaison n'a pas pour finalité de partager des résultats de recherches ou de proposer un nouveau programme historiographique⁶⁴¹. Les éditeurs cherchent au contraire à dévoiler des pratiques de recherches et à informer les abonnés sur les possibilités ouvertes par ces nouvelles pratiques comme sur les moyens de s'y former. En ce sens, les acteurs partagent ce que Swales a qualifié dans son article

⁶³⁹ Swales. Pour une réflexion plus récente sur cette définition initiale, voir John M. Swales, « Reflections on the Concept of Discourse Community », *ASp. La Revue Du GERAS*, n° 69 (9 mars 2016): 7-19.

⁶⁴⁰ Maingueneau, « La situation d'énonciation entre langue et discours », 6.

⁶⁴¹ Plusieurs publications, nées pendant cette période, s'inscrivent dans ce type d'ambition programmatique. On notera notamment *Histoire & Mesure*, déjà mentionnée dans les chapitres précédents, mais également, par exemple, la revue *Genèses*. Voir Vincent Heimendinger, « Genèses en train de se faire : écrire l'histoire d'une revue scientifique », *Zilsel* N° 5, n° 1 (2019): 271-314.

de 1987 d'objectifs publics communs (*common public goals*), éléments qu'il place au centre de la définition des communautés discursives.

Pour poursuivre ces objectifs, les éditeurs structurent la publication sous une forme qui s'articule avec cette finalité. Les logiques de regroupement des articles en numéros thématiques, sur lesquelles nous reviendrons dans la deuxième partie de ce chapitre, comme l'organisation interne de chaque numéro sous la forme de rubriques, nous révèlent que, dès le premier numéro, les éditeurs ont fait en sorte de structurer la publication afin de donner aux lecteurs, comme aux usagers, les clefs pour s'emparer de ces nouvelles pratiques de travail.

1.1.1. Finalité et structure interne du bulletin.

Si l'on met de côté les éditoriaux, le bulletin est constitué de cinq rubriques : 1) des « *Articles de fond* », qui abordent des projets en cours ; 2) des « *Notes techniques* », qui précisent un point mathématique, statistique ou informatique (le plus souvent assorti d'une bibliographie) ; 3) des comptes-rendus de visites dans des centres de calcul, qui présentent l'équipement, les programmes informatiques et les infrastructures disponibles ; 4) des informations, qui prennent la forme de calendriers des rencontres à venir ou de comptes-rendus d'événements passés, et enfin 5) des courriers de lecteurs, par le biais desquels les usagers et les lecteurs seraient à même de poser des questions et d'interagir avec la communauté.

Chacune de ces rubriques vise ainsi à mettre en relation les divers acteurs impliqués dans ces nouvelles pratiques selon des modalités différentes. Les articles de fonds ont pour objet de permettre aux collectifs impliqués dans ce type de projet de dévoiler les particularités des cultures de travail locales dans lesquels s'inscrivent leurs travaux. Nous reviendrons dans la deuxième partie de ce chapitre sur les stratégies mises en place par les éditeurs pour favoriser ce type de mise en relation de collectif à collectif, et d'usagers à usagers. Les notes techniques participent quant à elles d'un effort de formation et de persuasion auprès des lecteurs, à travers la présentation de procédures mathématiques, statistiques et informatiques, dont nous étudierons plus précisément la teneur dans la troisième partie de ce chapitre. Les comptes-rendus de visite au centre de calcul et les calendriers de rencontres donnent quant à eux des informations précises sur les lieux de la recherche (adresse, contact des responsables, etc.) et visent à inciter la mise

en relation directe des chercheurs entre eux. Enfin, la rubrique « *Courrier des lecteurs* » offre aux lecteurs un espace pour réagir aux articles et aux informations publiées, renforçant la dimension participative de cette publication.

L'analyse de la fonction de ces différentes rubriques met en lumière trois autres éléments caractéristiques des communautés discursives dégagés par Swales. Le premier concerne la fonction du bulletin de liaison dans le développement de cette communauté, en tant que mécanisme d'intercommunication entre les membres (*mechanisms for intercommunication*). Le bulletin vient ainsi s'ajouter à d'autres mécanismes préexistants, comme les rencontres et les colloques. Il contribue par là au développement de cette communauté en offrant non seulement des informations indispensables à la promotion de ce type de pratiques, mais également en permettant la publication des réactions des membres de la communauté, deux autres critères qui, selon Swales, caractérisent les communautés discursives.

1.1.2. Matérialité et succès.

La poursuite de ces différents objectifs (informer, former, mettre en relation directe, publier des réactions) s'incarne par ailleurs dans la forme matérielle du *Médiéviste et l'ordinateur*. Tout d'abord, les numéros sont courts. Contrairement à des revues académiques plus traditionnelles comme les *Annales ESC*, les *Cahiers de civilisations médiévales* ou *Histoire & Mesure*, qui proposent des livraisons de près de 400 pages, les numéros du *Médiéviste et l'ordinateur* comportent en moyenne vingt pages, et jamais plus de trente. Cette forme permet un rythme de parution régulier, avec deux numéros par an, un au printemps et un en automne.

Entre 1979 et 1989, les numéros sont composés, selon les témoignages des acteurs, de quelques dizaines de feuillets dactylographiés, agrafés ensemble⁶⁴². Une série d'entretiens, menés dans le cadre de cette thèse, entre 2016 et 2019, permet de mieux comprendre la spécificité du support d'impression de ce périodique. Parmi les lecteurs interrogés⁶⁴³, le bulletin est régulièrement qualifié de « *feuille de chou* », c'est-à-dire de

⁶⁴² Voir Le comité de Rédaction, « Trente ans d'Histoire & Mesure. Entretiens croisés avec Gérard Béaur, Jean-Philippe Genet et Jean Heffer », *Histoire & Mesure* XXXI, n° XXXI-2 (2016): 3-9.

⁶⁴³ Les lecteurs que nous avons pu interroger sont le plus souvent aussi des auteurs, voir des éditeurs de la revue : Christiane Klapisch-Zuber, Jean-Philippe Genet, Caroline Bourlet, mais aussi Alain Guerreau. Je profite de cette note pour remercier ces interlocuteurs précieux.

journal de peu de valeur. Ce jugement porte-t-il sur la taille de chaque numéro, sur le contenu, ou sur le support d'impression du bulletin ? Sans pouvoir trancher, il ressort de ces éléments sur le support d'impression que le statut de cette publication différait d'une revue classique.

Tout porte toutefois à croire que les considérations des acteurs sur le support allaient de pair avec la nature provisoire des informations contenues dans le bulletin. Cette publication, et nous y reviendrons dans la troisième partie de ce chapitre, visait en effet à donner aux lecteurs des informations précises, mais valables pour de courtes durées, comme des adresses de centre de recherches, des dates de colloques ou encore l'état de projets en cours de réalisation. Cette remarque s'applique aussi aux contenus portant sur les techniques et les équipements informatiques, qui avaient eux aussi vocation à évoluer rapidement. *Le Médiéviste et l'ordinateur* offre ainsi à l'observateur une image de la recherche en train de se faire, avec toutes les limites que cela peut supposer.

Ces considérations sur le support d'impression doivent encore être mises en perspective avec les modes de distribution de ce périodique. En effet, *Le Médiéviste et l'ordinateur* est distribué gratuitement aux lecteurs, qui souscrivent à un abonnement par simple envoi d'une demande auprès des éditeurs. À la différence des revues académiques plus traditionnelles dont il a été question plus haut, cette publication ne passe donc pas par les canaux traditionnels des maisons d'éditions scientifiques, et ne répond pas non plus aux logiques de l'édition commerciale. Pendant la période considérée (1979-1989), le bulletin de liaison est édité et imprimé par la section informatique de l'Institut de Recherche et d'Histoire des Textes (IRHT). Il est ensuite distribué directement par l'IRHT, mais aussi par l'Équipe de Recherche Autonome 713 (ERA 713) de l'université Paris 1 Panthéon-Sorbonne et, à partir de 1980, par le Centre de Recherches Historiques (CRH) de l'EHESS.

On peut d'ailleurs estimer approximativement le tirage du bulletin à partir des archives de l'IRHT conservées à Orléans, et en particulier à l'aide d'un document du carton intitulé « Liste d'abonnés/appel à réabonnement ». Un bilan, établi par Bourlet en 1989, nous renseigne sur le fait que le nombre d'abonnés est en constante augmentation depuis 1979, puisqu'il passe pendant cette période de 350 à 900. Il est donc plutôt élevé par comparaison avec des publications de nature différentes, mais s'adressant au même type de lectorat. La revue académique *Cahiers de Civilisation Médiévale*, comptait par

exemple, en 2002, 1037 abonnés⁶⁴⁴. Pour *Histoire & Mesure*, une autre revue académique lancée en 1986, « les abonnements n'ont jamais dépassé les 200 à 300 » et « seuls les numéros spéciaux se vendaient correctement » selon le témoignage de Genet, l'un des éditeurs⁶⁴⁵. Certains numéros du *Médiéviste et l'ordinateur* seront même distribués en dehors du cercle des abonnés, comme ce fut le cas de celui dédié à l'édition électronique (numéro 16), qui fera l'objet d'un second tirage pour satisfaire les demandes.

Le succès de cette publication nous éclaire sur la réussite du premier objectif poursuivi par les éditeurs, à savoir, d'étendre le réseau au-delà du premier noyau de 25 chercheurs identifiés grâce à l'enquête de 1977. Les ingrédients de ce succès reposent, selon nous, essentiellement sur le type de contenu partagé dans les pages du bulletin (information périssable, contenu technique, réactions) et sur les caractéristiques matérielles de ce dernier (nombre de pages, support d'impression, mode de distribution). Le bulletin est au centre de la constitution de cette communauté discursive : il est à la fois le support privilégié de la circulation des informations, des savoirs, des pratiques et l'instance de la mise en relation directe entre ses membres.

Sur la base de cette étude de cas, il nous apparaît que la spécificité du *genre* de périodique « bulletin de liaison » repose justement sur cette double fonction, en se distinguant des bulletins d'informations cités plus haut par cette volonté affichée par les éditeurs de favoriser la mise en relation directe des acteurs les uns avec les autres. Cet élément de définition nous apparaît d'autant plus pertinent que, nous allons le voir, le rôle des éditeurs dans le fonctionnement de cette communauté discursive dépasse largement le cadre de ces seules activités éditoriales, qui ne constituent qu'un des mécanismes d'intercommunication entre ses membres.

1.2. Bien plus qu'une publication.

Outre ces activités d'édition, d'impression et de distribution du bulletin, le secrétariat de rédaction du bulletin remplissait encore trois autres fonctions que nous allons analyser tour à tour. Il était 1) une interface de contact entre les acteurs de ce

⁶⁴⁴ Martin Aurell, « Les Cahiers de Civilisation Médiévale au service de l'interdisciplinarité. », in *Actes de la Table Ronde de l'Institut Historique Allemand* (Paris, France, 2009), 10-25.

⁶⁴⁵ Rédaction, « Trente ans d'Histoire & Mesure. Entretiens croisés avec Gérard Béaur, Jean-Philippe Genet et Jean Heffer ».

réseau ; 2) il fournissait un ensemble de conseils sur des aspects techniques de la recherche assistée par ordinateur, et 3) certains de ses membres étaient consultés en tant qu'experts sur des projets impliquant l'utilisation des ordinateurs en sciences humaines en France et à l'étranger.

1.2.1. Un annuaire de personnes et de lieux.

Trois types de sources nous renseignent sur l'importance du secrétariat comme point de contact entre les membres du réseau. Nous disposons, tout d'abord, de nombreuses lettres adressées par des chercheurs de différents collectifs aux membres du secrétariat. Dans la plupart des cas, un lecteur découvre un projet, ou une équipe de recherche, utilisant des moyens informatiques et se tourne vers le secrétariat du bulletin pour obtenir l'adresse de ses membres et les contacter. Parfois, ils entendent simplement obtenir de plus amples informations, dans la mesure où ils projettent une opération sur des sources de même type. Parfois aussi, ils rencontrent un problème précis pour la résolution duquel ils espèrent recevoir de l'aide.

C'est le cas d'une lettre du 17 février 1987 de B. S. Jackson, professeur de droit à l'université de Kent, retrouvée aux archives de l'IRHT à Orléans, dans la pochette « correspondances » du carton intitulé « Comité de rédaction (vol. 1-34) ». Jackson y exprime le souhait d'être mis en contact avec Françoise Micheau (1945-), maître de conférence en histoire médiévale à l'université Paris 1 Panthéon-Sorbonne, et Jean-Pierre Molénat (1940-), directeur de recherche à l'IRHT, qui co-signent un article dans le numéro sur l'édition électronique paru à l'automne 1986. Dans ce texte, ces deux médiévistes présentent le programme qu'ils ont utilisé, intitulé FONTastic, pour créer une police de caractères d'alphabet arabe pour un ordinateur personnel Macintosh. Dans sa lettre, Jackson cherche notamment à comprendre si ce programme lui permettrait de créer des caractères d'alphabet avec des points souscrits pour la translittération de l'hébreu.

Cet exemple nous montre comment le bulletin œuvre comme un point de contact, une interface entre les acteurs, dans la mesure où les membres du réseau, qu'ils soient usagers ou lecteurs, trouvent au secrétariat un annuaire des personnes susceptibles d'être contactées. Les mécanismes d'intercommunication mis en place via le bulletin de liaison se poursuivent ainsi par d'autres voies, en dehors des pages du *Médiéviste et l'ordinateur*.

Un autre exemple de mécanismes d'intercommunication préexiste par ailleurs à la création de ce périodique. En effet, comme nous l'avons vu dans le chapitre précédent, les éditeurs lancent en 1975 une première enquête visant à recenser l'ensemble des projets qui recourent à des moyens informatiques en France et en Italie. Ce type d'initiative se répétera d'ailleurs à plusieurs reprises entre 1975 et 1990, conduisant les éditeurs à découvrir de nouveaux collectifs et de nouveaux chercheurs utilisant des ordinateurs. Les archives du secrétariat de rédaction en gardent d'ailleurs la trace puisqu'on y retrouve un carton complet ayant pour intitulé « Enquête traitement informatique des documents médiévaux », composé de pochettes consacrées chacune à un chercheur ou à une équipe de recherche. On y trouve des lettres, des notes, des rapports d'activité et des articles imprimés présentant les projets de ces collectifs.

Certaines pochettes attestent de l'addition de nouvelles découvertes jusqu'en 1990, c'est-à-dire près d'un an après la retraite de Fossier. Le secrétariat de rédaction du bulletin lance d'ailleurs une nouvelle enquête, basée sur un nouveau questionnaire, en 1989, dans le but d'évaluer l'évolution du champ depuis le début de la publication. D'autres initiatives viennent par ailleurs compléter ce travail d'enquête, et attestent du rôle pivot que jouait ce secrétariat, mais aussi plus largement la section informatique de l'IRHT. En 1982, Bourlet, à la fois membre de la section informatique de l'IRHT et éditrice du *Médiéviste et l'ordinateur*, publie par exemple, en collaboration avec Serge Lusignan (1943-), responsable de la newsletter américaine *CAMDAP*, le premier tome d'une bibliographie à volonté exhaustive portant sur l'ensemble des travaux utilisant l'informatique en histoire médiévale depuis le fameux *Index Thomisticus* de Roberto Busa (1913-2011) entamé en 1948⁶⁴⁶.

Cette fonction de centralisation et de diffusions des informations jouée par le secrétariat de rédaction s'articule par ailleurs avec les activités éditoriales elles-mêmes. Les articles de la rubrique « *Voyage au centre de calcul* » témoignent de cette relation. Pour chaque numéro, un ou deux membres du comité de rédaction se rendent dans des centres de calculs de toute l'Europe pour visiter leurs infrastructures et rencontrer des ingénieurs et des historiens⁶⁴⁷. Ces rencontres leur permettent non

⁶⁴⁶ Caroline Bourlet, Charles Doutrelepon, et Serge Lusignan, *Ordinateur et études médiévales : Bibliographie* (Montréal, Canada: Institut d'études médiévales, 1982).

⁶⁴⁷ Les membres du comité éditorial visitent ainsi des centres de calculs situés dans les villes suivantes : Poitiers, Nancy, Oxford, Cambridge, Orsay, Montpellier, Orléans, Nice, Bordeaux, Marseille, Limoges,

seulement d'élargir leur réseau, en découvrant notamment de futurs auteurs pour le bulletin, mais également d'ouvrir la voie vers de futures collaborations entre les membres de cette communauté discursive. Pour favoriser ce type d'interaction, chaque article de cette rubrique indique une adresse postale et le nom d'un référent à contacter au cas où un lecteur désirerait avoir des informations complémentaires, ou lancer un projet de traitement informatique en collaboration avec le centre de calcul concerné.

1.2.2. Consultations techniques.

Les archives de l'IRHT nous renseignent sur une deuxième fonction assurée par le comité de rédaction, en particulier par Fossier et Bourlet, qui sont par ailleurs toutes les deux membres de la section informatique de l'IRHT. Dans deux lettres identifiées dans la pochette « correspondances » du carton intitulé « Comité de rédaction (vol. 1-34) », des chercheurs contactent les éditeurs pour avoir des conseils sur l'achat de matériel informatique. Dans le premier cas, pour lequel la lettre n'est pas datée, un historien contacte le secrétariat pour demander des conseils sur l'acquisition d'un micro-ordinateur. Quel ordinateur doit-il choisir ? Avec quels périphériques ? En l'absence de carbone de réponse, il est cependant impossible de savoir 1) si ce dernier a écrit au comité suite à la lecture du bulletin et 2) si un membre du comité de rédaction a bien pris en charge cette demande.

Dans le second cas, qu'atteste une autre lettre qui n'est pas non plus datée, le lien établi avec la publication du bulletin est plus clair, puisque le chercheur mentionne qu'il écrit en réaction à la lecture du numéro 16 du bulletin. La question est elle-aussi plus précise, puisqu'il s'agit pour le chercheur de résoudre un problème de connectique entre du matériel Thomson et Apple, dans le but de transférer des fichiers textes. Là encore, il est impossible de savoir si une réponse lui est effectivement parvenue, mais on retrouve au dos de la lettre un schéma qui indique comment réaliser ce transfert. Il paraît donc fort probable qu'un membre du comité de rédaction ait répondu à sa demande.

En plus d'une consultation technique sur l'aspect *hardware*, les archives révèlent qu'à plusieurs reprises le comité de rédaction a donné des conseils sur l'aspect *software*.

Saint-Etienne et Besançon. Pour une étude précise de la répartition de ces « pôles scientifiques » sur le territoire français voir Michel Grossetti et Pierre-Eric Mounier-Kuhn, « Les débuts de l'informatique dans les universités », *Revue française de sociologie* XXXVI-n°2 (1995): 295-324.

À ce titre, la correspondance entre Georges Jehel et Fossier, dont le chapitre 3 a déjà permis d'esquisser le déroulement, en fournit l'illustration la plus intéressante. Dans la pochette « Gênes » du carton d'archives intitulé « Enquête traitement informatique des doc médiévaux », on retrouve un fascicule communiqué par Jehel, non daté, et titré « Projet de traitement informatique des données établies à partir des minutiers notariaux génois du XIII^e siècle conservés à l'*Archivio di Stato* de Gênes ». La dernière partie de ce fascicule, intitulé « Opérations demandées à l'ordinateur », est d'un intérêt tout particulier. Jehel y présente les éléments des documents qu'il souhaite enregistrer et indique les opérations statistiques qu'il ambitionne de réaliser sur la base de ces données.

On retrouve sur ce document des annotations de Fossier au crayon, qui indiquent dans la marge le programme à utiliser pour chacune des opérations : en l'espèce, CLIO, un des programmes conçus à l'IRHT, permettrait la réalisation d'une partie des traitements imaginés par Jehel, auquel elle indique qu'il serait nécessaire de combiner un autre programme, développé par le médiéviste Michel Balard (1936-) pour son travail sur les archives notariales de Gênes⁶⁴⁸. Dans ce cas, Fossier indique non seulement à Jehel les programmes informatiques à utiliser pour réaliser les différentes opérations statistiques imaginées par l'historien, mais elle organise aussi un rendez-vous avec Balard pour assurer la poursuite du projet. Ces éléments éclairent donc les activités de conseil technique, sur le matériel informatique comme sur les logiciels, assurées par certains membres du comité de rédaction.

Il faut cependant insister ici sur le fait que la configuration archivistique dans laquelle nous nous trouvons ne permet pas de distinguer clairement les activités qui sont du ressort de la section informatique de l'IRHT de celles qui sont du ressort du comité de rédaction du bulletin, du fait des titres de certains cartons. Dans ce cadre, il semble important d'insister sur le fait qu'il est possible que ce type d'activités extra-éditoriales ne soient assurées que par les membres de la section informatique. Toutefois, il reste probable que d'autres membres du comité apportent également des conseils techniques, mais que ces activités ne puissent être observées dans les archives de l'IRHT.

⁶⁴⁸ Ce travail est présenté dans le numéro 3 du *Médiéviste et l'ordinateur* : Michel Balard, « Les actes notariés de Gênes », *Le médiéviste et l'ordinateur* 3, n° 1 (1980): 3-5.

1.2.3. Expertise.

Une troisième fonction du comité de rédaction apparaît encore à la lecture des archives conservées à l'IRHT. Elle concerne le rôle d'expert joué par Fossier ou Bourlet dans des projets extérieurs à leur cadre institutionnel. Trois cas attestent de ce type d'activité. Le premier concerne l'expertise du réseau FRANTIQ, que Fossier réalise à la fin des années 1980. Ce projet, lancé en 1984, avait pour ambition d'élaborer un catalogue collectif des fonds de plusieurs bibliothèques françaises contenant des ressources pour l'étude de la période ancienne. En 1987, une première version du catalogue est disponible sur un serveur de l'Éducation Nationale et interrogeable par Minitel. En 1988, Fossier dirige un audit, commandé par le département des Sciences Humaines et Sociales du Centre National de la Recherche Scientifique (CNRS), qui conclut à la prise en charge du financement de la poursuite de ce projet⁶⁴⁹. Malheureusement, l'absence d'éléments dans les archives de l'IRHT ne nous permet pas de développer le rôle que joue Fossier dans cette aventure.

Nous nous intéresserons donc plutôt au deuxième cas, qui consiste en une demande d'expertise formulée par le Ministère de l'industrie et de la recherche auprès de Fossier. Dans ce cadre, l'objectif du ministère est de publier un rapport sur les besoins informatiques des Sciences humaines et sociales pour la période 1984-1989, afin de mettre en avant les besoins des équipes de recherches des universités et du CNRS. À travers un bref échange de lettres, conservées dans la pochette intitulée « MST Besoins de la recherche en moyens informatiques » du carton « Archives diverses – Lucie Fossier », le responsable de cette initiative au ministère lui demande d'évaluer le contenu du rapport, en insistant par ailleurs sur les possibles oublis ou sur les insuffisances qu'elle serait en mesure d'identifier. Dans les deux cas présentés ci-dessus, le rôle d'experte joué par cette médiéviste auprès du ministère atteste de la position prise par la section informatique de l'IRHT, et par Fossier en particulier, dans le champ plus général de l'informatisation des sciences humaines et sociales au milieu des années 1980.

Une autre demande de ce type, adressée cette fois à Bourlet en 1990, éclaire la position dont jouissent les membres de la section informatique de l'IRHT après la retraite de Fossier l'année précédente. Ce sont ici des linguistes anglo-saxons, Lou Burnard

⁶⁴⁹ Ces informations sont tirées de la page « Historique » du site web du réseau FRANTIQ : <https://www.frantiq.fr/frantiq/a-propos/historique/>

(1946-) et Michael Sperberg-McQueen (1954-), qui demandent à Bourlet d'évaluer, en amont de sa publication, un document de travail, le *Guideline for the Encoding and Interchange of Machine-Readable Texts (TEI P1)*. Ce document vise à permettre l'harmonisation à l'échelle internationale des normes d'annotation pour les enregistrements en *full-text*. L'évaluation de ce projet, aujourd'hui connu sous le nom de Text Encoding Initiative (TEI)⁶⁵⁰, s'inscrit donc dans la continuité du rôle joué par la section informatique de l'IRHT, qui s'impose pour les collectifs de l'étranger comme l'un des partenaires de référence pour ce type de projet.

Il s'avère cependant difficile, pour chacun de ces trois cas, de saisir dans quelle mesure la renommée du service informatique de l'IRHT est corrélée avec la publication du *Médiéviste et l'ordinateur*. Les liens qui unissent le comité de rédaction et la section informatique sont, comme nous l'avons vu, très ténus, et difficiles à démêler. Il ressort cependant de l'analyse de ces deux dernières fonctions que l'organisation de cette communauté discursive repose sur le haut degré de compétences d'au moins certains membres du secrétariat de rédaction du bulletin. Cette caractéristique est d'ailleurs mise en avant par Swales dans son article de 1987, comme l'un des critères de définition des communautés discursives⁶⁵¹.

Le dévoilement du rôle fondamental que joue les éditeurs du *Médiéviste et l'ordinateur* dans la constitution et le développement d'un réseau de chercheurs utilisant des ordinateurs en histoire médiévale ne saurait cependant être complet si l'on ne s'intéressait pas à leurs activités proprement éditoriales. En effet, en tant qu'administrateurs de la production textuelle de cette communauté de discours, les éditeurs sont en mesure de 1) choisir les discours qui circulent et de 2) façonner les modalités de cette circulation. Nous nous intéresserons ici au premier de ces deux points, à travers l'analyse des scénarios qui aboutissent à la publication des articles dans le bulletin. Nous nous pencherons ensuite dans les deuxième et troisième parties de ce chapitre sur quelques stratégies mises en place par les éditeurs pour façonner les modalités de la circulation des contenus dans les pages du *Médiéviste et l'ordinateur*.

⁶⁵⁰ Pour une présentation synthétique de la TEI en des termes actuels voir : Lou Burnard, « La TEI et le XML », in *Qu'est-ce que la Text Encoding Initiative ?*, Encyclopédie numérique (Marseille: OpenEdition Press, 2015). Pour une approche historique de ce projet, dans les termes des acteurs, et sous la forme d'une communication voir : Lou Burnard, « Histoire de la TEI : un cas d'étude dans l'évolution des méthodes et pratiques scientifiques dans les SHS ? », *Les Dossiers du Grihl*, 20 avril 2012.

⁶⁵¹ Swales, *Approaching the Concept of Discourse Community*.

1.3. Qui écrit dans le *Médiéviste et l'ordinateur* ?

Pendant les dix ans qui nous intéressent ici, 136 auteurs, aussi bien des médiévistes, des modernistes, des informaticiens, des linguistes, des mathématiciens, des archéologues, des archivistes que des sociologues, venant de huit pays différents⁶⁵², publient presque 200 articles dans *Le Médiéviste et l'ordinateur*. Pourtant, parmi ces auteurs, on remarque assez vite que certains noms reviennent à plusieurs reprises, alors que d'autres n'apparaissent qu'une fois. Sur les 136 auteurs mentionnés ci-dessus, 21 écrivent ainsi plus d'une fois dans la revue. Certains écrivent même plus de 25 fois.

Comme nous allons le voir, la compréhension du fonctionnement de la revue passe donc par 1) l'analyse de la répartition des textes par auteurs, qui nous permettra de montrer l'immense contribution de certains membres du comité éditorial dans la poursuite de la mission d'information qu'ils se sont fixés et 2) l'analyse des processus qui conduisent à la publication des articles, qui nous conduira à montrer comment cet aspect de l'activité éditoriale est absolument fondamental dans le développement du réseau voulu par les éditeurs.

1.3.1. Trois scénarios de sélection des auteurs.

Trois types de scénarios conduisent à la publication de textes dans les différentes rubriques du bulletin. Dans un très grand nombre de cas, ce sont les membres du comité de rédaction eux-mêmes qui prennent en charge l'écriture de ces textes. Ce canal de contribution concerne 41% des textes du bulletin, soit plus de 83 articles pour huit auteurs⁶⁵³. Les 113 textes restants sont quant à eux écrits par 128 auteurs. L'animation, comme l'existence du bulletin, passe donc avant tout par l'implication des éditeurs, dont

⁶⁵² Ces auteurs travaillent aussi bien en France, qu'en Italie, en Allemagne, en Belgique, en Suisse, au Royaume-Uni, aux États-Unis et au Canada.

⁶⁵³ Aux sept éditeurs mentionnés plus haut, il faut ajouter Hélène Millet (1944-), membre de l'ERA 713, qui rejoint le comité de rédaction au cours de la période (à une date que nous n'avons pas pu déterminer).

les activités de rédaction constituent une base solide qui soutient la poursuite des objectifs formulés dès 1979.

En allant plus en avant dans l'analyse, on remarque cependant qu'au sein même du comité éditorial, la répartition du nombre d'articles par auteur est elle aussi déséquilibrée.

Figure 18. Nombre d'articles rédigés par les membres du comité éditorial du *Médiéviste et l'ordinateur* (1979-1989).

Membres	Articles individuels	Articles collectifs	Nombre total d'articles
L. Fossier	26	4	30
J-P. Genet	21	2	23
C. Bourlet	9	6	15
C. Klapisch	7	2	9
G-P. Zarri	5	4	9
J. Metman	4	2	6
H. Millet	4	1	5
J. Lefort	2	1	3

Le tableau ci-dessus présente le nombre de textes rédigés par chacun des huit membres du comité de rédaction sur la période 1979-1989. On remarque immédiatement que certains membres sont plus actifs que d'autres dans l'animation du bulletin (Fossier et Genet particulièrement, mais aussi Bourlet). On remarque encore qu'un des membres du comité, Lefort, n'a signé que deux articles, mis à part l'éditorial du premier numéro. Ces disparités dans la contribution au bulletin reflètent plus précisément le fonctionnement de la publication. Trois des membres du comité de rédaction assurent l'essentiel du travail d'écriture, puisqu'ils cumulent à eux trois près de 60 textes sur les 199 publiés sur la période.

Les membres du comité de rédaction se chargent essentiellement de présenter un numéro à travers la rédaction d'éditoriaux, de rédiger un compte-rendu de visite d'un

centre de calcul (rubrique « *Voyage au centre de calcul* »), d'une conférence ou d'un colloque (rubrique « *Information* »). Leurs contributions se concentrent ainsi sur des rubriques particulières, qui sont attachées aux objectifs de circulation des informations poursuivies par les éditeurs.

Cependant, on retrouve aussi les textes de certains d'entre eux dans la rubrique « *Note technique* ». Zarri, seul informaticien du comité éditorial, publie par exemple trois articles dans cette rubrique. Mais c'est aussi le cas, dans une moindre mesure, de Genet et Millet, qui publient chacun une note technique sur l'analyse factorielle des correspondances dans le numéro 5 du printemps 1981. Dans certains cas, comme nous le verrons dans la troisième partie de ce chapitre, les éditeurs assument donc aussi un rôle de spécialiste de certaines techniques statistiques ou informatiques auprès des lecteurs, sur lequel nous reviendrons dans la troisième partie de ce chapitre. Plus surprenant, on retrouve encore les contributions des éditeurs dans la rubrique « *Courrier des lecteurs* ». Cinq textes, sur les 29 que compte le corpus, sont de la plume des éditeurs eux-mêmes. Ces nombreux indices éclairent pourquoi et comment l'activité de rédaction des huit membres du comité éditorial est fondamentale pour la vie du bulletin dans ces premières années.

Le deuxième scénario le plus répandu se caractérise par une demande directe d'un membre du comité de rédaction à un chercheur. Ces articles « commandés » par les éditeurs du bulletin s'inscrivent, pour près des deux tiers d'entre eux, dans la rubrique « *Articles de fonds* ». Ils apparaissent donc là où on les attend, puisque cette rubrique a pour vocation de présenter aux lecteurs des projets en cours ou achevés. De même, plus des trois quarts des « *Notes techniques* », proviennent de commandes (18 textes sur 23). Ce scénario de recrutement des auteurs constitue ainsi un matériel formidable pour étudier les modalités du développement du réseau qu'aspirent à constituer les éditeurs. L'analyse des articles de la rubrique « *Information* » donne, à ce titre, de très bons indices sur l'un des aspects de ce phénomène.

Dans cette rubrique, près de la moitié des articles sont commandés à des chercheurs dont il est possible d'identifier clairement les relations avec des membres du comité éditorial. Dans les six premiers numéros, on retrouve par exemple de nombreux articles rédigés par les collaborateurs de certains éditeurs. Tout d'abord, les articles proviennent des membres des mêmes laboratoires : deux articles sont écrits par des

membres de l'IRHT, et deux par des membres du Centre de Recherches Historiques (CRH), le laboratoire de Klapisch-Zuber. Pour la période qui nous intéresse, il est aussi possible de repérer deux comptes-rendus de Michel Parisse (1936-2020), directeur de la section diplomatique du CRAL de Nancy, fondée par Fossier, et deux autres articles écrits par Lusignan, éditeur du *CAMDAP* et collaborateur du *Médiéviste et l'ordinateur* depuis 1979.

D'autres types de relation entre les auteurs et les éditeurs peuvent encore être identifiés sur la base des archives de l'IRHT. On constate par exemple que certaines activités des membres du comité éditorial, comme leurs visites au centre de calcul ou leurs participations à des rencontres, aboutissent à l'identification de nouveaux auteurs potentiels. Un ensemble de documents, retrouvé aux archives dans la pochette intitulée « Textes en réserve » du carton « Textes en réserve - colloque de 1989 - hommage à Lucie Fossier », illustre par exemple le lien entre le recrutement d'un auteur et la participation de l'un des éditeurs à un colloque. Le premier document qui nous intéresse ici est le programme d'un colloque de 1986 intitulé « *Des ordinateurs et des mots. Épistémologie de la recherche informatisée* », sur lequel le titre de l'une des communications est entouré.

Cette communication est donnée par Alain Bresson, historien antiquisant qui prépare un doctorat d'État à l'université de Besançon⁶⁵⁴. Il y présente deux programmes informatiques développés au sein de son unité de recherche : ONOMASTIKON et PETRAE. Une lettre lui est envoyée, vraisemblablement au cours de l'année 1987, mais seule sa réponse, datée du 29 décembre 1987, a été conservée dans la pochette susmentionnée. Dans cette lettre, il accepte de soumettre un article pour la rubrique « *Article de fonds* », qui ne sera jamais publié. En dépit de l'absence de publication dans ce cas, on aperçoit le processus qui conduit des activités extra-éditoriales à l'identification d'auteurs potentiels.

Les articles publiés par le bulletin parvenaient parfois en suivant un troisième parcours tout à fait différent, et beaucoup plus rare, puisqu'on ne trouve dans les archives du comité de rédaction qu'un seul cas de ce type. En 1983, Fossier reçoit ainsi une soumission spontanée de Inge Schoups, historienne du droit, dont nous n'avons

⁶⁵⁴ Alain Bresson, « Recherches sur la société Rhodienne (480 avant J.-C.-100 après J.-C.) » (Lille, France: A.N.R.T., 1994).

malheureusement pas pu retrouver l'affiliation institutionnelle. Fossier lui répond en 1984 :

« *c'est en effet votre genre de réaction que nous souhaiterions trouver chez l'ensemble de nos lecteurs* ».

Dans ce cas, c'est donc une lectrice qui contacte directement le comité de rédaction pour proposer un article dans le bulletin, ce qui atteste alors du succès de la publication : le bulletin a touché directement un chercheur, sans qu'aucun membre du comité de rédaction n'ait à s'impliquer personnellement.

Ces trois canaux de contributions nous montrent ainsi que sur la période, le fonctionnement du bulletin repose avant tout sur les activités et l'environnement institutionnel de certains membres du comité éditorial, dont deux sont des membres du service informatique de l'IRHT (Fossier et Bourlet). Ils assurent la rédaction des articles, mais aussi, par les activités associées à l'édition (rencontres, enquêtes, consultations techniques, expertises), rencontrent de futurs auteurs.

1.3.2. Processus de révisions des articles.

Une autre particularité du *Médiéviste et l'ordinateur* consiste à publier des articles quasiment sans aucun processus de révision. On ne trouve en effet dans les archives aucune trace d'aller-retour de différentes versions entre les auteurs et le comité de rédaction. Il apparaît même, à la lumière des archives, que les éditeurs ne refusaient presque jamais de texte. Dans une pochette intitulée « textes en réserve », mentionnée ci-dessus, on retrouve ainsi trois articles, dont il est difficile, pour au moins deux d'entre eux, de savoir pourquoi ils sont restés « en réserve ». Pour le troisième, la raison est plus claire, puisqu'il est annoté, sur la première page, avec la mention « *article non valable* ». Avons-nous là la trace d'une évaluation qui avait cours pour tous les articles, ou est-ce une exception ? Impossible de répondre avec certitude.

Un autre élément qui atteste de la particularité du travail éditorial des membres du comité nous est révélé grâce aux nombreuses lettres accompagnant les contributions. Pour une grande majorité d'entre elles, les auteurs indiquent aux éditeurs qu'ils leur laissent le soin de déterminer s'il faut pratiquer des coupes ou non. Pour prendre seulement un

exemple, retrouvée dans la pochette intitulée « Correspondances relatives aux articles » du carton « Premiers adhérents – courriers – colloques » des archives de l'IRHT à Orléans, Ezio Ornato, spécialiste de codicologie quantitative au sein de l'Équipe de Recherche 52 du CNRS, adresse en octobre 1984 une lettre concernant l'article qu'il co-écrit avec l'informaticien de l'IRHT Denis Muzerelle (1945-2001) pour le treizième numéro du bulletin. Il y indique :

« Nous vous laissons carte blanche pour les coupures si l'article est trop long ».

L'autre carte blanche que les auteurs laissent aux éditeurs vis-à-vis de leurs articles concerne la publication des annexes ou des figures, dont plusieurs lettres attestent dans les archives de l'IRHT. Nous en trouvons un exemple dans une lettre de l'historienne américaine spécialiste de la France médiévale Elizabeth A. R. Brown (1932-), adressée à Fossier en 1984 et retrouvée dans la pochette « Correspondances relatives aux articles mentionnée ci-dessus ». Brown y précise, à propos de son article pour le onzième numéro du printemps 1984, qu'elle lui laisse le choix de publier ou non les exemples d'index produits par les ordinateurs et l'arbre généalogique joints à l'article. Dans ce cas, seuls les exemples des documents de sortie seront publiés, ce qui, nous y reviendrons, permet de saisir l'importance que les éditeurs attachent à la publication des documents de travail dans les pages du bulletin.

Finalement, ce qu'on constate, à la lumière des éléments dégagés ci-dessus, c'est qu'à la différence de ce qui existe aujourd'hui dans les revues scientifiques à comité de lecture : 1) le choix des auteurs et des textes publiés repose sur les décisions des éditeurs, sans que celles-ci soient attachées à des processus de révision formalisés des articles et que 2) la version finale de l'article semble être, dans une certaine mesure, à la discrétion des éditeurs et non des auteurs.

À la lumière des objectifs poursuivis par les éditeurs du bulletin, il faut considérer, selon nous, ces deux éléments comme autant de moyens dont ils disposent pour façonner les modalités de la circulation des contenus dans les pages du bulletin. Rappelons ici qu'il en existe d'autres, que nous avons évoqués plus haut, comme l'organisation thématique de la collection et l'organisation en rubrique de chaque numéro. Ces considérations appellent deux questions, auxquelles nous répondrons dans les deux parties suivantes : 1)

comment les éditeurs ont-ils coordonné la publication des discours produits par les auteurs à travers l'élaboration de thématiques et de rubriques ? et 2) comment orientaient-ils les auteurs pour obtenir des articles qu'ils jugeraient intéressants de publier en accord avec les objectifs affirmés du *Médiéviste et l'ordinateur* ?

2. Coordonner la production textuelle d'une communauté discursive.

Nous avons examiné jusqu'ici l'organisation du bulletin de liaison, en analysant notamment les activités des membres du comité de rédaction, et les logiques qui conduisent à la publication des articles. Ces éléments nous ont permis de montrer comment *Le Médiéviste et l'ordinateur* fonctionnait comme l'un des mécanismes d'intercommunication d'une communauté discursive, dont les membres partageaient l'objectif commun de créer un réseau. Pour poursuivre notre analyse, il nous apparaît nécessaire de nous intéresser non plus à l'organisation du bulletin, mais à son contenu. En effet, ce périodique ne nous offre pas seulement un panorama des membres de ce réseau, de leurs méthodes de recherches, des types de sources qu'ils étudient, de l'équipement qu'ils mobilisent. Il nous fournit également une image des stratégies de coordination mises en place par le comité éditorial pour structurer des axes de discussions communs.

Nous regarderons comment les éditeurs coordonnent, par la définition de « thème » propre à chaque numéro, la publication d'articles présentant des projets comparables, que ce soit au regard de leurs méthodes d'enregistrement ou de traitement des données, du type de sources qu'ils utilisent ou du type d'instruments de travail qu'ils cherchent à élaborer. Comme nous le verrons, cette étude des « thèmes » présente

cependant des limites, que nous tenterons de dépasser en proposant une analyse automatique des relations de cooccurrences des termes les plus fréquents de ce corpus.

2.1. Diversité des collectifs engagés et élaboration de thèmes communs.

2.1.1. 70 collectifs et autant de cultures locales.

Le Médiéviste et l'ordinateur constitue un espace d'échanges pour un grand nombre de chercheurs, qui travaillent au sein de collectifs développant des cultures de travail locales extrêmement disparates. Comme nous l'avons déjà mentionné plus haut, le bulletin donne la parole, pendant ces dix premières années, à 136 auteurs, dont 128 ne sont pas membres du comité éditorial. Ces 128 auteurs ne sont d'ailleurs pas tous des médiévistes : ils travaillent dans des disciplines différentes, comme l'histoire, la linguistique, l'informatique, la musicologie, l'archéologie, la sociologie, les mathématiques. La plupart d'entre eux s'inscrivent même aux frontières entre plusieurs d'entre elles. Certains d'entre eux assument des fonctions de professeur des universités, d'autres de chercheurs indépendants, mais on rencontre également parmi les auteurs des ingénieurs de recherche ou des techniciens exerçant dans des centres de calcul.

Les institutions dans lesquelles ces auteurs sont accueillis sont elles aussi très diversifiées. Parmi les auteurs, nombreux sont ceux qui sont employés au sein d'équipes de recherche affiliées à de grandes institutions, comme le CNRS. Pour les chercheurs français, on retrouve d'ailleurs de nombreuses petites équipes de la forme de celle étudiée au chapitre 2, organisées sous la forme d'Équipes de Recherche Associée, ou de collectifs constitués pour la mise en œuvre d'un projet particulier, comme celui de la Recherche Coopérative sur Programme analysée au chapitre 1. Il faut ajouter à celles-ci des structures de recherche plus grande, comme l'IRHT ou le CRAL, qui sont composées de plusieurs équipes et dont nous avons présenté certains caractères au chapitre 3.

En dehors du contexte français, les projets présentés par des historiens d'autres pays européens sont en majorité le fruit de collectifs de recherche organisés autour d'un projet spécifiquement informatique, comme le *Lessico Intelletuale Europeo*, mené par une équipe à l'université *La Sapienza* de Rome. En Allemagne, d'autres logiques institutionnelles président à la formation de ce type de collectifs. C'est par exemple le cas

du groupe mené par Joachim Wollasch (1931-2015) sur les obituaires et les nécrologies, organisé sous la forme d'un centre de recherches collaboratives, le *Sonderforschungsbereich 7*. Il est d'ailleurs intéressant de noter que la majorité des articles qui présentent les projets de ce type de structure sont signés collectivement. En tout, ce sont près de 70 collectifs de recherches qui peuvent être ainsi repérées pour la période qui nous intéressent dans ce chapitre.

Comme nous l'avons montré dans les deux premiers chapitres, les recherches de ces équipes s'inscrivent aussi dans des environnements matériels spécifiques, dont dépendent, entre autres, le type de financement obtenus, le matériel informatique à disposition et l'autonomie vis-à-vis des informaticiens. On rencontre dans les pages du bulletin des configurations matérielles très différentes, qui vont des analyses factorielles faites à la calculatrice programmable de poche du médiéviste du CNRS Alain Guerreau (1948-)⁶⁵⁵, que nous étudierons au chapitre 6, au programme ALINE de l'ERA 713 écrit pour fonctionner sur l'un des IBM 360 du CIRCE⁶⁵⁶, en passant par le travail d'indexation automatique pratiquée par les équipes du CRAL sur les ordinateurs du centre de calcul de Nancy⁶⁵⁷ ou le travail sur un micro-ordinateur Apple présenté par l'historien Jean-Claude Cheynet (1947-)⁶⁵⁸.

À ces disparités institutionnelles et matérielles, il faut encore ajouter celles, révélées dans les trois premiers chapitres, qui tiennent aux types de sources étudiées, aux objectifs de recherche poursuivis, aux relations interdisciplinaires qu'entretenaient les acteurs de ces équipes avec d'autres chercheurs et à la finalité des travaux entrepris, qui pouvaient viser, comme nous l'avons vu, la constitution d'instruments de travail génériques ou spécifiques. À la lumière de ces nombreuses variables, on conçoit donc volontiers que les 70 équipes de recherches qui présentent leur travail dans *Le Médiéviste et l'ordinateur* entre 1979 et 1989 rendent compte d'au moins autant de cultures locales d'utilisation des ordinateurs.

⁶⁵⁵ Alain Guerreau, « Analyse factorielle au moyen d'une calculatrice de poche programmable », *Le médiéviste et l'ordinateur* 2, n° 1 (1979): 10-12.

⁶⁵⁶ Jean-Philippe Genet, « Une application de l'analyse factorielle à l'étude du vocabulaire », *Le médiéviste et l'ordinateur* 5, n° 1 (1981): 11-15.

⁶⁵⁷ Caroline Bourlet et Michel Parisse, ... « ... En direct de Nancy : l'automate et les chartes », *Le médiéviste et l'ordinateur* 2, n° 1 (1979): 8-10.

⁶⁵⁸ Jean-Claude Cheynet, « À la maison : c'est Byzance ! », *Le médiéviste et l'ordinateur* 9, n° 1 (1983): 7-8.

2.1.2. Fonctions de l'organisation thématique.

Partant de ce constat, il nous appartient de révéler les stratégies éditoriales mises en place par les éditeurs pour coordonner les discussions et les échanges entre ces collectifs. L'une de ces stratégies, qui nous intéressera ici, consiste à construire des « thèmes » pour chaque numéro. Cette structuration générale du périodique permet, selon les mots des éditeurs, de publier ensemble des « *projets analogues* »⁶⁵⁹, et donc, d'établir des ponts entre les travaux de collectifs ayant développé des cultures de travail locales différentes. Le tableau suivant regroupe, en quatre catégories, les thèmes des dix-neuf premiers numéros du bulletin (le numéro 20 étant un numéro spécial, et sans thème).

Figure 19. Typologie des numéros du *Médiéviste et l'ordinateur* (1979-1989).

Catégories	Nombre de numéros	Thèmes des numéros
<i>Outils</i>	5	Traitement de texte, analyse factorielle des correspondances, classification automatique, système experts, onomastique et indexation documentaire
<i>Méthodes</i>	4	Prosopographie, édition électronique, recherche documentaire, thesaurus et lexique
<i>Types de sources</i>	6	Chartes, sources sérielles (x2), testament, le livre médiéval, cadastres
<i>Équipements</i>	4	Lecteur optique, micro-informatique, micro et lourde, entrées-sorties

La classification que nous proposons dans ce tableau fait apparaître différentes logiques d'organisations des numéros. On remarque d'abord que près d'un quart des livraisons regroupent des articles qui portent sur ce que Fossier appelle des « *instruments méthodologiques* »⁶⁶⁰. À travers ce type de thématiques, les éditeurs entendent focaliser les discussions autour des outils informatiques ou statistiques apparus au contact des

⁶⁵⁹ Bourlet et al., « Éditorial », 2.

⁶⁶⁰ Lucie Fossier, « Éditorial », *Le médiéviste et l'ordinateur* 20, n° 1 (1988): 2.

ordinateurs, comme l'analyse factorielle des correspondances, la classification automatique, le traitement de texte ou les systèmes experts. Ces instruments peuvent encore être présentés en relation avec des domaines de l'étude historique, comme c'est le cas pour le numéro intitulé « Onomastique et indexation documentaire ».

La deuxième catégorie regroupe des thématiques qui portent sur des méthodes déjà anciennes en histoire, comme la prosopographie ou la constitution de thésaurus pour l'étude du lexique, ou sur les transformations qu'impliquent l'utilisation des ordinateurs dans certains pans de l'activité des médiévistes, comme l'édition ou la recherche documentaire.

Ces deux catégories de livraisons ont plusieurs objectifs. Il s'agit tout d'abord de présenter aux lecteurs diverses formes d'applications de ces outils ou de ces méthodes assistées par ordinateur. L'enjeu consiste aussi, pour les éditeurs du *Médiéviste et l'ordinateur*, à faire dialoguer les différents utilisateurs de ces outils ou de ces méthodes par articles interposés. Le bulletin permet alors aux usagers 1) de découvrir, dans un même numéro, d'autres applications d'un outil ou d'une méthode, d'autres programmes informatiques permettant leurs mises en œuvre et 2) de comparer leurs utilisations de ces outils ou leurs méthodes avec celles de collègues travaillant sur d'autres ensembles documentaires, qui poursuivent d'autres objectifs et travaillent dans d'autres environnements matériels.

Les numéros consacrés à des types de sources permettent d'autres formes de convergence entre les pratiques des usagers. Nous l'avons vu dans le chapitre 3, les discussions entamées au cours du colloque de Rome de 1975 s'articulent autour de l'examen des caractères des genres d'ensembles documentaires, dans la mesure où ils appelaient des traitements pour lesquels les ordinateurs jouaient des rôles différents. Ces ensembles différaient le plus souvent en raison de leurs logiques d'archivage, de l'homogénéité des documents qui les composaient ou du caractère répétitif de leur contenu. En consacrant des numéros thématiques à des genres d'ensembles documentaires, les éditeurs structurent les échanges autour d'autres points communs que les outils ou les méthodes.

Il faut ici insister sur le fait que si ces documents sont traités ensemble dans les numéros du bulletin, c'est du fait qu'ils sont comparables au regard des prises qu'ils offrent à un traitement informatique. L'une des logiques qui se dégage à la lecture de ces

numéros repose notamment sur la similitude entre la structure des informations qu'elles contiennent. À ce titre, le choix fait par les éditeurs de proposer deux numéros distincts sur les « sources sérielles » est très éclairant.

Le premier de ces numéros regroupe des études portant sur des sources comme des actes notariés, des livres de comptes ou des cadastres, c'est-à-dire des documents qui présentent une structure de l'information répétitive, mais complexe. Une distinction y est faite entre, d'une part, les corpus constitués de matériaux dits homogènes, qui sont définis comme des corpus composés sur la base d'une source unique et, d'autre part, les corpus constitués de matériaux dits hétérogènes, définis comme ceux pour lesquels l'historien a lui-même fait le choix de rassembler des sources pour mener son étude, en dépit du fait qu'elles pouvaient avoir été produites par différents acteurs à différentes périodes.

Le second numéro regroupe, quant à lui, uniquement des études portant sur des collections de listes, c'est-à-dire, des documents pour lesquels les textes présentent directement une information structurée de manière répétitive. Les ensembles documentaires sont en l'espèce les *Tratte fiorentini* (liste de candidats potentiels aux offices communaux), les registres de baptême (listes des baptêmes au sein d'une paroisse), et les nécrologes du bas Moyen-Âge (listes de décès). Pour chacun de ces deux types de corpus est présenté une et une seule technique de méthode de traitement informatique : l'analyse documentaire pour les sources à structure complexe, et le couplage automatique de données pour les listes.

Les six numéros traitant de types de sources exposent ainsi des liens privilégiés entre les outils informatiques et les structures des corpus étudiés. Pour chaque type de sources, on retrouve dans les numéros un ou deux modes de traitement informatique privilégiés, que nous avons présentés au chapitre précédent : enregistrement intégral pour les textes dits « littéraires », analyse documentaire pour les textes dits « documentaires » à structure complexe, couplage automatique de données pour les listes ou encore constitution de base de données et échantillonnage pour les collections de livres médiévaux. Les éditeurs, en choisissant de structurer les discussions autour de types de sources, permettent donc aux lecteurs d'identifier des méthodes adaptées aux différents genres d'ensembles documentaires, et aux usagers de saisir les déclinaisons de l'application des outils et des méthodes à la lumière des conditions matérielles et des particularités des sources.

2.1.3. Évolutions des équipements : micro-informatique, lecteurs optiques, disquettes.

Un dernier ensemble de numéros montre un autre type de discussions que les éditeurs entendent développer par le biais du bulletin. Elles concernent les problématiques liées aux équipements informatiques, qui évoluent très rapidement dans les années 1980. C'est dans cette catégorie que s'inscrivent les deux livraisons consacrées directement aux usages du micro-ordinateur dans les pratiques de recherches (« La micro-informatique à l'ordre du jour » au printemps 1983 et « Micro-informatique et informatique lourde » à l'automne 1985).

Ce nouveau type d'ordinateurs, résultat des progrès de la miniaturisation des circuits intégrés, s'était massivement diffusé à partir de la fin des années 1970⁶⁶¹. Au sein de la discipline historique en France, les premières rencontres consacrées à leurs utilisations n'apparaissent pourtant qu'en 1983, et les premiers témoignages des médiévistes utilisant ces technologies récentes sont à rechercher dans le premier des deux numéros cités ci-dessus. On y découvre que ces nouvelles machines bouleverse en profondeur les modalités de l'utilisation des ordinateurs en histoire médiévale, en ouvrant de nouvelles possibilités pour les chercheurs : travail chez soi, parfois connecté aux ordinateurs plus puissants d'un centre de calcul ; conservation et échanges des données sur des disquettes ou des *céderom* - même s'il faut préciser que les capacités de stockage de ces nouveaux supports ne permettent alors essentiellement d'échanger que des fichiers très légers, comme des articles, des bibliographies ou de petits programmes - et coût de fonctionnement quasi-nul de cette utilisation, par rapport aux tarifs proposés dans les centres de calcul⁶⁶².

Les médiévistes qui s'en emparent font particulièrement remarquer l'intérêt que présentent les micro-ordinateurs pour les activités de saisie des données, étape très chronophage, et pour la possibilité qu'ils offrent de multiplier de façon plus souple les opérations d'analyses de ces données à l'aide de programmes simples. Ils remarquent

⁶⁶¹ Pour un article de cette époque qui présente le développement de ces technologies voir par exemple Laurent De Mautort, « L'industrie des micro-ordinateurs (les principales lignes de force) », *Revue d'économie industrielle* 38, n° 1 (1986): 64-73.

⁶⁶² Au début du numéro 9 du *Médiéviste et l'ordinateur*, dédié à la micro-informatique, un premier article écrit par un informaticien de l'IRHT reprend synthétiquement les avantages et les inconvénients de ce nouveau type d'ordinateurs : Jean-Luc Minel, « Le micro-ordinateur est-il une panacée ? », *Le médiéviste et l'ordinateur* 9, n° 1 (1983): 2.

également l'intérêt particulier que présentent ces nouveaux outils pour certains types d'approches, comme l'édition électronique ou la prosopographie.

Ces auteurs soulignent toutefois que l'utilisation de ces technologies soulevaient trois types de difficultés : 1) l'investissement financier nécessaire pour l'achat du micro-ordinateur lui-même (entre 30 000 francs et 40 000 francs en moyenne en 1984, soit l'équivalent de 4 000 - 5 000 euros), mais également pour celui des périphériques (un disque dur de 10 méga-octets coûtait alors environ 30 000 francs, soit environ 4 800 euros, un lecteur de disquette 4 500 francs, soit environ 670 euros) et des logiciels (1 166 francs, par exemple, pour le système d'exploitation MS DOS d'un IBM-PC, soit environ 180 euros)⁶⁶³ ; 2) les connaissances techniques requises, en particulier pour la programmation et 3) les modalités de l'articulation de ces nouveaux instruments avec l'informatique dite « lourde », que rendaient presque indispensable les faibles capacités de mémoire et de calcul des modèles de micro-ordinateurs disponibles à cette époque⁶⁶⁴.

Pour guider les lecteurs à travers ces difficultés, et lui permettre de faire « *le bon choix* »⁶⁶⁵, ces deux numéros donnent la parole à des universitaires, pionniers dans l'achat de ces nouvelles machines, à des amateurs passionnés, membres de club de micro-informatiques, ou à des ingénieurs de recherches, travaillant dans des centres de calcul mettant à disposition ce type de machines, comme c'est le cas au Laboratoire d'Informatique pour les Sciences de l'Homme (LISH) par exemple⁶⁶⁶.

Le même type de numéros est constitué autour du lecteur optique, avec pour objectif de présenter une avancée technologique très attendue par certains médiévistes, puisque cet équipement devait leur permettre d'enregistrer des textes au format numérique à un rythme incomparable avec celui de la saisie à la main⁶⁶⁷. Comme pour

⁶⁶³ Ces valeurs sont tirées de l'analyse de deux fiches techniques publiées en 1984 dans le numéro 7 de la revue commerciale *Votre ordinateur. Le magazine de l'informatique à la maison*. Ces fiches présentent les caractéristiques et les prix de deux ordinateurs - un IBM PC et un Apple, commercialisés respectivement en 1981 et 1984 - et de leurs périphériques. Les conversions des francs à l'euro ont été effectuées à l'aide de l'un des outils du site fxtop.com

⁶⁶⁴ Minel, « Le micro-ordinateur est-il une panacée ? ».

⁶⁶⁵ Cette expression est empruntée aux éditeurs eux-mêmes, qui intitulent ainsi l'une des rubriques du numéro, composée de trois articles qui visent à éclairer les différents environnements de travail dans lesquels il est possible d'utiliser des micro-ordinateurs (dans un centre de calcul, dans un laboratoire ou à domicile).

⁶⁶⁶ Michael Hainsworth, « Dans un centre de calcul : le LISH », *Le médiéviste et l'ordinateur* 9, n° 1 (1983): 3-5.

⁶⁶⁷ Dans l'éditorial du numéro 12 publié à l'automne 1984, Fossier annonce en note que le LISH vient de faire l'acquisition d'un modèle de lecteur optique Kurzweil, très attendu par la communauté. Elle commente

les deux numéros sur la micro-informatique, cette livraison était construite à partir de témoignages d'utilisateurs pionniers, qui y partageaient leurs expériences et tentaient d'apporter à la communauté des points de vue sur l'intérêt et les limites de ces nouveaux périphériques.

À l'automne 1984, un numéro recoupant ces trois préoccupations est publié. Il est consacré à ce que les acteurs qualifient de « Problème des « entrées-sorties » », c'est-à-dire aux interfaces entre le chercheur et la machine. Dans l'éditorial de ce numéro, Fossier insiste notamment sur la désuétude des cartes perforées, du fait de l'apparition de nouvelles formes de saisie magnétique directement sur supports numériques, comme les cassettes, les disquettes ou les disques durs des micro-ordinateurs. Ces innovations techniques opéraient, selon elle, comme une « *révolution complète* »⁶⁶⁸ des pratiques de saisies des données, par contraste avec celles mises en œuvre par certains médiévistes jusqu'au début des années 1980.

Cette révolution touchait encore la matérialité des documents de sorties résultant de l'application des programmes, puisqu'elle substituait à l'impression de ces résultats sur papier, sous forme de listing, un terminal vidéo permettant de les visualiser. Certains types de logiciels exploitaient d'ailleurs, depuis le début des années 1970, toutes les possibilités de cette innovation en proposant à l'utilisateur de voir directement à l'écran les résultats de l'éventuelle impression, ce qui bouleversait profondément certains domaines d'activités des médiévistes comme l'édition ou la publication⁶⁶⁹. Dans l'éditorial du numéro 12, Fossier insiste encore sur l'évolution des périphériques d'impression au début des années 1980, qui permettent d'obtenir des documents de sorties moins « *frustres* » et moins « *rébarbatifs* » que ceux obtenus à l'aide des imprimantes de la génération précédente⁶⁷⁰.

On constate, sur la base de cette brève analyse, que les numéros thématiques de cette catégorie répondent à d'autres logiques que celles des catégories précédentes. Ils ne visent en effet pas à susciter des échanges d'auteurs à auteurs, ou à tenter d'illustrer des

cette actualité en ces termes : « *C'est tout de même un grand évènement* ». Lucie Fossier, « De la carte perforée au texte numérisé », *Le médiéviste et l'ordinateur* 12, n° 1 (1984): 2.

⁶⁶⁸ Fossier, 1.

⁶⁶⁹ Ce type de logiciel a notamment été qualifié à l'aide de l'expression « What you see is what you get » (WYSIWYG). Dans notre corpus, on trouve par exemple reproductions de capture d'écrans dans les articles à partir de l'année 1986. Voir par exemple Françoise Micheau et Jean-Pierre Molénat, « Création d'alphabets sur Macintosh », *Le médiéviste et l'ordinateur* 16, n° 1 (1986): 13-18.

⁶⁷⁰ Fossier, « De la carte perforée au texte numérisé », 1.

tendances générales dans le type de traitement associé à certains genres d'ensembles documentaires, mais au contraire à informer les lecteurs sur les possibilités et les limites de ces nouveautés techniques. À l'échelle de la période que nous étudions ici (1979-1989), on remarque que ces évolutions techniques, et en particulier le développement massif de la micro-informatique, modifient fondamentalement les besoins des utilisateurs, et donc, la situation de communication dans laquelle s'inscrit le bulletin de liaison. Il nous est cependant difficile d'identifier avec certitude les effets de ce changement sur les stratégies éditoriales des éditeurs. Nous verrons toutefois, en conclusion, qu'il est possible que ce phénomène ait conduit à une réorganisation de cette communauté autour d'initiatives d'autres acteurs.

L'ensemble des « thèmes » que nous avons analysés jusqu'ici donne une image claire des choix faits par les éditeurs pour structurer les discussions de cette communauté discursive. Ils illustrent aussi les stratégies mises en place pour faire dialoguer des équipes de recherches aux objectifs, aux méthodes, aux sources et aux environnements de travail tous différents. Cependant, cette approche par les « thèmes » ne saurait donner une image pleinement satisfaisante de la teneur des contenus échangés dans les pages du bulletin. En effet, si les « thèmes » structurent les numéros, ils ne concernent que les seuls « articles de fonds » et les « notes techniques », et non les autres rubriques, qui sont pourtant elles aussi essentielles dans les échanges entre les acteurs. Pour saisir plus complètement la teneur de ces échanges, mais aussi pour mesurer l'impact de cette stratégie éditoriale, une analyse complémentaire paraît nécessaire.

2.2. Cartographie des sujets abordés dans le bulletin (1979-1989).

Dans le but de représenter plus largement la teneur des échanges qui occupent les quelques 600 pages du *Médiéviste et l'ordinateur* entre 1979 et 1989, nous aurons ici recours à un outil de fouille de texte appelé CorText. Développé par l'Unité de Recherche 1326 du CNRS INRA SenS entre 2010 et 2014, puis par le laboratoire LISIS depuis 2015, cet outil a notamment fait ses preuves pour extraire les principaux « sujets » de corpus

bien plus vastes que le nôtre⁶⁷¹. Cette approche, pour laquelle nous avons optée avec l'aide des ingénieurs de recherches du laboratoire LATTICE⁶⁷², permet de réaliser des graphes de cooccurrences. Elle nous permettra de donner une représentation graphique du corpus du *Médiéviste et l'ordinateur* pour la période 1979-1989, éclairée non à la lumière des « thèmes » des numéros ou de l'organisation sous forme de rubriques, mais de l'ensemble de son vocabulaire.

2.2.1. Graphe de cooccurrences des termes les plus fréquents du corpus.

Le graphe que nous présenterons ici propose une représentation des mots les plus fréquents du corpus sous la forme de triangles aux dimensions variables en fonction du nombre d'occurrences de ces termes. Les relations de cooccurrences entre ces termes, qui sont établies sur la base de la présence simultanée de ces mots dans le même environnement textuel, y sont illustrées sous la forme d'une liaison entre ces triangles. L'élaboration de ce type de graphe procède d'un traitement du corpus en plusieurs étapes qu'il nous faut ici rappeler. Après un nettoyage et une annotation sommaire des articles du corpus (début et fin d'article, rubrique), CorText permet l'extraction des 1 000 mots non grammaticaux les plus fréquents. Il faut ici préciser que l'absence de lemmatisation de notre corpus entraîne, de fait, l'absence de verbes dans cette liste. L'étape suivante consiste à nettoyer la liste à la main, c'est-à-dire dans notre cas, à la débarrasser des noms des acteurs et des institutions, afin de l'indexer sur l'ensemble du corpus. Cette dernière décision était motivée par la volonté de nous intéresser aux termes employés par les auteurs et non à la distribution des mentions des acteurs dans le corpus.

Une fois ces opérations de préparation des données effectuées, il est alors possible de demander au logiciel de produire un graphe qui cartographie les relations de cooccurrences entre les 882 noms communs restants. Par ce dernier procédé, l'outil CorText révèle, sous la forme de nuages colorés de triangles-mots, l'existence

⁶⁷¹ Pablo Ruiz et Thierry Poibeau, « Mapping the Bentham Corpus: Concept-based Navigation », *Journal of Data Mining and Digital Humanities* Special Issue: Digital Humanities between knowledge and know-how (Atelier Digit_Hum) (mars 2019).

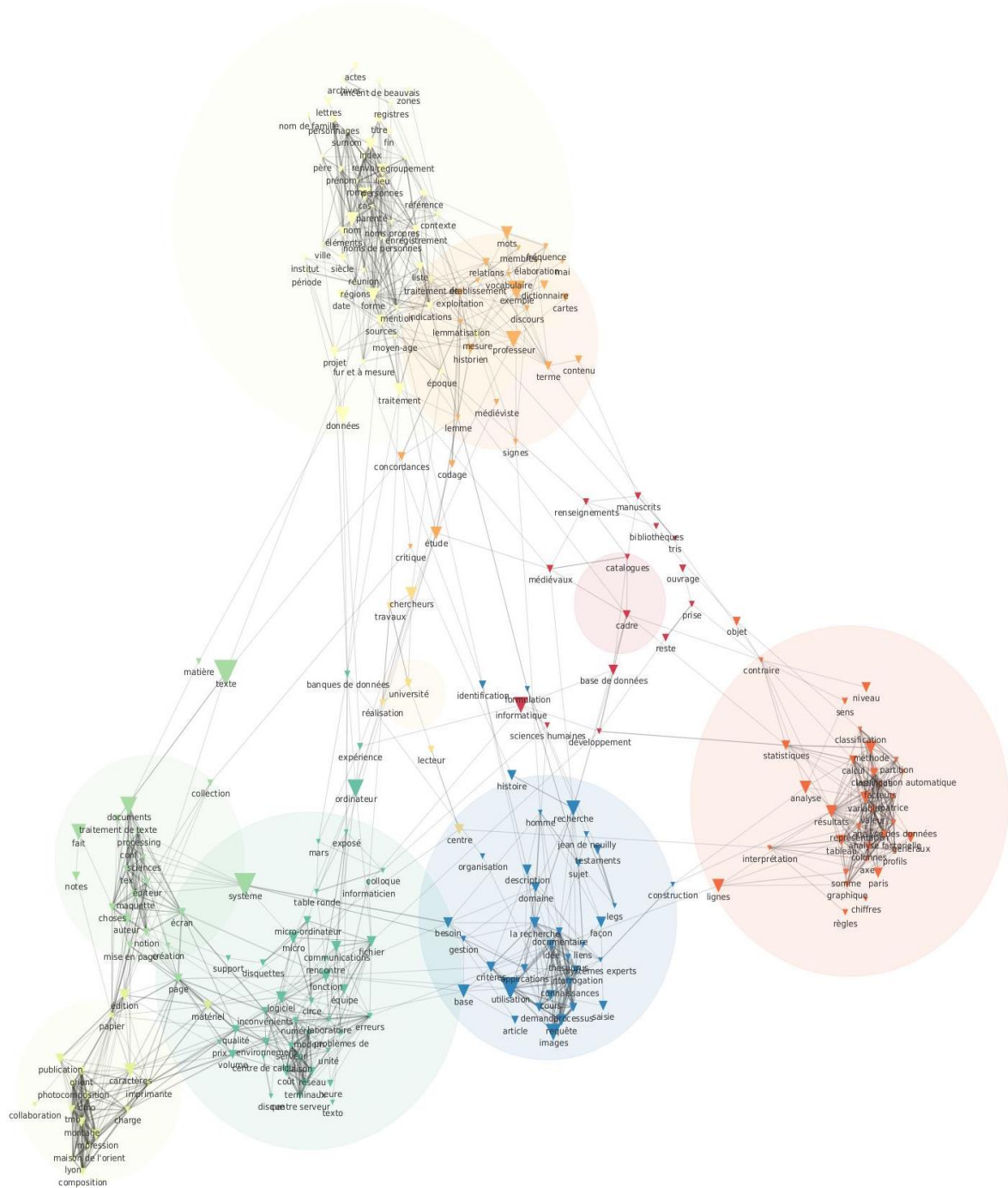
⁶⁷² J'aimerais ici remercier chaleureusement Frédérique Mélanie et Clément Plancq, ingénieurs de recherche au laboratoire LATTICE, qui m'ont, entre autres choses, présenté cet outil et guidé dans son utilisation avec patience et disponibilité. Le travail de ce chapitre ne saurait être apprécié à sa juste valeur sans prendre en compte leur aide précieuse.

d'ensembles de mots fortement co-occurents les uns avec les autres. Dans ce processus, plusieurs éléments doivent être paramétrés par le chercheur, et notamment les modalités de définition de l'environnement textuel, qui repose sur la mesure de la proximité lexicale⁶⁷³. La principale valeur d'ajustement est notamment le nombre de nœuds, c'est-à-dire le nombre de liens de cooccurences entre des termes retenus pour la représentation. Après divers essais, il a été retenu que le graphe présentant 250 nœuds était à la fois assez précis et assez lisible⁶⁷⁴ :

⁶⁷³ Pour une présentation des possibilités offertes par CorText sur ce point, nous renvoyons à la page 28 du manuel d'utilisation en français : Frédérique Mélanie, Pablo Ruiz, « Comment utiliser CorText », janvier 2017 disponible ici : <https://docs.cortext.net/wp-content/uploads/CommentUtiliserCortext.pdf> Pour la réalisation de ce graphe, nous avons choisi pour le seuil de proximité la valeur conventionnelle en statistique textuelle : 0,1.

⁶⁷⁴ En faisant varier la valeur du nombre de nœud entre 80 et 500, il a été possible d'établir que le nombre de nuages variait entre sept et neuf. Au-dessus de 300 noeuds, les neufs nuages tendaient à voir leurs liens se renforcer.

Figure 20. Graphe de cooccurrences des 882 noms communs les plus fréquents du corpus.



Une première lecture du graphe permet d'identifier clairement sept ensembles interconnectés de termes fortement co-occurents. On remarquera, dans la suite de l'analyse, que ces ensembles de termes, que nous qualifierons de « sujets » pour plus de

commodité, se confondent parfois avec des « thèmes » de numéro, mais que le graphe permet toujours d'en donner une image plus fine, mais aussi plus proche du texte.

Un premier sujet se dégage en bleu-vert, en bas sur la gauche du graphique, avec des termes comme « *micro-ordinateur* », « *logiciel* », « *centre de calcul* », « *environnement* », « *support* », « *terminal* », « *disquettes* », « *serveur* », « *réseau* », ... mais aussi un deuxième groupe de mots comme « *inconvenients* », « *problèmes de modem* », « *prix* », « *coût* » ou « *erreurs* ». Dans le corpus, les questions relatives aux infrastructures comme à la matérialité de l'informatique se retrouvent donc, en accord avec nos analyses précédentes, associées aux difficultés que rencontrent les acteurs dans la mise en œuvre des technologies informatiques.

Un troisième type de termes, plus surprenant pour nous, se retrouve dans cet ensemble. Il renvoie à la place des modalités du travail collectif dans les discours sur ces questions : « *équipe* », « *rencontres* », « *communications* », « *laboratoire* », « *informaticien* », « *colloque* », etc. Le graphique permet donc d'identifier ici un premier pôle de discussion, qui articule trois champs sémantiques différents et qui porte sur les relations entre la matérialité des opérations, les difficultés qui lui sont associées et les organisations collectives qui permettent de s'en accommoder. Ce premier sujet recoupe fortement les thèmes des numéros que nous avons regroupés plus haut sous la catégorie « *Équipement* », et englobe également des considérations qui se retrouvent dans les textes réunis dans les rubriques « *Information* » et « *Voyage au centre de calcul* ».

En élargissant la focale, on note que ce premier nuage bleu-vert présente des relations fortes avec deux autres nuages, qui touchent à deux aspects distincts de la production de textes : l'édition électronique, en vert à l'extrême gauche du graphique (« *mise en page* », « *maquette* », « *notes* », « *écran* », « *TeX* », etc.) et la publication assistée par ordinateur, en jaune-vert, tout en bas à gauche sur le graphique (« *impression* », « *composition* », « *photocomposition* », « *publication* », « *imprimante* », etc.). Ces deux nuages de relations de cooccurrences rendent compte de la publication, dans deux numéros distincts, de deux ensembles de textes qui traitent de ces questions.

Le numéro 16 tout d'abord, publié au printemps 1986, nous l'avons vu, est entièrement dédié à l'édition électronique. Il existe par ailleurs, dans le numéro 14 du printemps 1985, un compte-rendu d'une visite chez un éditeur, la Maison de l'Orient, qui

entreprend à cette période d'informatiser ses processus de publication. Dans ce second cas, on constate qu'un sujet apparaît distinctement sur le graphique sans pour autant qu'un numéro y soit consacré. On constate donc la fécondité du travail sur les cooccurrences puisqu'il permet de dégager l'existence d'une thématique sur laquelle aucune livraison du *Médiéviste et l'ordinateur* ne porte explicitement.

Nous remarquons, en bleu, un quatrième nuage, à droite de ce premier ensemble tricéphale. Il regroupe des termes comme « *utilisation* », « *systèmes experts* », « *saisie* », « *requêtes* », « *testaments* », « *base* », « *critères* » ou « *domaine* ». Nous pouvons l'interpréter comme renvoyant à l'ensemble des problématiques liées à l'indexation et à la recherche documentaires. Ce nuage matérialise donc les discussions qui prennent forme autour de cette branche de l'informatique qui s'intéresse aux modalités de l'accès au contenu des sources et des documents de la façon la plus efficace possible. À ce titre, il n'est pas étonnant de retrouver, au centre de cet ensemble, des mots aux occurrences extrêmement nombreuses comme « *utilisation* », « *recherche* » ou « *requête* », termes caractéristiques de cette approche documentaire, ou comme « *thesaurus* », instrument de travail spécifique à ces approches, comme nous l'avons vu dans le chapitre 3.

On distingue, sur la droite du graphe, en orange, un cinquième nuage clairement identifiable. Extrêmement cohérent, il regroupe les termes qui renvoient à l'analyse de données (« *axe* », « *profil* », « *variable* », « *matrice* », « *colonnes* », « *lignes* »), et, en particulier, à l'utilisation de l'informatique pour la mise en œuvre de deux procédures : les analyses factorielles et la classification automatique. Au centre de ce nuage, les termes les plus fréquents sont « *analyse* », « *résultats* » et « *tableau* », technologie intellectuelle au cœur du fonctionnement de ces méthodes (auquel il faut ajouter, quoique présentant moins d'occurrences, le terme « *graphique* »).

De plus, on remarque que la proximité de ce nuage avec celui touchant à l'indexation et la recherche documentaire décrit ci-dessus s'opère via le terme « *lignes* ». Or, si ce terme est commun à ces deux contextes, il est nécessaire d'insister sur le fait qu'il renvoie en fait à deux types de technologies intellectuelles distinctes, propres à ces différentes pratiques. En effet, les discussions dans le corpus de texte relatifs à l'analyse de données portent sur des tableaux, dont la construction des lignes s'articule avec celle de colonnes. Au contraire, les pratiques d'indexation documentaire supposent la constitution de liste de termes et de listes de descripteurs, ce pour quoi l'ensemble de

termes qui figure dans ce nuage présente des relations de cooccurrences plus fortes avec le terme « *ligne* ».

Le graphe présente encore deux autres nuages, que nous distinguons en jaune et en orange clair en haut du graphique. Le premier d'entre eux nous apparaît lui aussi très cohérent. Il se caractérise par un vaste ensemble de termes qui renvoie à un domaine que nous avons présenté dans le chapitre précédent : l'onomastique (« *personnes* », « *nom* », « *prénom* », « *surnom* », « *parenté* », « *ville* », « *région* »). Ce premier ensemble de termes s'articule avec un autre représenté dans ce même nuage, qui touche aux instruments de travail privilégiés pour mener l'étude de problématiques onomastiques (« *index* », « *référence* », « *liste* », « *regroupement* », « *enregistrement* »). Ce sujet rejoint ainsi l'un des domaines largement discuté au cours du colloque de Rome de 1975, et il n'est pas anodin de retrouver parmi les termes de ce nuage le mot « *Rome* », qui semble indiquer la référence directe faite par les auteurs à cet évènement.

Le second nuage qui apparaît en haut du graphique est étroitement lié au précédent. Il regroupe, quant à lui, des termes renvoyant aux approches des sources inspirées des méthodes de la linguistique quantitative, dont nous avons montré l'intérêt qu'elles suscitaient chez certains médiévistes dans le chapitre 3. On y retrouve des termes comme « *mots* », « *lemmatisation* », « *fréquence* », « *vocabulaire* », « *terme* », « *discours* », caractéristiques des pratiques d'enregistrement et de traitement des données textuelles. On retrouve de plus, là aussi, certains des instruments de travail ayant fait l'objet de discussions au colloque de Rome, et notamment les « *concordances* » et les « *dictionnaires* ». Pour ce sujet, il est sans doute significatif que le nuage englobe, parmi les termes les plus fréquents, le mot « *exemple* », dans la mesure où ce terme atteste de la conduite des raisonnements corrélés à ces approches. Comme nous l'avons notamment analysé dans le chapitre 2, l'utilisation des statistiques lexicales suppose en effet un ensemble d'opérations d'aller-retours entre l'analyse des indices de fréquences et l'étude d'exemples d'utilisations des mots en contexte.

Le nuage de termes présente toutefois une anomalie. Pourquoi y trouve-t-on le terme « *professeur* » ? On peut l'expliquer en faisant remarquer que ce type d'approches est plus spécifiquement pratiqué, dans notre contexte, par les membres de l'Équipe de Recherche Associée 713 (ERA 713), constitué essentiellement de chercheurs en poste à

l'université Paris 1 Panthéon-Sorbonne, autour, notamment, de Genet. Nous reviendrons sur ce point dans la suite de ce chapitre.

2.2.2. Rubriques, thèmes, sujets.

Pour nous résumer, l'analyse du graphe montre que sept sujets principaux se démarquent dans les pages du bulletin entre 1979 et 1989. Il faut cependant ajouter à cet ensemble deux autres nuages dont nous n'avons jusque-là pas parlé. Le premier, en rouge, au milieu à droite sur le graphe, concerne les discussions relatives à la fabrication et la consultation de « catalogues » donnant accès à des « manuscrits » ou des « archives ». Le second et dernier nuage, en jaune au centre du graphique, ne se compose que de six termes (« centre », « lecteur », « réalisation », « université », « travaux », « chercheurs »), ne peut, selon nous, être considéré comme pertinent pour établir l'existence d'un sujet au sein du corpus, puisqu'il concerne des objets transversaux à l'ensemble des pratiques.

Vu cette fois du point de vue des termes employés par les acteurs, voici comment nous pouvons cartographier les sujets abordés dans *Le Médiéviste et l'ordinateur* :

Figure 21. Liste des sujets du *Médiéviste et l'ordinateur* établis sur la base du graphe de cooccurrences (1979-1989).

Édition électronique
Publication assistée par ordinateur
Analyse de données
Statistiques lexicales
Indexation documentaire
Catalogues et bibliothèques
Équipement et organisations collectives
Onomastique

Cette analyse a permis d'affiner les catégories dégagées à partir des « thèmes » des numéros dans la sous-partie précédente. Insistons sur le fait qu'elle a porté sur

l'ensemble des rubriques du bulletin, et pas seulement sur les « *Articles de fonds* » qui sont regroupés selon la logique thématique analysée plus haut. En comparant la liste de ces sujets avec celle des « thèmes » présentée au 2.1., plusieurs remarques doivent être développées. Tout d'abord, on constate que si certains « sujets » ne sont traités que dans un seul numéro thématique, comme c'est le cas pour l'édition électronique ou le traitement de texte, ou dans une seule partie d'un numéro, comme c'est le cas pour la publication assistée par ordinateur, d'autres sont plus transversaux.

C'est notamment le cas des discours sur les outils d'analyse de données, pour lesquels les éditeurs prévoient deux numéros thématiques pendant la période qui nous intéresse : un sur la classification automatique, et l'autre sur l'analyse factorielle des correspondances. D'autres sujets sont encore plus transversaux, comme les discussions sur l'équipement et les organisations collectives. Ceci s'explique par l'existence de rubriques dédiées à ces questions, comme « *Voyage au centre de calcul* », « *Note technique* ». ou « *Information* », autant que par des stratégies éditoriales visant à décrire aux lecteurs des expériences concrètes⁶⁷⁵.

Ce dernier point nous permet de montrer que les grands axes de discussions qui structurent les discours de cette communauté discursive dépendent à la fois de l'organisation thématique de la collection et de l'organisation des numéros sous forme de rubriques. L'articulation entre ces deux niveaux de structuration permet aux éditeurs de favoriser la circulation, dans les pages du bulletin, de différents types d'informations portant sur différents aspects de l'utilisation des ordinateurs en histoire médiévale.

Enfin, il faut revenir à la finalité que poursuit cette double structuration. Cette stratégie éditoriale permet, nous l'avons dit, d'établir des ponts entre des collectifs d'usagers des ordinateurs qui, publiant dans le même numéro, peuvent y découvrir le travail d'autres collectifs étudiant le même type de sources, utilisant les mêmes outils statistiques ou poursuivant le même type d'objectifs. Elle permet encore d'orienter les lecteurs dans leur approche de ces nouvelles pratiques informatiques, en leur présentant plusieurs facettes d'un même sujet. Comme nous allons le montrer maintenant, il apparaît que la formation à ces pratiques est elle aussi, quoique de manière moins explicite, au cœur de la finalité du *Médiéviste et l'ordinateur*.

⁶⁷⁵ Nous nous concentrerons plus précisément sur cet aspect dans le chapitre 5 de cette thèse, qui propose une étude des caractères des textes de la rubrique « *Articles de fonds* » du *Médiéviste et l'ordinateur*.

3. Favoriser la circulation de nouvelles pratiques.

Nous avons analysé dans la partie précédente l'une des stratégies éditoriales mise en œuvre par les éditeurs dans le but de coordonner la production textuelle des auteurs du *Médiéviste et l'ordinateur*. Nous n'avons cependant développé notre analyse qu'à l'échelle du corpus tout entier, sans véritablement rentrer dans le détail du contenu des numéros des bulletin. Nous ne nous sommes par ailleurs concentrés uniquement sur la fonction que jouait le bulletin vis-à-vis des auteurs, laissant à l'écart de l'analyse la dimension du lectorat. Pourtant, et nous l'avons évoqué en introduction, le bulletin vise, selon les termes des éditeurs, à favoriser les relations entre les utilisateurs des ordinateurs autant qu'à permettre de convaincre les médiévistes qui hésitent à s'emparer des moyens électroniques.

Dans l'éditorial du vingtième numéro du *Médiéviste et l'ordinateur*, publié en 1989, Fossier revient rétrospectivement sur les intentions initiales des éditeurs. Elle insiste sur le fait que le bulletin devait offrir aux lecteurs « *une information qui les touchât de près* »⁶⁷⁶. Pour ce faire, il est possible de repérer que les éditeurs avaient mis en place plusieurs stratégies éditoriales, dont deux nous intéresserons particulièrement ici : la construction de chemin d'apprentissage pour faire passer aux lecteurs des contenus techniques et la reproduction, dans les pages du bulletin, de documents de travail provenant des pratiques de recherche des auteurs.

3.1. Un lectorat hétérogène aux capacités inconnues.

Un rapide coup d'œil sur la composition du lectorat, sur la base de la liste des abonnés conservés dans les archives de l'IRHT, nous permet de mettre en lumière les difficultés auxquelles les éditeurs font face. Parmi les abonnés du bulletin de liaison, on retrouve 35 bibliothèques universitaires, de nombreuses équipes de recherche et laboratoires en histoire, linguistique et archéologie, des départements d'archives, ainsi que des particuliers parmi lesquels des historiens, des linguistes et des archéologues. Fait plus intéressant pour nous, cette hétérogénéité des positionnements disciplinaires et

⁶⁷⁶ Fossier, « Éditorial », 3.

institutionnels se superpose à une autre. Elle tient aux disparités qui peuvent exister entre les différents lecteurs quant aux différents domaines que mobilisent les auteurs de cette publication : historiographie médiévale, mathématiques, statistiques, archivistique, informatique, onomastique ou encore diplomatique.

Le défi que les éditeurs relèvent apparaît d'autant plus ambitieux que ceux-ci se proposent, comme nous l'avons dit, de publier un bulletin qui s'adresse autant aux utilisateurs aguerris des ordinateurs qu'aux néophytes. Pour créer un réseau de chercheurs, il faut aussi bien intéresser les premiers que convaincre les sceptiques ou les hésitants de l'intérêt de ces méthodes.

3.1.1. Quel lectorat-cible pour le bulletin de liaison ?

Dans cette situation, les éditeurs essaient de poser des limites quant au lectorat cible, afin d'éviter que le bulletin devienne trop généraliste, ou qu'il rentre en concurrence avec d'autres publications existantes. Cette tension se fait sentir dès le deuxième numéro, avec la publication dans l'éditorial d'une lettre d'un lecteur qu'il ne nous a pas été possible d'identifier⁶⁷⁷. Tout en félicitant les éditeurs pour la création du bulletin, cet auteur anonyme les met en garde contre le « *manque de recul et de perspective* » que manifeste le premier numéro. Parmi les raisons qu'il avance pour préciser son propos, il affirme notamment que les

*« articles ne permettent pas de comprendre s'il existe des exigences informatiques propres aux médiévistes qui se différencient par rapport à celles d'autres chercheurs en sciences humaines »*⁶⁷⁸.

Dans la continuité de cette première critique, il avertit ensuite les éditeurs du risque de pasticher un autre bulletin technique qui existe déjà dans ce contexte :

⁶⁷⁷ À ce jeu de devinette, et sur la base du peu d'indice révélé par l'auteur dans le texte, nous pouvons émettre l'hypothèse qu'il s'agirait de l'archéologue et spécialiste de l'automatique documentaire Jean-Claude Gardin (1925-2013), auquel Sébastien Plutniak a récemment consacré une grande partie de son travail de doctorat. Voir Sébastien Plutniak, « L'opération archéologique. Sociologie historique d'une discipline aux prises avec l'automatique et les mathématiques. France, Espagne, Italie, 2e moitié du XXe siècle » (Thèse de doctorat, Paris, EHESS, 2017).

⁶⁷⁸ Bourlet et al., « Éditorial », 1.

« ne faites pas du bulletin un double de celui du Centre Inter Régional de Calcul Électronique d'Orsay »⁶⁷⁹.

La réaction de ce lecteur avisé nous éclaire sur les difficultés que rencontrent les éditeurs dans la mise en œuvre d'un bulletin qui, tout en voulant faire circuler des contenus techniques, se doit d'en proposer une approche en accord avec les méthodes et les préoccupations d'une discipline spécifique. Dans ce cas, la réaction des éditeurs est de tout premier intérêt, puisqu'ils décident de modifier l'organisation thématique des numéros qui ne porteront plus seulement sur des « *types de traitement* », mais aussi sur des « *types de documents* »⁶⁸⁰.

Une autre réaction du même genre que celle de cet auteur anonyme, quoique tout à fait opposée, est apparue encore dans l'éditorial du numéro 6. Les éditeurs y répondent à André Zysberg (1947-), historien moderniste reconnu pour son utilisation des ordinateurs dans le cadre de son doctorat d'État sur les galériens aux XVII^e et XVIII^e siècles⁶⁸¹. Dans une lettre qu'il avait adressée aux éditeurs, il leur demandait de rendre compte du fait qu'ils n'avaient pas intitulé leur bulletin « *L'historien et l'ordinateur* ». Les deux interventions des lecteurs relatives au lectorat cible du bulletin sont des occasions pour les éditeurs de réaffirmer l'objectif du *Médiéviste et l'ordinateur* : il ne s'agit pas créer un réseau *ad hoc*, mais plutôt de mettre en relation des chercheurs faisant déjà partie d'une communauté existante, celle des médiévistes.

Pour ce faire, l'une des difficultés centrales qu'ils rencontrent est, nous l'avons dit, de saisir les compétences de leurs lecteurs en matière de nouvelles méthodologies. Cette préoccupation est d'autant plus fondamentale que ces méthodes recoupent de nombreux savoirs extérieurs (mathématiques et informatiques) à leurs disciplines, et le plus souvent absente de leur formation.

⁶⁷⁹ Bourlet et al., 2.

⁶⁸⁰ Bourlet et al., 2.

⁶⁸¹ André Zysberg, *Les galériens: vies et destins de 60 000 forçats sur les galères de France, 1680-1748* (Paris, France: Seuil, 1987). Pour un aperçu de la réception élogieuse de ce travail par les pairs voir Alain Cabantous, « André Zysberg, Les galériens. Vies et destins de 60 000 forçats sur les galères de France (1680-1748) », *Annales* 44, n° 6 (1989): 1544-47.

3.1.2. Exemple d'une stratégie envisagée : le choix de l'ordre des articles.

Une lettre de Genet, retrouvée dans la pochette intitulée « Le Médiéviste et l'ordinateur - numéro 9 » du carton « Comité de rédaction (vol 1-34) », adressée à Fossier et Bourlet en 1986, nous renseigne sur l'une des stratégies de publication mises en place pour tenter de gérer cette difficulté. Alors que Genet prend en charge l'édition du numéro 15 du bulletin sur l'intelligence artificielle et les systèmes experts, il affirme qu'il a choisi l'ordre des articles « *en fonction des capacités (supposées, bien sûr !) de nos [leurs] lecteurs* ». Il écrit :

« j'ai pensé l'article de Raiman facile, mais Zarri montre ensuite la complexité de ce qu'il y a derrière ; l'article IRHT vient ensuite conforter les médiévistes et leur montrer sur un point qu'ils savent difficile que l'IA apporte quelque chose, puis viennent les articles plus « techniques », mais dont j'ai peur qu'ils rebutent nos lecteurs. ».

Le choix de l'ordre des articles apparaît donc comme l'un des moyens permettant d'apporter une attention particulière à ce problème. Ce stratagème n'est d'ailleurs pas propre à ce numéro 15, puisqu'on perçoit cette problématique éditoriale de manière plus systématique dans les comptes-rendus des réunions du comité de rédaction du bulletin conservées dans les archives de l'IRHT. On y voit de nombreux brouillons simulant les différents ordres possibles, ce qui atteste du fait que la réflexion sur le plan de chaque numéro est considérée par les éditeurs comme un outil propre à résoudre la question des capacités du lectorat. De fait, le choix des auteurs est lui aussi dicté par la manière dont les éditeurs tentent de trouver des solutions à leurs manques d'informations sur les compétences de leurs lecteurs. Ces stratégies de communication constituent ainsi l'une des facettes de la mise en œuvre, dans les pages du *Médiéviste et l'ordinateur*, d'un chemin d'apprentissage pour des contenus techniques.

3.2. Chemin d'apprentissage vers l'analyse factorielle des correspondances.

Pour mieux comprendre comment ce type de stratégies de communication sont mises en place en contexte par les éditeurs, il est particulièrement intéressant d'analyser celles qui visent à exposer aux lecteurs des contenus techniques. Pour les illustrer le plus précisément possible, nous prendrons ici un seul exemple : la présentation de l'analyse factorielle des correspondances, méthode d'analyse de données multidimensionnelle développée pendant les années 1960 au sein de l'université de Rennes, puis à Jussieu, par le mathématicien Jean-Paul Benzécri (1932-2019) et son équipe⁶⁸². Cette procédure d'analyse des données fait l'objet d'un numéro complet du bulletin en 1981.

Dans l'éditorial de cette livraison, Genet indique que le choix de ce thème est dicté par trois considérations. Tout d'abord, la méthode suscite un véritable engouement dans les années 1980. Elle aurait même « *gagné la rue* »⁶⁸³ selon l'expression de l'auteur, de par la publication de ces fameux graphiques dans des journaux à destination du grand public. Pour autant, elle paraît hors d'atteinte pour un médiéviste moyen, du fait du haut degré de sophistication mathématique sur lequel elle repose. Enfin, selon Genet, plusieurs chercheurs

« *en contestent le bien-fondé, soit sur le plan mathématique, soit sur celui de l'utilité de ses applications à l'histoire* »⁶⁸⁴.

Cette cinquième livraison du *Médiéviste et l'ordinateur* propose six articles qui s'articulent avec ces considérations.

⁶⁸² Pour une histoire faite par les acteurs eux-mêmes voir Jean-Paul Benzécri, *Histoire et préhistoire de l'analyse des données* (Paris, France: Dunod : Bordas, 1982). On se reportera encore au chapitre 6 de cette thèse, qui porte précisément sur l'utilisation de cette méthode dans le contexte de l'histoire médiévale en France.

⁶⁸³ Jean-Philippe Genet, « Éditorial », *Le médiéviste et l'ordinateur* 5, n° 1 (1981): 1.

⁶⁸⁴ Genet, 1.

3.2.1. Six articles aux approches variées.

Le premier article consiste en une présentation des bases mathématiques de cette procédure⁶⁸⁵. Son auteur, Jean-Louis Robert (1945-), est un historien contemporanéiste qui a préparé un doctorat d'État sous la direction d'Antoine Prost (1933-) à l'université Paris 1 Panthéon-Sorbonne. En 1981, il vient de publier son doctorat de 3^e cycle, qui porte sur la scission syndicale de 1921⁶⁸⁶, dans lequel il utilise des analyses factorielles des correspondances pour mener son étude. Dans une recension de 1983 qui porte sur ce livre, Michel Launay (1933-2012), maître-assistant en histoire contemporaine à l'université Paris 1, insiste sur l'importance de la présentation de la méthode proposée par Robert dans ce livre, qui constitue, selon lui, l'un des seuls exposés « *qui permette au néophyte d'être quelque peu initié à l'analyse factorielle* »⁶⁸⁷. De plus, la même année, Robert est l'un des auteurs d'un manuel de statistiques destiné aux historiens⁶⁸⁸, le premier du genre à notre connaissance, qui résulte de son activité d'enseignement de ces méthodes.

Les deux articles suivants proposent des exemples d'application de cette procédure à différents types de données. Guerreau, médiéviste rattaché au CNRS, présente l'une des applications possibles de l'analyse factorielle des correspondances à un tableau de données publié par l'un de ses collègues l'année précédente⁶⁸⁹. Ce tableau porte sur un problème classique de l'historiographie médiévale, à savoir le développement du système féodal entre le VII^e et le XI^e siècle. Il y reprend sous forme synthétique les conclusions de six monographies régionales publiées depuis 1965⁶⁹⁰. Avec cet exemple, Guerreau vise à montrer l'intérêt de cette procédure pour traiter d'une problématique bien connue des médiévistes, qui soulève des difficultés liées aux opérations de synthèse qu'elle engendre.

⁶⁸⁵ Jean-Louis Robert, « L'analyse factorielle des correspondances : ses bases », *Le médiéviste et l'ordinateur* 5, n° 1 (1981): 3-5.

⁶⁸⁶ Jean-Louis Robert, « La scission syndicale de 1921: essai de reconnaissance des formes », éd. par Centre de recherches sur l'histoire des mouvements sociaux et du syndicalisme (Paris, France: Centre de recherches d'histoire des mouvements sociaux et du syndicalisme, 1980).

⁶⁸⁷ Michel Launay, review of *Review of La Scission syndicale de 1921. Essai de reconnaissance des formes*, par Jean-Louis Robert, *Le Mouvement social*, n° 122 (1983): 125-28.

⁶⁸⁸ Jean Heffer, Jean-Louis Robert, et Pierre Saly, *Outils statistiques pour les historiens* (Paris, France: Publications de la Sorbonne, 1981).

⁶⁸⁹ Alain Guerreau, « Le Haut Moyen Âge factorialisé », *Le médiéviste et l'ordinateur* 5, n° 1 (1981): 7-11.

⁶⁹⁰ Nous reviendrons longuement sur cet article de Guerreau dans le sixième chapitre de cette thèse, qui sera consacré à une controverse née de cette contribution.

L'autre exemple concerne une application de l'analyse factorielle des correspondances à l'étude de vocabulaire⁶⁹¹, du type de celle que l'on a analysée dans le deuxième chapitre de cette thèse. Dans cet article, Genet détaille comment utiliser l'ordinateur pour en tirer des graphiques sur la base de tableaux de fréquences lexicales, comment lire ces graphiques et comment y repérer les éléments clefs pour leur interprétation.

Le quatrième article de ce numéro est rédigé par un mathématicien, Jean-Pierre Fénelon (1940-2002), qui publie la même année un ouvrage destiné à présenter les procédures de l'analyse des données à un public non-mathématicien⁶⁹². Dans son article du *Médiéviste et l'ordinateur*, il ne revient pas sur les aspects mathématiques de la méthode, mais propose un dialogue fictif entre un statisticien, un mathématicien et un médiéviste, qui vise à justifier l'importance épistémologique de la démarche de l'analyse des données pour le médiéviste⁶⁹³.

Les deux derniers articles de ce numéro peuvent être quant à eux qualifiés de retours d'expériences. Le premier est rédigé par un statisticien du LISH, Philippe Cibois (1941-), qui assiste depuis 1975 des chercheurs en sciences humaines utilisant des analyses factorielles dans le cadre de leurs recherches⁶⁹⁴. Sur la base de ses expériences, il prodigue dans ce texte une série de conseils à destination des utilisateurs avérés ou potentiels de cette méthode⁶⁹⁵. Le deuxième article est le fait non plus d'un statisticien mais d'une médiéviste de l'ERA 713, Hélène Millet (1944-), qui a mis en œuvre cette procédure dans le cadre de son doctorat. Elle est aussi parmi les premiers médiévistes à avoir publié des résultats d'analyses factorielles des correspondances en 1981⁶⁹⁶. Dans son article pour le *Médiéviste et l'ordinateur*, elle propose quelques « conclusions

⁶⁹¹ Genet, « Une application de l'analyse factorielle à l'étude du vocabulaire ».

⁶⁹² Jean-Pierre Fénelon, *Qu'est-ce que l'analyse des données ? : exposé accessible aux non-mathématiciens* (Paris, France: Lefonen, 1981).

⁶⁹³ Jean-Pierre Fénelon, « Justification mathématique des correspondances », *Le médiéviste et l'ordinateur* 5, n° 1 (1981): 15-16.

⁶⁹⁴ Cibois témoigne notamment à propos de son expérience au LISH dans son Habilitation à Diriger des Recherches, soutenue en 1993, qui est consultable sur sa page web personnelle : <https://cibois.pagesperso-orange.fr/HDR93.pdf>

⁶⁹⁵ Philippe Cibois, « L'utilisation de l'analyse des correspondances », *Le médiéviste et l'ordinateur* 5, n° 1 (1981): 16-18.

⁶⁹⁶ Voir Hélène Millet, « La composition du chapitre cathédral de Laon au XIV^e siècle : résultats d'une analyse factorielle », *Annales* 36, n° 1 (1981): 117-38. Il faut cependant préciser que, quelques mois plus tard, dans la même revue, Guerreau a lui aussi publié des graphes d'analyses factorielles : Alain Guerreau, « Analyse factorielle et analyses statistiques classiques : le cas des ordres mendiants dans la France médiévale », *Annales* 36, n° 5 (1981): 869-912.

pratiques » tirées de sa propre expérience, qui vont dans le sens de l'intérêt de cette méthode pour la discipline⁶⁹⁷.

3.2.2. Six fonctions auprès des lecteurs.

Ces six articles, qui se présentent comme six propositions pour comprendre l'analyse factorielle des correspondances et ses usages, nous permettent de saisir comment les éditeurs ont conçu différents chemins d'apprentissage pour les nombreux types de lecteurs intéressés par ces questions. En effet, on constate que chacun de ces articles a sa propre fonction, et vise un certain lectorat. Dans l'éditorial de ce cinquième numéro, Genet présente par exemple la fonction de son article et de celui de Guerreau comme devant permettre à chaque lecteur d'apprécier, à l'aide des exemples d'application qu'ils développent, « *si le recours à l'analyse factorielle est possible dans le cadre de sa propre recherche* »⁶⁹⁸.

Dans ce but, l'article de Genet s'adresse à des lecteurs ou des usagers intéressés par l'étude d'un certain type de données, les données textuelles, qui permettent notamment l'étude du vocabulaire des textes. L'application de l'analyse factorielle qu'il présente repose de plus sur l'exploitation d'un type de représentation des données bien particulier, les tableaux de contingence, qui peuvent être définis comme le résultat d'opérations de tris croisés. Par ailleurs, la fonction de cet article est manifestement didactique, puisque l'auteur détaille, étape par étape, les façons de lire les différents documents de sortie produits par l'ordinateur pour mener ce type d'analyse, aspect sur lequel nous reviendrons plus en détail dans la suite de ce chapitre.

L'article de Guerreau se distingue de celui de Genet sur ces trois points, et ne s'adresse donc pas au même type de lecteurs. Tout d'abord, l'application de l'analyse factorielle qu'il illustre ne porte pas sur des données textuelles. Ensuite, ces données ne sont pas présentées sous forme d'un tableau de contingence, mais codées en 1 et en 0, ce qui constitue ce que l'on appelle un tableau d'incidence. Enfin, par contraste avec l'article de Genet, Guerreau ne détaille pas toutes les étapes de la construction des graphiques d'analyse factorielle, mais se concentre sur la lecture des résultats. De plus, cet article remplit une autre fonction, celle de montrer comment les analyses factorielles des

⁶⁹⁷ Hélène Millet, « Pour l'analyse factorielle », *Le médiéviste et l'ordinateur* 5, n° 1 (1981): 18-19.

⁶⁹⁸ Genet, « Éditorial », 1.

correspondances permettent de dégager la structure d'un ensemble de données, même si celui-ci est de faible dimension (14 lignes et 8 colonnes dans ce cas).

L'article de Fénelon est peut-être celui dont la fonction est la plus facilement reconnaissable, puisqu'il a pour rôle, selon les mots de l'éditorial, de réaffirmer « *la légitimité mathématique* » de la méthode par la plume de « *l'un de ses propagateurs mathématiciens les plus qualifiés et les plus expérimentés* »⁶⁹⁹. En ce sens, sa proposition est d'autant plus intéressante qu'il ne répond pas du tout à cette injonction des éditeurs par une démonstration mathématique. Fénelon se contente au contraire de réaffirmer, depuis sa position d'autorité, l'intérêt de la méthode pour les sciences humaines. Cet article se tourne ainsi vers les lecteurs sceptiques quant à la validité mathématique de la méthode.

L'article de Millet rend compte lui aussi de cette stratégie éditoriale, mais la poursuit sous un autre angle, puisqu'il vise surtout à insister sur l'apport de cet outil pour la recherche et sur l'intérêt qu'il peut y avoir à consacrer du temps à s'y former. Comme pour l'article de Fénelon, Genet introduit ce texte dans l'éditorial du numéro en se concentrant non pas sur son contenu, mais sur la position de son auteure dans le champ, puisqu'il la présente comme l'« *une des rares médiévistes à avoir pratiqué l'analyse factorielle* »⁷⁰⁰.

La fonction de l'article de Robert s'éclaire elle aussi si l'on prend en compte la position particulière de son auteur dans ce contexte. Tout d'abord, il faut rappeler que Robert enseigne les statistiques aux historiens à l'université Paris 1. De plus, ce dernier écrit son article en tant que spécialiste de la période contemporaine, depuis l'extérieur de la communauté des médiévistes. Selon nous, il profite ici de l'aura dont peuvent disposer, dans les années 1980, les modernistes et les contemporanéistes dans les discussions sur l'utilisation des outils mathématiques en histoire. Un troisième élément central doit cependant être précisé pour saisir la fonction de son article.

Dans l'éditorial, Genet introduit cette contribution comme un effort pour parler « *la langue du littéraire* »⁷⁰¹. Cet indice précise le type d'approche mise en œuvre par cet article dans la perspective des éditeurs, à savoir de parler aux médiévistes en tant qu'ils ont pour la grande majorité d'entre eux suivi des études plus littéraires que scientifiques.

⁶⁹⁹ Genet, 2.

⁷⁰⁰ Genet, 2.

⁷⁰¹ Genet, 2.

Mais en regardant l'article de plus près, il apparaît que cette présentation de Genet n'est pas tout à fait fidèle au texte de Robert, puisque ce dernier est largement composé de formules, de concepts mathématiques et de représentations graphiques abstraites. Le seul effort notable en ce sens consiste, selon nous, dans une note de bas de page, qui résume l'analyse factorielle des correspondances en ces termes :

« cela revient exactement à chercher la meilleure façon d'embrocher un poulet »⁷⁰².

Au-delà du caractère comique, quoique juste, de la métaphore, cette indication nous éclaire sur le fait que les stratégies des éditeurs reposent non seulement sur la présentation d'angles de vue différents sur le même sujet, mais également sur une attention quant au langage de ces articles.

L'analyse de ces différents articles nous permet de mettre en évidence comment les éditeurs d'un numéro fabriquent ce que nous avons qualifié plus haut de chemin d'apprentissage. À l'issue de l'analyse, il apparaît cependant possible de préciser que ce n'est pas un chemin qui est tracé par les éditeurs, mais qu'ils en imaginent plusieurs, qui peuvent se recomposer au gré des compétences et des intérêts des lecteurs. Pour poursuivre leurs objectifs, les éditeurs développent ainsi des stratégies éditoriales qui fondent la spécificité de ce bulletin de liaison.

3.3. Publication de nouveaux types de textes et stratégies didactiques.

On peut saisir de la façon peut être la plus claire le caractère singulier de ce périodique en observant la manière dont les articles y incorporent la reproduction de documents de travail. Comme nous l'avons vu dans les trois derniers chapitre, l'usage de l'informatique entraîne, à l'époque dont nous parlons, la production et la manipulation de textes de nouveaux types. Pour entrer les données en machine, il peut par exemple être nécessaire d'élaborer un bordereau du type de celui que nous avons étudié dans le chapitre 1. Les données sont ensuite enregistrées sur cartes perforées, bandes magnétiques,

⁷⁰² Robert, « L'analyse factorielle des correspondances », 5.

disquettes ou CD-Rom. Au cours de ces opérations, les historiens peuvent aussi produire des documents de travail plus traditionnels, mais dont les fonctions sont renouvelées par la pratique des ordinateurs, comme des tableaux, des listes ou des thesaurus.

L'usage des ordinateurs implique également l'emploi par les acteurs de documents de travail qui se rapportent au traitement des données ainsi produites. L'utilisation des programmes informatiques a pour effet majeur de munir les utilisateurs d'un ensemble de formes textuelles parmi lesquelles des index, des dictionnaires, des tableaux, des graphiques et des cartes. Ces documents de travail constituent autant d'outils d'analyse, qui proposent aux historiens de nouvelles représentations des corpus qu'ils étudient.

3.3.1. Des documents de sortie hermétiques.

Le Médiéviste et l'ordinateur se distingue de publications périodiques comparables du fait qu'il propose aux lecteurs la reproduction de nombre de ces documents. Comme nous allons le voir, ces types de textes ne sont pas présentés aux lecteurs comme éléments de démonstration ou de justification de la validité des procédures de traitement informatique des sources. Leur mise à disposition vise un objectif presque didactique. Il s'agit aux yeux des éditeurs de rendre compte de la réalité de la pratique des ordinateurs en histoire médiévale. Leurs publications attestent de l'ambition de constituer, au-delà d'un réseau d'information, une *culture épistémique commune* qui se caractériserait par le réemploi, la réutilisation, le recyclage de manières de traiter des sources, d'interpréter des documents de travail, et donc, de raisonner⁷⁰³.

Dans cette optique, il faut insister à nouveau, comme nous l'avons montré plus haut, sur le fait que dans de très nombreux cas, le choix de publier ou non ces documents

⁷⁰³ Le concept de cultures épistémiques est ici emprunté au travail de la sociologue des sciences Karin Knorr Cetina. Dans un article de 1991, elle définissait cette notion (« *epistemic cultures* ») comme : la somme des « *"untidy" goings-on of various businesses of experimentation [...] in a particular domain form what one might call an epistemic culture.* ». À la différence de ce qu'elle proposait dans cet article, nous n'utilisons cependant pas ce concept pour mettre en lumière la désunion de la science (« *disunity of science* ») ou, en d'autres termes, la diversité des pratiques locales de la rationalité scientifique, puisque nous avons justement établie cette diversité, pour notre sujet, dans la première partie de cette thèse. Nous proposons au contraire d'utiliser ce concept pour réfléchir aux manières par lesquelles certains acteurs organisent ce désordre (« *untidy goings-on* ») dans un domaine de recherche donnée, afin que l'ensemble des chercheurs de ce domaine partage un bagage épistémique commun. Cette utilisation explique à la fois l'utilisation de l'expression au singulier, et l'ajout de l'adjectif commun à cette notion. Voir Karin Knorr

relève en dernière instance des éditeurs eux-mêmes. Pour montrer comment ces documents sont intégrés dans les textes, et examiner plus précisément comment les auteurs partagent des manières de lire et d'interpréter les documents, nous nous intéresserons ici à la deuxième catégorie de documents mentionnés ci-dessus : les documents de sortie.

Remarquons d'entrée de jeu que les lecteurs du *Médiéviste et l'ordinateur* n'ont pas de raison d'être familiers des nouveaux types de textes que les historiens produisent du fait de leur nouvel environnement de travail. On peut identifier deux raisons fondamentales qui expliquent cet état de fait. D'une part, les documents de travail employés pour un projet donné ne sont pas tous mis en circulation au-delà des portes des laboratoires ou des collectifs de recherches. Nous pouvons notamment citer ici les résultats intermédiaires des traitements informatiques, qui n'ont pour objet, dans le cas d'un traitement statistique, que d'explorer le contenu d'un corpus ou, dans le cas d'une indexation documentaire, de permettre l'établissement d'une liste de descripteurs. Les revues académiques traditionnelles, comme *Les Annales ESC* ou *Les Cahiers de Civilisations Médiévales*, nous le verrons dans le chapitre 5, ne publient jamais ce type de documents alors même qu'elles proposent des études produites à l'aide d'ordinateurs.

D'autre part, ces documents dérivent d'environnements de travail très différents. Pour des raisons qui tiennent aux capacités de mémoires des technologies informatiques des années 1980, les documents de sortie des programmes statistiques présentent par exemple les noms des individus et des variables sous forme de descripteurs ou de codes qui ne sont composés que de quelques lettres (ex : « *PERSI* », « *_NRP* », « *TYS* »)⁷⁰⁴. Il s'avère donc presque impossible pour le lecteur de comprendre par lui-même le sens précis de ces sigles et de se saisir ainsi pleinement de ces documents, qui restent relativement hermétiques.

Ce type de problèmes de lecture se produit à la fois pour des technologies intellectuelles connues, comme des listes ou des tableaux, mais peut aussi concerner des objets textuels nouveaux pour la plupart des lecteurs, comme des graphiques d'analyse factorielle des correspondances. Dans ce dernier cas, les difficultés de lecture sont redoublées : il ne faut plus seulement aider le lecteur à déchiffrer le sens à attribuer aux

Cetina, « Epistemic Cultures: Forms of Reason in Science », *History of Political Economy* 23, n° 1 (1991): 105-22.

⁷⁰⁴ Ces exemples sont tirés de Genet, « Une application de l'analyse factorielle à l'étude du vocabulaire ».

sigles des variables ou des individus statistiques, mais également lui expliquer comment interpréter ces documents.

3.3.2. Un exemple de stratégie didactique : le recours aux questions directes.

Pour mieux comprendre comment les éditeurs entendent communiquer à leurs lecteurs ces façons de lire, prenons le cas, évoqué au 3.2., d'une application de l'analyse factorielle à l'étude du vocabulaire, présentée par Genet, dans le numéro 5 du bulletin⁷⁰⁵. Dans ce texte, l'auteur montre, à l'aide de cinq documents de travail, les opérations faites sur la base des sources pour comparer le vocabulaire de sept textes politiques anglais de la fin du Moyen-Âge. À cette fin, il propose d'appliquer la méthode de l'analyse factorielle de correspondances à un tableau rassemblant les 40 occurrences des mots les plus fréquents de son corpus, selon une méthode que nous avons étudiée au chapitre 1. À chaque étape du raisonnement, il montre comment cette approche met en œuvre l'utilisation de documents de travail produits par l'ordinateur et insiste explicitement sur les modalités par lesquelles le raisonnement qu'il développe s'appuie sur les documents de travail.

Nous reproduisons ci-dessous le tableau des occurrences publié dans l'article de Genet (figure 22). À partir de ce document, Genet propose aux lecteurs

« un petit exercice salubre [qui] consiste à observer le tableau pour y chercher les traits notables, et à classer ceux-ci par ordre d'importance »⁷⁰⁶.

L'objectif de cet exercice est double. Il s'agit d'abord d'engager son lectorat à le suivre dans la lecture des documents, pour qu'il se confronte lui-même à la difficulté que représente l'analyse de ce tableau d'occurrences⁷⁰⁷.

⁷⁰⁵ Genet.

⁷⁰⁶ Genet, 12.

⁷⁰⁷ La linguiste norvégienne Kjersti Fløttum, spécialiste des textes scientifiques, affirme même que ce rôle actif des lecteurs est l'un des présupposés du genre article académique (« *presupposes the active role of [its] readers* »). Voir Kjersti Fløttum, Trine Dahl et Torodd Kinn, *Academic Voices* (John Benjamins Publishing Company, 2006).

Figure 22. Tableau du nombre d'occurrences des 40 termes les plus fréquents reproduit par Genet.

	DMC	CON	TDR	APP	GOE	RPS	DPY		DMC	CON	TDR	APP	GOE	RPS	DPY
KING	80	61	86	7	249	20	31	ESTATE	4	31	26	25	35	6	4
HAVE	44	127	34	67	130	17	44	OWN	27	42	7	11	29	3	2
MAN	50	48	141	52	103	18	21	LAND	7	29	6	1	60	14	3
MAY	36	50	39	71	163	13	39	CHRIST	33	2	60	8			
DO	39	118	49	25	84	7	25	CONSEIL	3	36		8	39	7	2
GREAT	18	117	2	28	128	29	21	SELF	4	36	11	5	26	8	2
GOOD	15	130	20	52	68	11	22	RIGHT	5	34	2	25	20	4	2
GOU	33	70	127	50	14	11	10	OFFICE	2	17	24	2	37	4	3
SAY	46	99	35	6	58	22	37	MANNER	14	21	28	3	19		4
MAKE	31	33	24	17	163	9	15	KEEP	1	27	8	21	27		3
PRINCE	29	178		2	30	15	5	CAUSE	6	18	8	9	23	8	11
LORD	22	65	49	14	48	21	18	JUSTICE		51		3	13	3	10
LAW	36	13	73	19	32	6	7	COME	4	21	8	8	25	5	8
RESUME	2	22	60	5	72	10	15	REASON	5	19	15	11	18	3	8
HELL	9	84	10	31	21	11	8	PUT	13	21	9	10	16		8
PEOPLE	11	61	11	13	36	21	16	POWER	37	1	14	5	8	3	6
THING	11	54	15	27	32	10	11	WORLD	7	30	16	3	8	5	4
GIVE	31	50	5	13	42	6	8	KNOW	26	26	6	5	4		5
TIME	19	44	7	20	34	16	14	LRDSHIP	13	21	14	4	18		1
TAKE	37	20	21	19	31	10	9	CHARGE	5	2		2	58	3	

Source : Genet, 11.

L'auteur donne alors sa propre réponse à ce petit exercice :

« il est manifeste que King a une fréquence anormalement faible dans APP si on la compare à celle du même mot dans les six autres textes. Travail qui devient vite fastidieux, mais dans lequel on s'embrouille aussi très vite : comment savoir si à priori la différence notée à propos d'APP dans l'emploi de King compte plus ou moins que les différences observées à propos du mot Christ »⁷⁰⁸.

On découvre ici le second objectif de cet exercice, qui vise également à convaincre le lecteur du caractère indispensable de l'analyse factorielle dans la poursuite de l'analyse du tableau.

Genet reproduit alors un deuxième document de sortie dans son article, le tableau des valeurs propres, qui montre la représentativité de la variance pour chaque axe factoriel, et qui doit permettre de saisir à quel point la représentation graphique « résume » bien les oppositions qui structurent le jeu de données.

⁷⁰⁸ Genet, « Une application de l'analyse factorielle à l'étude du vocabulaire », 12.

Figure 23. Tableau des valeurs propres reproduit par Genet.

VAL PROPRE	POURCENT	CUMUL	HISTOGRAMME DES VALEURS PROPRES DE LA MATRICE
0.15683436	41.813	41.813	*****
0.10268337	27.377	69.190	*****
0.05062842	13.498	82.688	*****
0.03706568	9.882	92.570	*****
0.01763836	4.702	97.272	*****
0.01023180	2.728	100.000	*****

Source : Genet, 12.

Il indique dans son texte, à la suite de cette figure :

« Ici, l'information intéressante se trouve dans la colonne « pourcent » »⁷⁰⁹.

Pour guider la lecture de ce premier document de sortie, inconnu de ceux qui n'auraient jamais pratiqué d'analyse factorielle des correspondances, Genet oriente le lecteur vers la bonne manière de lire les valeurs de ce tableau. La colonne « pourcent » indique en effet les valeurs de la variance attachée à chacun des axes factoriels calculés par le programme. Schématiquement, elle permet donc au chercheur d'évaluer à quel point la représentation graphique de chaque plan « résume » les oppositions majeures du jeu de données⁷¹⁰.

Il poursuit en reproduisant un troisième document de travail (reproduit à la figure 24 ci-dessous) : le graphique représentant la structure des données selon les axes 1 et 2, qui représentent respectivement 41.813% et 27.377% de la variance (première et deuxième lignes du tableau de la figure 23).

Une fois les graphiques présentés, il pose explicitement la question au lecteur « Que lit-on sur ces graphiques ? », pour répondre lui-même :

« On discerne d'abord sur le premier graphique un paquet dans lequel figurent au centre RPS et DPY, entourés de mots comme time, self [...]. Dans le coin en haut à gauche, les mots prince, just, well, et CON »⁷¹¹.

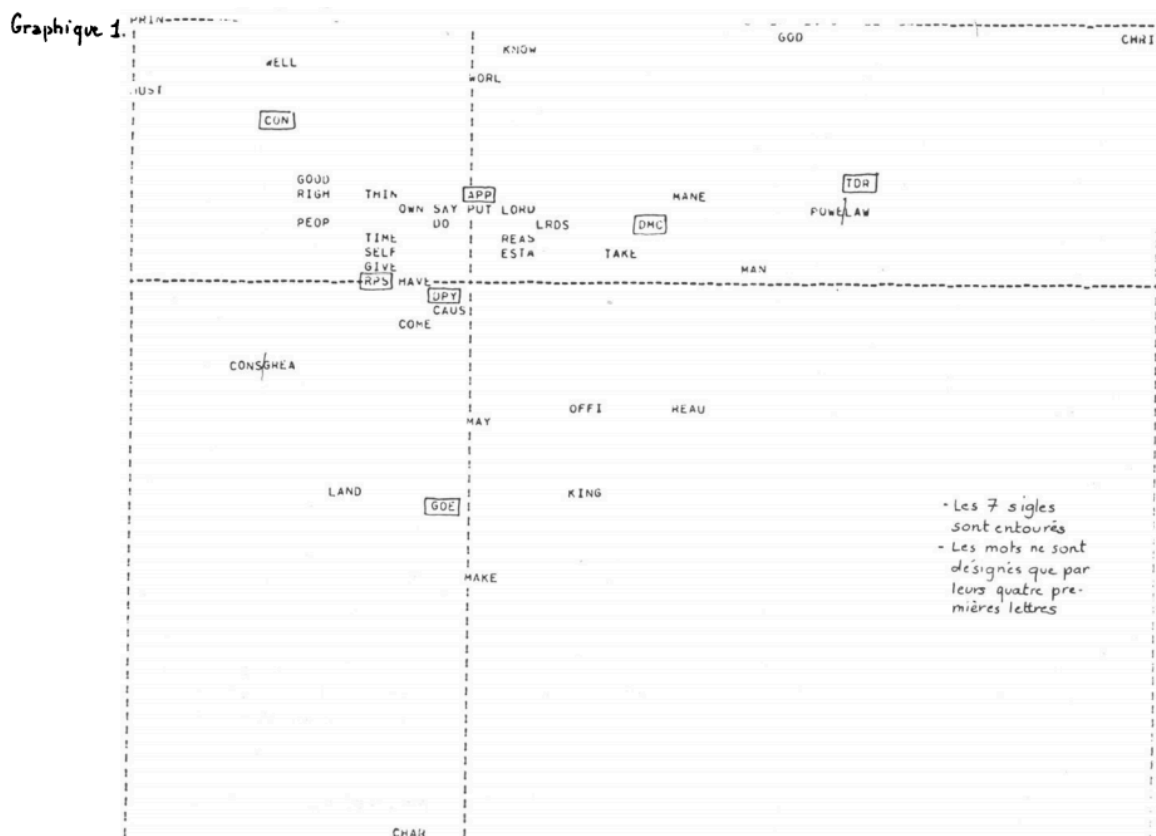
⁷⁰⁹ Genet, 12.

⁷¹⁰ Pour rappel, nous détaillons plus précisément les étapes de la procédure d'analyse factorielle des correspondances dans les chapitres 2 et 6. Il est encore possible, pour une présentation plus complète, de se reporter à Philippe Cibois, *L'analyse factorielle: analyse en composantes principales et analyse des correspondances* (Paris, France: Presses universitaires de France, 2000).

⁷¹¹ Genet, 13.

Plus loin, il introduit une nouvelle question directe invitant le lecteur à revenir aux données qui sous-tendent cette construction graphique, c'est-à-dire au tableau des occurrences que nous avons reproduit à la figure 22. « *Comment lire ces tableaux de chiffres ?* » demande-t-il, avant de proposer une analyse qui met en relation la position de quelques-uns des mots du graphique d'analyse factorielle avec les valeurs du nombre d'occurrences de ces mêmes termes dans le tableau.

Figure 24. Graphique d'analyse factorielle des correspondances reproduit par Genet.



Source : Genet, 13.

Sans revenir plus précisément sur cette analyse, qui nous éloignerait du sujet de ce chapitre, on remarque ici comment l'auteur guide la lecture des documents de travail tout comme le raisonnement mené à leur contact en émaillant son propos de plusieurs questions directes. Certains linguistes anglo-saxons se sont justement intéressés à ce type

de procédés rhétoriques⁷¹², et montre que l'intérêt de cette stratégie langagière est multiple. Elle permet aux lecteurs de suivre le raisonnement de l'auteur, aussi bien que de naviguer plus facilement entre les différents éléments du texte. La question directe incite en effet le lecteur à se reporter au graphique, et donc, à l'examiner.

Cependant, les questions de ce type assument également une troisième fonction : elles permettent une présentation didactique des raisonnements⁷¹³. Dans son texte, Genet va au-devant des difficultés du lecteur, conscient de l'étrangeté de ce type de textes pour la grande majorité des médiévistes. En effet, si l'analyse factorielle des correspondances est à la mode à cette période, comme le notent les éditeurs au début du numéro, les documents de travail sur lesquels elles s'appuient ne circulent pas dans les publications académiques. Certes, quelques graphiques d'analyse factorielle des correspondances ont été publiés en 1981 par Millet et Guerreau dans les *Annales ESC*⁷¹⁴, mais ces deux articles ne reproduisent ni les tableaux des valeurs propres ni les matrices qui sous-tendent la construction des graphiques.

L'analyse de cet article de Genet nous permet d'illustrer comment l'auteur guide la manière par laquelle le lecteur peut s'appropriier la lecture et l'interprétation des documents de travail. La stratégie rhétorique mise en œuvre par Genet dans ce cadre nous révèle une volonté de transmettre aux lecteurs des modes de raisonnement attachés à des documents de travail produits dans le contexte de l'utilisation des ordinateurs. Ce type d'articles nous paraît ainsi caractéristique de la volonté des éditeurs de favoriser l'émergence d'une culture épistémique commune.

⁷¹² Ce type de recherche a essentiellement été mené, à notre connaissance, sur la base de corpus d'article d'économie. Pour une étude sur l'anglais uniquement, voir Ken Hyland, « What Do They Mean? Questions in Academic Writing », *Text - Interdisciplinary Journal for the Study of Discourse* 22, n° 4 (2002). Pour une étude comparative plus large de l'utilisation des questions directes et indirectes dans des textes publiés en anglais, en espagnol et en français, voir la thèse récemment soutenue par Niall Curry, « Academic Writing and Reader Engagement: Contrasting Questions in English, French and Spanish Corpora » (University of Limerick, 2019).

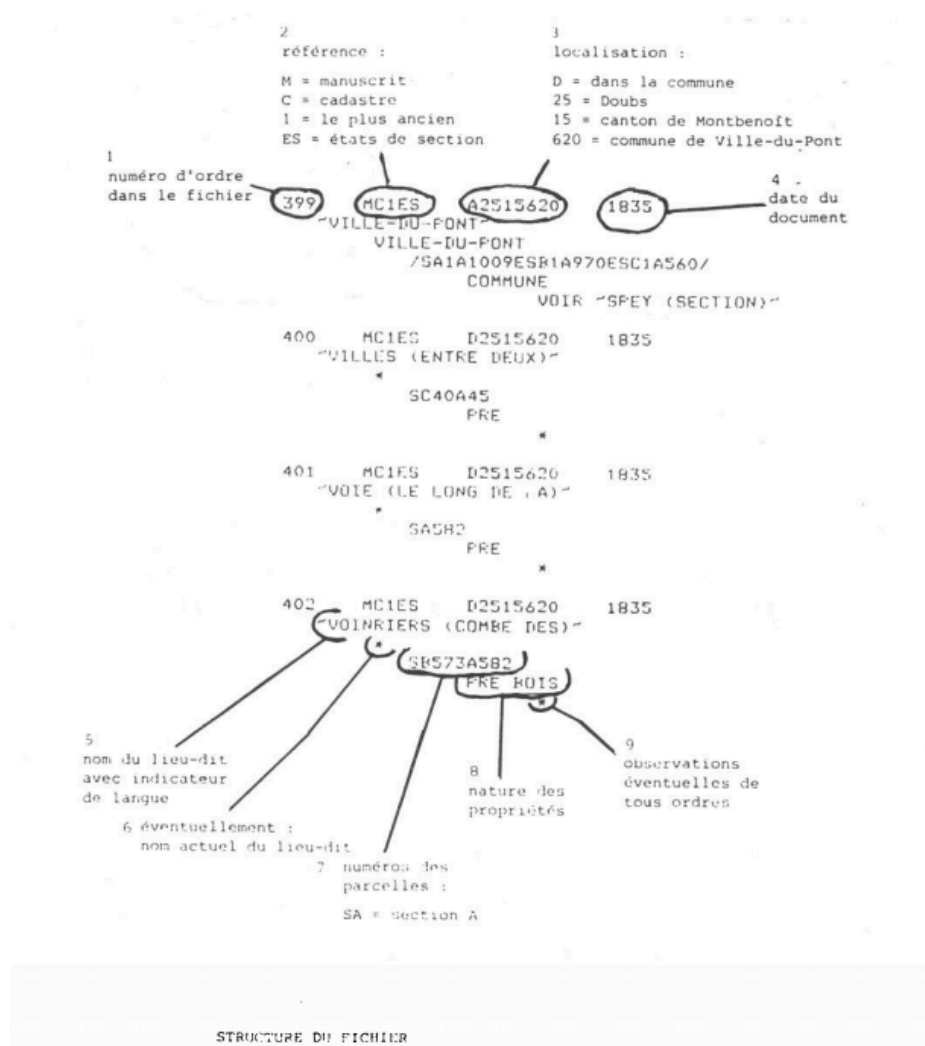
⁷¹³ Cette fonction didactique n'a, à notre connaissance, pas encore été identifiée ailleurs, ce qui s'explique selon nous par la nature des corpus étudiés par les linguistes. En effet, leurs travaux s'appuient sur des corpus d'articles académiques traditionnels de recherches, qui ont pour finalité de communiquer des résultats. Au contraire, le type de corpus qui nous intéresse ici a pour finalité de faire circuler des pratiques de recherches, et donne ainsi à voir d'autres usages scientifiques de cette stratégie rhétorique.

⁷¹⁴ Ces deux études ont déjà été citées. Voir Millet, « La composition du chapitre cathédral de Laon au XIV^e siècle » ; Guerreau, « Analyse factorielle et analyses statistiques classiques ».

3.3.3. Un autre exemple : l'annotation de documents de sortie.

D'autres auteurs du bulletin recourent eux aussi à des procédés permettant de guider le lecteur dans sa découverte de l'utilisation de nouveaux types de textes. L'une des stratégies de ce type consiste à insérer la reproduction de documents produits par des machines en les annotant. Ils peuvent ainsi rendre compte directement aux lecteurs des manières de les lire. L'exemple ci-dessous, emprunté à un article de l'historien François Lassus (1944-) publié dans *Le Médiéviste et l'ordinateur* en 1982, illustre l'une de ces stratégies d'annotation⁷¹⁵.

Figure 25. Exemple d'annotations sur un document reproduit par François Lassus.



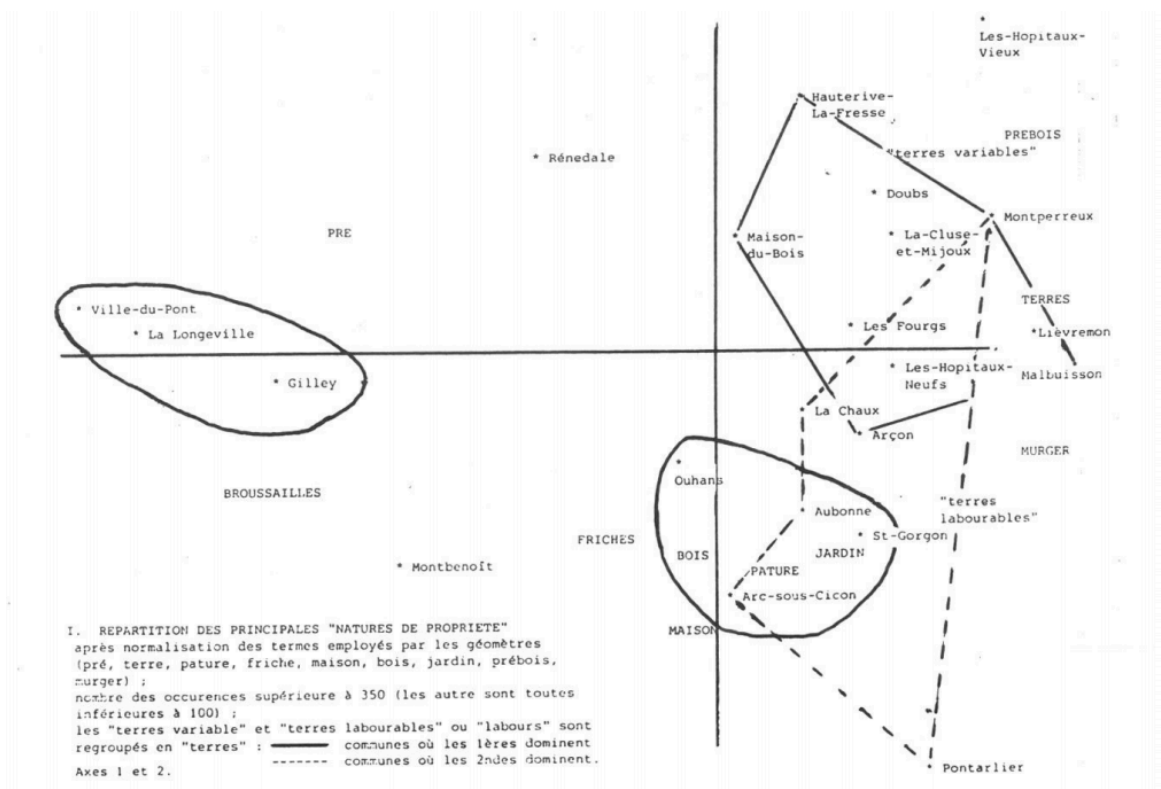
Source : Lassus, 5.

⁷¹⁵ François Lassus, « Toponymie et microtoponymie », *Le médiéviste et l'ordinateur* 8, n° 1 (1982): 4-8.

Dans ce cas, l'annotation opère comme une légende qui permet de rendre compréhensible la « *structure du fichier* »⁷¹⁶, c'est-à-dire les variables choisies par l'historien et le codage des modalités qui leur sont associées. Elle permet de révéler au lecteur comment lire ce document de sortie en dépit du caractère hermétique des suites de caractères alphanumériques qui s'y trouve représentées, dont le sens est inaccessible à ceux qui ne connaissent pas les règles de codages utilisées. L'annotation du document permet ici une description claire de ces éléments, certainement plus efficacement que ne le ferait un texte l'accompagnant.

Lassus recourt, dans le même article, à une autre forme de graphique annoté qui illustre une autre manière de guider la lecture. Sur trois graphiques d'analyse factorielle, des paquets d'individus sont entourés et, pour l'un d'entre eux, une légende précise le sens de cette opération. Nous reproduisons ce dernier ci-dessous⁷¹⁷.

Figure 26. Graphique d'analyse factorielle de correspondance annoté.



Source : Lassus, 6.

⁷¹⁶ Lassus, 6.

⁷¹⁷ Lassus, 6.

Sans revenir ni sur les données que Lassus exploite pour établir cette analyse factorielle, ni sur les conclusions qu'il tire de ce graphique, il nous apparaît que les annotations et la légende remplissent ici des fonctions évidentes. Comme l'indique la légende, les traits pleins regroupent les communes dans lesquelles les « *terres variables* » dominent, tandis que les pointillés rassemblent les communes pour lesquelles ce sont les « *terres labourables* » qui sont les plus nombreuses.

Comme nous l'indique l'auteur, ce graphique reconstitue ainsi « *un zonage déjà bien connu entre régions à vocation forestière et régions où l'élevage domine* »⁷¹⁸, que les annotations viennent souligner pour le lecteur. Les tracés permettent d'« imager » la lecture qu'en fait l'auteur. Dans ce cas, ces tracés ne sont cependant pas uniquement destinés aux lecteurs du bulletin. Ils révèlent aussi les opérations matérielles qui sont faites par cet historien sur le document de sortie produit par le programme informatique. En effet, ces annotations lui permettent de faire apparaître, pour lui-même, les articulations entre les « *natures de propriété* » (pré, terre, pâture, friche, maison, etc.) et les localités (La Longeville, Saint-Gorgon, etc.) qui lui paraissent les plus saisissantes. Cette seconde fonction de l'annotation montre ainsi aux lecteurs comment il est possible d'utiliser un graphique d'analyse factorielle des correspondances pour le rendre plus facilement lisible pour soi-même.

Un troisième exemple nous renseigne sur cette volonté de transmettre des usages des documents de travail. Nous l'avons déjà souligné : ces illustrations permettent tout à la fois de communiquer des méthodes et des raisonnements. On trouve un autre exemple de cette double fonction de présentation des documents de travail dans un article du médiéviste italien Michele Luzzati (1939-2014) publié dans le quatrième numéro du bulletin⁷¹⁹. Dans ce texte, l'auteur décrit le travail manuel rendu possible à partir de listes d'individus produites par un programme destiné à reconstruire automatiquement des familles à partir des registres de baptêmes pisans. Avant de présenter un premier document de sortie - un extrait de liste des baptisés portant le nom de famille Ferraio - Luzzati écrit :

⁷¹⁸ Lassus, 5.

⁷¹⁹ Michèle Luzzati, « L'identification des individus dans les registres de baptême pisans », *Le médiéviste et l'ordinateur* 4, n° 1 (1980): 3-6.

« Par un travail manuel effectué sur ces listes et par une série de déductions, de comparaisons, d'éliminations, [...] que l'on ne pouvait confier à la machine, nous avons réussi à isoler une série de baptisés qui appartiennent probablement à une même lignée généalogique »⁷²⁰.

Dans ce cas, le travail à la main est combiné aux tris effectués par l'ordinateur, dans le but de résoudre le problème posé par l'historien. Le document produit par l'ordinateur est donc lui-même manipulé et corrigé sans intervention de la machine. Luzzati présente cet aspect du travail en ces termes :

« L'ordinateur n'a pu pourvoir à ces opérations étant donné le grand nombre de variantes possibles (qu'on ne pouvait toujours déterminer à l'avance) et la nécessité de tenir compte d'éléments qui n'étaient pas entrés en mémoire, comme les erreurs matérielles des scribes des registres et les erreurs de lecture faites par nous-mêmes. ».

On remarque donc ici que l'utilisation des documents produits par l'ordinateur intervient à des moments différents du raisonnement. Contrairement aux cas des articles de Genet ou Lassus présentés plus haut, une articulation entre travail à la main et le travail automatique apparaît. Les documents de sorties sont donc aussi des supports qui permettent au médiéviste de vérifier les erreurs qui auraient pu être faites au moment de l'entrée des données en machine, autant que de passer outre les limites attachées aux sources elles-mêmes ou aux capacités des ordinateurs.

Cet exemple insiste sur l'intérêt de ces documents, non plus comme résultat à analyser, mais là encore, comme outils de raisonnement. Les trois exemples développés ci-dessus nous permettent ainsi de conclure que le bulletin permet non seulement de partager les manières de lire les documents de sorties des ordinateurs, mais que le maniement des listes, des graphiques ou des tableaux entraîne la circulation de raisonnement, qui ont vocation à être recyclé par d'autres membres de la communauté discursive. En s'attachant à partager ces raisonnements par le biais des textes sur lesquels ils s'appuient, *Le Médiéviste et l'ordinateur* dépasse un simple rôle d'information.

⁷²⁰ Luzzati, 4.

Sans préjuger des effets de ce mécanisme, il est possible de conclure que le partage des documents de travail dans les pages du bulletin vise à l'élaboration d'une culture épistémique commune, fondée sur l'utilisation de nouveaux types de documents de travail. Cette ambition s'accorde ainsi avec l'un des objectifs de cette communauté discursive : rendre accessibles et reproductibles les expériences singulières des médiévistes utilisant des ordinateurs pour favoriser une plus large utilisation de ce type de méthodes.

Conclusion.

Dès le chapitre précédent, qui constituait le premier volet de la deuxième partie de cette thèse, nous avons mis en avant certaines conséquences liées à l'utilisation des ordinateurs en histoire médiévale à une échelle différente de celle à laquelle nous nous étions attachés dans la première partie. Il s'agissait en effet de comprendre comment ce phénomène avait pu avoir des conséquences sur les relations entre les collectifs de médiévistes qui avaient mis en œuvre ces nouvelles approches entre 1970 et 1979, et donc, plus généralement, sur l'évolution de certains aspects de la discipline pendant cette période.

Dans ce quatrième chapitre, nous nous sommes intéressés à ce type de phénomène pour la période suivante (1979-1989), en nous attachant à étudier les activités d'un groupe de chercheurs organisés autour de la publication du périodique *Le Médiéviste et l'ordinateur*. Pour conclure cette partie, nous proposons d'examiner dans le même temps 1) comment les premières intentions d'organisation méta-collective, identifiées dans le chapitre précédent, se sont développées pendant la période qui nous intéresse ici et 2) quel était l'état de ce champ de recherche à l'issue de cette deuxième période.

Le changement de focale que nous avons opéré dans le chapitre précédent nous a conduit à mettre en avant les ambitions attachées à l'utilisation des ordinateurs au début des années 1970 en histoire médiévale. Ces ambitions reposaient sur la possibilité qu'offraient ces moyens électroniques de constituer des instruments de travail (index, éditions, concordanciers, programmes informatiques, bases de données, etc.) permettant d'étudier de grands ensembles de documents. Ces instruments devaient notamment

permettre de dépasser les limites qu'imposaient les logiques d'archivage des sources telles qu'elles avaient été mises en œuvre depuis la deuxième moitié XIX^e siècle, et offrir de nouvelles perspectives pour la recherche documentaire et l'étude de ces corpus.

Dans la conclusion du colloque organisé à Rome en 1975, Léopold Genicot (1914-1995) présentait quelques « *propositions précises* » pour organiser la production de ces instruments à l'échelle du groupe des premiers utilisateurs d'ordinateurs :

« Il faudrait d'abord décrire attentivement les genres d'ensembles [...] dans leurs caractères externes et surtout dans leur contenu et leur texture [...] pour décider de la manière de les traiter, juger de la nécessité et de la possibilité de les enregistrer ou non [...]. On ferait ensuite un second pas : élaborer de façon précise, avec le concours de techniciens, le mode de traitement, le « programme », le mieux adapté à chaque genre. Pour obtenir ou permettre que toute entreprise se coule désormais dans un moule uniforme et approprié. Il s'agirait enfin de faire connaître les projets achevés, en cours ou envisagés. »⁷²¹

Si certaines des étapes de ce programme avaient déjà été engagées au moment du colloque de Rome ou même en amont, nous avons montré dans le présent chapitre que plusieurs numéros du *Médiéviste et l'ordinateur* s'inscrivaient dans la continuité de cet agenda.

Nous pouvons même affirmer, au terme de notre analyse, que le mode de traitement adapté à certains de ces genres d'ensembles documentaires est défini dans les pages du bulletin. C'est ce que nous avons montré avec l'exemple des deux numéros dédiés aux sources sérielles, dans lesquels apparaissait une corrélation forte entre la structure des informations contenues dans les sources (éléments répétitifs mais structure complexe, liste) et le choix de la méthode d'analyse ou d'enregistrement (indexation documentaire, couplage automatique).

On pourrait y ajouter les considérations développées par Genet dans l'éditorial du numéro 11 qui paraît au printemps 1984. Alors qu'il postulait que les testaments, l'un des rares types de sources sérielles à la disposition des médiévistes, devaient pouvoir être traités selon les méthodes de l'indexation documentaire, il constatait, sur la base des

⁷²¹ « Conclusion générale », *Publications de l'École Française de Rome* 31, n° 1 (1977): 430.

contributions de ce numéro, que les médiévistes leur appliquaient dans la majorité des cas un enregistrement plein-texte⁷²².

Ces deux exemples nous montrent que les échanges qui prennent forme dans les pages du *Médiéviste et l'ordinateur* conduisent les acteurs à repérer des tendances générales et, ainsi, à avancer sur le chemin tracé par Genicot en 1975. Cependant, à la différence de ce que proposait Genicot, l'identification de ces corrélations et la circulation de ces informations ne résultent pas d'une approche normative, mais au contraire de la coordination de la production textuelle autour de thèmes fédérateurs, définis par les éditeurs du *Médiéviste et l'ordinateur*.

Sur ce dernier point, ce chapitre présente selon nous un intérêt méthodologique plus général que celui que l'on porte dans cette thèse à l'histoire de l'utilisation des ordinateurs par les médiévistes, de par la reprise d'un concept clef de l'analyse du discours, celui de *communauté discursive*. Selon nous, l'étude des logiques d'organisation des communautés scientifiques, sujet transversal en histoire des sciences, a beaucoup à gagner de la proposition faite par Swales lorsqu'il distingue les *speech communities* des *discourse communities*. Cette approche méthodologique permet en effet à la fois 1) de s'affranchir de logiques institutionnelles ou territoriales dans la construction de ce type d'objet de recherche et 2) de fonder cette construction sur la base des textes partagés par leurs membres et donc, d'examiner comment l'appartenance à une communauté scientifique se définit par la maîtrise de savoirs et de savoir-faire liés à ces textes (production, manipulation, écriture, lecture, etc.).

Un deuxième élément qui était au centre des préoccupations des acteurs du colloque de Rome consistait à définir de nouveaux moyens qui soient à même de faire circuler les instruments de travail ainsi produits. À la suite de Rome, les propositions en ce sens avaient notamment été articulées autour de la création d'un « *modèle de description* » standard pour tous les projets impliquant l'utilisation des ordinateurs⁷²³. Sans qu'il nous soit possible d'affirmer si Genicot a effectivement rassemblé des documents de ce type dans la deuxième moitié des années 1970, nous avons vu dans ce

⁷²² Jean-Philippe Genet, « Éditorial », *Le médiéviste et l'ordinateur* 11, n° 1 (1984): 1-2.

⁷²³ Pour rappel, ce modèle de description est présenté dans Léopold Genicot, « Pour une organisation de la recherche en histoire médiévale, II. L'information sur les recours aux ordinateurs », *Francia* 3 (1975): 643-47.

quatrième chapitre que les médiévistes s'étaient dotés, avec le bulletin de liaison *Le Médiéviste et l'ordinateur*, d'une tribune pour les porteurs de projets de ce type.

À partir de 1979, les pages du bulletin de liaison constituaient ainsi autant un lieu du développement d'un modèle standard de description, sur lequel nous reviendrons dans le chapitre 5, qu'un canal pour la diffusion de ces informations. Les activités connexes à l'édition et la publication du bulletin des membres du comité éditorial (enquêtes, mises en relation des chercheurs, consultations techniques, expertises) permettaient par ailleurs d'assurer une forme de recensement des projets indispensable à la poursuite de leurs objectifs.

Il faut toutefois préciser que le groupe de médiévistes qui nous intéresse ici n'est pas le seul à se mobiliser pour favoriser les discussions sur la circulation de ces nouveaux instruments de travail. Depuis le début des années 1980, de nombreuses autres initiatives méta-collectives s'étaient en effet développées en lien avec ces mêmes problématiques. C'était par exemple le cas du lancement en 1985, à l'initiative du spécialiste du traitement informatique de données historiques allemand Manfred Thaller (1950-), ingénieur au sein du Max Planck Institut für Geschichte de Göttingen, d'une série de conférences qui concernaient la standardisation et l'échange de bases de données historiques. En 1987, la troisième édition de cette série de rencontres a d'ailleurs lieu au LISH, à Paris, à l'invitation de l'un des éditeurs du *Médiéviste et l'ordinateur* : Genet⁷²⁴.

En plus de cette diffusion d'informations à propos des projets achevés ou en cours d'élaboration, qui s'inscrivait là aussi dans la droite ligne de la conclusion de Genicot au colloque de Rome, nous avons par ailleurs montré que *Le Médiéviste et l'ordinateur* offrait à ses lecteurs des informations sur bien d'autres aspects de l'utilisation des ordinateurs : sur les infrastructures existantes, sur les rencontres en lien avec ces méthodes ou sur les nouveaux équipements à la disposition des médiévistes pendant cette période (micro-informatique, lecteur optique, disquette, etc.). Ce second aspect de la fonction informative jouée par le bulletin de liaison était notamment assuré par l'existence de

⁷²⁴ Voir Elisabeth Mornet, « Standardisation et échange des bases de données historiques. Actes de la troisième Table Ronde internationale tenue au L.I.S.H. (C.N.R.S.), Paris, 15-16 mai 1987. Jean-Philippe Genet éd.-Paris, C.N.R.S., 1988 », *Le médiéviste et l'ordinateur* 20, n° 1 (1988): 4-5. Si l'on prend le cas plus général de l'organisation du partage des instruments de travail en sciences sociales, les initiatives sont encore plus pléthore, avec la création de nombreux comité, conseil ou fédération aux États-Unis et à l'échelle internationale dès la deuxième moitié des années 1970. Voir Erwin K. Scheuch, « Les services de données en sciences sociales : historique et perspectives », *Revue internationale des sciences sociales* n° 177, n° 3 (2003): 433-49.

différentes rubriques qui visaient à donner aux lecteurs les clefs de la vie de cette communauté discursive (« *Voyage au centre de calcul* », « *Information* », « *Courrier des lecteurs* »).

À travers cette fonction, l'ambition affichée par les éditeurs était de créer des liens entre les membres de cette communauté discursive dans le but de constituer un réseau. Cela passait notamment par la publication dans les pages du bulletin d'informations susceptibles de provoquer des rencontres entre les lecteurs et les auteurs : calendrier des événements consacrées à ces questions (séminaires, colloques, tables rondes), contacts des responsables des centres de calculs, références de livres permettant de se former à certaines méthodes, etc. Ce réseau se construisait encore à travers les activités extra-éditoriales des éditeurs - en particulier les membres de la section informatique de l'IRHT - : mise en contact direct de certains chercheurs partageant les mêmes problématiques, correspondance entre les éditeurs et de nombreux chercheurs en Europe et aux États-Unis ou organisation à Paris d'une table de ronde en 1989⁷²⁵.

L'une des évolutions que l'on observe sur ce point concerne le développement du réseau des médiévistes français pendant cette période (1979-1989). Tout d'abord, on remarque que les multiples activités des éditeurs du bulletin se corrélaient avec un renforcement des liens entre les médiévistes utilisant des ordinateurs en France. Ils organisent notamment des colloques autour de problématiques particulières (traitement des cadastres, prosopographie, base de données, etc.)⁷²⁶ ou se regroupent sous la forme de nouveaux cadres administratifs, comme l'Action Thématique Programmée « Prosopographie » créée en 1983 qui réunissait des membres de l'ERA 713, de l'IRHT et du CNRS⁷²⁷.

Ensuite, alors que le colloque de Rome témoignait essentiellement des contacts entre chercheurs français, belges et italiens, les pages du bulletin attestaient de contacts

⁷²⁵ Lucie Fossier, éd., *Le médiéviste et l'ordinateur : actes de la table ronde, Paris, CNRS, 17 novembre 1989* (Paris, France: Institut de Recherche et d'histoire des textes, 1990).

⁷²⁶ Voir par exemple : Centre d'histoire urbaine, École française de Rome, et Centre national de la recherche scientifique, *Les cadastres anciens des villes et leur traitement par l'informatique: actes de la table ronde, 31 janvier-2 février 1985, Saint-Cloud*, éd. par Jean-Louis Biget, Jean-Claude Hervé, et Yvon Thébert (Rome, Italie: École française de Rome, 1989, 1989); Jean Glénisson, « Prosopographie et informatique », *Publications de l'École Française de Rome* 31, n° 1 (1977): 227-29; Mornet, « Standardisation et échange des bases de données historiques. Actes de la troisième Table Ronde internationale tenue au L.I.S.H. (C.N.R.S.), Paris, 15-16 mai 1987. Jean-Philippe Genet éd.-Paris, C.N.R.S., 1988 ».

⁷²⁷ Cette structure est notamment évoqué dans Lucie Fossier et al., « Vingt ans d'informatique en histoire médiévale », *Actes des congrès de la Société des historiens médiévistes de l'enseignement supérieur public* 20, n° 1 (1989): 501-25.

entre les éditeurs et des collectifs en Allemagne, au Royaume-Uni ou aux États-Unis. La publication du *Médiéviste et l'ordinateur* participe ainsi dans les années 1980 au développement de relations plus larges que celles de collectifs à collectifs que nous avons mises en lumière pour le début des années 1970.

Il faut toutefois insister ici sur les évolutions technologiques qui ont lieu pendant cette décennie. Elles modifient en profondeur non seulement les modalités de ces collaborations mais également les sources disponibles pour les étudier. En 1989, à l'issue de la deuxième enquête lancée par le service informatique de l'IRHT, Fossier et Bourlet constatent que 75% des médiévistes adeptes de l'informatique travaillent désormais sur leurs propres micro-ordinateurs. Aux conséquences techniques de l'utilisation massive de ces nouveaux instruments, que nous avons analysés dans ce chapitre, il faut ajouter que les acteurs font le constat d'un danger qui tient à cette nouvelle situation : « *l'éparpillement des entreprises* » voire même le « *cloisonnement néfaste des projets* »⁷²⁸.

Dès lors, ce n'est plus uniquement la *situation de communication* dans laquelle se trouvent les éditeurs du bulletin qui évolue, du fait des nouvelles informations que recherchent ces médiévistes, ainsi que les logiques de collaboration entre les acteurs qui se transforment. À l'exception de quelques articles présentant des logiciels pour les micro-ordinateurs, comme SAS, JEUEMO, TEXTTO, TeX, etc. le corpus mobilisé dans ce chapitre ne permet pas d'observer comment ce tournant technologique a impacté l'organisation méta-collective entre ces chercheurs dans la deuxième moitié des années 1980. L'une de nos ambitions est ainsi, pour la suite, de poursuivre notre enquête en nous intéressant plus particulièrement à la fin des années 1980 et aux années 1990, afin de saisir cette évolution en lien avec les nouveaux canaux de circulation des instruments de travail et des informations qui se développent alors (club de micro-informatique, *cédérom*, *mail list*, intranet, internet, forum, etc.).

En plus de la fonction informative du bulletin que nous avons examinée, nous avons identifié que les éditeurs mettent en place des stratégies qui visent à faire circuler de nouvelles manières de faire de la recherche en histoire médiévale. Cette dimension

⁷²⁸ Lucie Fossier et al., « Vingt ans d'informatique en histoire médiévale », *Actes des congrès de la Société des historiens médiévistes de l'enseignement supérieur public* 20, n° 1 (1989): 517.

rejoint un troisième élément largement discuté au colloque de Rome, qui concernait le partage des documents de travail, et en particulier de certains d'entre eux, comme les thesaurus, les programmes informatiques ou les modèles de grille d'analyse destinés à constituer des bases de données statistiques. Sur ce point, on constate dans les pages du *Médiéviste et l'ordinateur* une profonde évolution par rapport à la période précédente. Les éditeurs favorisent en effet non seulement le partage de ce type de documents de travail, pléthore dans les pages du bulletin, mais font également en sorte que les auteurs publient des documents de sorties des programmes et discutent des modes de lecture et de l'utilisation qu'ils font de ces nouveaux types de textes.

Nous avons montré que cette volonté s'inscrivait plus largement dans ce que nous avons qualifié de fonction didactique du bulletin de liaison. En effet, l'un des objectifs assumés par les éditeurs était de favoriser la circulation de savoirs et de savoir-faire attachés aux utilisations des moyens électroniques de calcul. L'organisation thématique qu'ils avaient mise en place nous renseigne, à l'échelle de l'ensemble de la collection, sur la poursuite de cet objectif. On en trouve encore une autre manifestation dans la constitution, pour certains numéros, de chemins d'apprentissage permettant à des lecteurs aux compétences hétérogènes de s'approprier des contenus techniques ou mathématiques.

Plus fondamentalement, nous avons ainsi pu affirmer que les auteurs poursuivaient la volonté de constituer une *culture épistémique commune*, composée d'un ensemble de pratiques, d'outils ou de méthodes utilisés à l'échelle locale par les collectifs ou les chercheurs isolés. Pour illustrer la volonté qu'avaient les éditeurs de fournir aux lecteurs des éléments constitutifs d'un bagage commun, les numéros du bulletin que nous avons catégorisés comme relevant de ce que Fossier appelait des « *instruments méthodologiques* » sont les plus éclairants : traitement de texte, analyses factorielles des correspondances, classification automatique, indexation documentaire, systèmes experts. Il faut y ajouter l'ensemble des discours portant sur la matérialité des opérations, les énoncés présentant les modalités d'utilisation des instruments de travail du type de ceux que nous avons présentés dans le chapitre précédent (base de données, index informatisé, programme) ou encore les techniques de lecture des documents de sorties des ordinateurs qui sont attachés à ces nouvelles méthodes.

Le renversement de l'utilisation du concept de cultures épistémiques, proposée par Knorr Cetina pour déconstruire le modèle de l'unité de la rationalité des sciences au

début des années 1990⁷²⁹, nous a ainsi permis de rendre compte, pour un domaine particulier, de l'existence de stratégies visant à harmoniser certaines dimensions des pratiques de recherches développées dans des cultures plus locales. Nous avons encore montré que ces stratégies ne s'articulaient ni avec une volonté normative ni avec des ambitions d'uniformisation de ces manières de faire, mais que les acteurs qui les mettaient en œuvre reconnaissaient au contraire l'intérêt de cette diversité en même temps qu'ils tentaient d'en assurer l'unité.

Au terme de ce chapitre, nous souhaitons revenir sur la méthodologie que nous avons suivie, inspirée d'un article de Maingueneau dans lequel celui-ci établissait une grille d'analyse composée des différents éléments que proposaient les linguistes qui s'intéressaient à la *situation de communication*. Cette approche nous a permis de rendre compte d'une série d'éléments de contexte à même de soulever des questions historiques sur le rôle des éditeurs de ce bulletin dans la mise en œuvre d'une organisation méta-collective, sur les objectifs qu'ils s'étaient fixés et sur les stratégies éditoriales qu'ils avaient mises en œuvre pour y parvenir.

Elle nous a encore permis de définir les fonctions que poursuivaient ce bulletin de liaison, et de faire de cette étude de cas un exemple qui pourra servir de base à une réflexion plus large sur les fonctions de ce genre de périodiques dans la constitution de cultures et de communautés scientifiques. La fonction didactique que nous avons étudiée est-elle spécifique à notre contexte ? Ou peut-on la retrouver dans d'autres publications ? Les éditeurs de ce type de périodiques assurent-ils toujours des fonctions extra-éditoriales en lien avec le champ de recherche dans lequel ils s'inscrivent ? Ou ce type d'activités se développe-t-il seulement en lien avec l'irruption dans la discipline de nouveaux instruments techniques ou technologiques ? Plus largement, cet exemple pourrait encore servir de base pour une étude des fonctions des *newsletters*, dont nous avons montré en introduction qu'elles partagent, selon la définition qu'en donne Kelty, de nombreux points communs avec les bulletins de liaison.

Ces questions nous apparaissent d'autant plus pertinentes que certains acteurs ont reconnus l'importance de cette forme d'organisation dès les années 1980 dans notre

⁷²⁹ Cetina, « Epistemic Cultures ».

contexte. Dans une table ronde intitulée « Histoire et informatique »⁷³⁰, organisée conjointement par l'Institut d'Histoire du Temps Présent et l'Institut d'Histoire Moderne et Contemporaine, Zysberg affirmait :

« L'idée serait, au départ, de créer un petit réseau d'informations. Je suis très frappé par l'avance qui a été prise, en ce domaine, par les médiévistes, avec leur bulletin : « le Médiéviste et l'ordinateur » »⁷³¹.

La mise en œuvre de ce projet sera un peu plus tardive pour les modernistes et les contemporanéistes, avec la création en 1988 de la branche française de l'Association for History and Computing : l'Association Française pour l'histoire et l'informatique. Elle publiera, à partir de 1989, un bulletin bi-annuel intitulé *Mémoire Vive*. Ainsi, si l'avance avait été prise par les historiens modernistes dans le traitement des sources conservées en séries, comme dans l'utilisation de la mécanographie et des ordinateurs dans les années 1960, les médiévistes deviennent, en France, un modèle d'organisation pour tous les historiens qui utilisent des ordinateurs.

Afin de saisir certains des effets de ce modèle d'organisation, nous proposons dans les deux chapitres suivants des études de certaines modalités des échanges qui prennent forme dans les pages de ce bulletin de liaison : 1) le genre d'article qui s'y développe en relation avec la nécessité de faire circuler de nouveaux types de pratiques de recherches et 2) les difficultés qu'ont pu rencontrer les utilisateurs de ces méthodes dans le partage de ce type de savoirs avec certains de leurs collègues.

⁷³⁰ Il est intéressant de noter, en écho au reste de ce chapitre, que Fossier, Klapisch-Zuber, Zarri et Genet, tous les quatre éditeurs du *Médiéviste et l'ordinateur*, sont dans le public de cette table ronde.

⁷³¹ André Zysberg, « Les historiens et l'apprentissage de l'informatique », in *Table ronde « Histoire et informatique » du 8 mars 1980* (École Normale Supérieure: Édition du CNRS, 1980), 16-28.

PARTIE III

Chapitre 5 : Genre et fonctions des « articles de fond » du *Médiéviste et l'ordinateur* (1979-1989).

Le chapitre précédent a permis de conclure qu'il existait, pendant les années 1980, une communauté discursive originale, qui se développait, entre autres, du fait des activités des membres du comité éditorial du *Médiéviste et l'ordinateur*. Les éditeurs de ce périodique avaient articulé leur projet autour de deux objectifs : 1) la mise en relation directe entre les nombreux acteurs impliqués ou intéressés par les utilisations des ordinateurs en histoire médiévale, et 2) la formation d'une culture épistémique commune. Si l'on a pu voir comment, et par qui, cette publication était organisée, nous n'avons pas encore analysé pleinement la spécificité des textes publiés dans ce bulletin de liaison. Cet aspect est d'autant plus primordial pour notre travail que les acteurs se sont eux-mêmes attachés, dès le début des années 1970, à réfléchir à la forme des textes adaptée à la circulation des méthodes informatiques en histoire.

Le premier témoignage publié en ce sens est, à notre connaissance, le fait du médiéviste belge Léopold Genicot (1914-1995)⁷³². Dans la conclusion du colloque de Rome « Informatique et Histoire médiévale » de 1975, il affirme qu'il faut « *présenter les réalisations sous une forme normalisée* »⁷³³. Le même son de cloche se fait entendre, cinq ans plus tard, en histoire moderne et contemporaine, dans les actes de la Table Ronde « Histoire et Informatique »⁷³⁴, organisée en 1980 par les tout jeunes Institut d'Histoire du Temps Présent et Institut d'Histoire Moderne et Contemporaine. André Zysberg (1947-), figure importante de l'introduction de méthodes informatiques en histoire

⁷³² Le lecteur curieux trouvera dans le chapitre 3 de cette thèse des détails biographiques sur ce médiéviste, et une analyse du « modèle de description » qu'il propose en 1975.

⁷³³ « Conclusion générale », *Publications de l'École Française de Rome* 31, n° 1 (1977): 430.

⁷³⁴ Institut d'histoire moderne et contemporaine et Institut d'histoire du temps présent, éd., *Table ronde « Histoire et informatique »*, *École normale supérieure*, 8 mars 1980 (Paris, France: CNRS, 1980).

moderne⁷³⁵, décrit, d'une façon différente de Genicot, le problème associé à la publication de ce type d'études :

« *L'information. Elle passe rarement par le canal des revues spécialisées d'histoire. Lorsque vous écrivez un article à propos d'une recherche qui a nécessité l'emploi de l'informatique, vous ne pouvez presque jamais détailler votre méthode, ou exposer les difficultés que vous avez rencontrées : tout est rejeté dans les notes, et encore ...* »⁷³⁶

Ce diagnostic proposé par le moderniste, permet de soulever une série de questions qui se doivent d'être examinées en détail.

Le premier ensemble de ces questions concerne les particularités du type d'articles du *Médiéviste et l'ordinateur*, aperçues dans le chapitre précédent au prisme de la reproduction des documents de travail. Existe-t-il un *genre* de textes académiques propre à l'exposition d'études utilisant les ordinateurs en histoire médiévale ? Dans quels contextes de publication, ou en d'autres termes, dans quels types de périodiques, ce genre se développe-t-il ? Quelles sont alors les spécificités des articles publiés dans le *Médiéviste et l'ordinateur* par rapport à ce genre ?

L'utilisation de la notion de *genre* en histoire des sciences n'est pas une nouveauté⁷³⁷. Cependant, son utilisation s'est inscrite jusque-là, à notre connaissance, dans la continuité de l'approche développée par le linguiste américain Douglas Biber (1952-), dont les travaux avaient pour objet de regrouper des textes en se fondant sur l'analyse de certaines de leurs caractéristiques communes⁷³⁸. Dans notre cas, ce n'est pas

⁷³⁵ Pour un aperçu des travaux de ce moderniste, voir sa thèse, citée dans le chapitre précédent : André Zysberg, *Les galériens: vies et destins de 60 000 forçats sur les galères de France, 1680-1748* (Paris, France: Seuil, 1987). On pourra encore, pour découvrir ses réflexions sur les usages des ordinateurs en histoire, regarder André Zysberg, « Impact de l'informatique sur la recherche historique », *Matériaux pour l'histoire de notre temps* 10, n° 1 (1987): 35-37.

⁷³⁶ André Zysberg, « Les historiens et l'apprentissage de l'informatique », 24.

⁷³⁷ Il est notamment possible de citer ici certains travaux du groupe « Histoire des sciences, histoire du texte » du laboratoire SPHERE (CNRS & université de Paris – UMR 7219), et plus particulièrement l'article de Florence Bretelle-Establet, « The Issue of Textual Genres in the Medical Literature Produced in Late Imperial China », in *Texts, Textual Acts and the History of Science*, éd. par Jacques Virbel et Karine Chemla (Springer Verlag, 2015).

⁷³⁸ Voir Douglas Biber, *Variation across Speech and Writing* (Cambridge: Cambridge University Press, 1988). Pour une illustration d'une approche de ce type, développée par une historienne des sciences, voir Bretelle-Establet, « The Issue of Textual Genres in the Medical Literature Produced in Late Imperial China ».

la classification des textes considérés qui nous intéresse, mais plutôt leurs fonctions au sein de la *formation discursive* constituée de l'ensemble des énoncés portant sur l'utilisation des ordinateurs en histoire médiévale pendant les années 1980.

Poursuivant cette problématique, nous emprunterons dans ce chapitre une autre approche de la notion de genre, qui s'inscrit dans la continuité des travaux du linguiste John Swales (1938-). Dans un livre de 1990, il définit ce concept en ces termes :

« *A genre comprises a class of communicative events, the members of which share some set of communicative purposes. These purposes are recognized by the expert members of the parent discourse community, and thereby constitute the rationale for the genre. This rationale shapes the schematic structure of the discourse and influences and constraints choice of content and style* »⁷³⁹.

Dans cette définition, Swales met en avant trois éléments déterminants, qui guideront notre analyse : 1) un genre de texte se caractérise en lien avec les objectifs de communication (*communicative purposes*) qu'il poursuit ; 2) ces objectifs sont attachés à une communauté discursive (*the parent discourse community*), et donc, aux relations qui existent entre auteurs et lecteurs et 3) ces objectifs sont constitutifs de la logique (*rationale*) du genre et façonne la structure schématique, le choix du contenu et le style des articles.

Pour répondre aux questions formulées ci-dessus, nous comparerons les articles du *Médiéviste et l'ordinateur* à d'autres corpus, composés de textes publiés par les mêmes acteurs dans des revues académiques plus traditionnelles. À la lumière des éléments sur lesquels Swales insiste, nous mènerons cette comparaison sur deux fronts. Elle portera d'une part sur les objectifs de communication que poursuivent les auteurs ou, en d'autres termes, sur ce qu'ils cherchent à communiquer dans leurs articles. Elle impliquera d'autre part une analyse des contraintes spécifiques que suppose la publication d'articles dans chacun de ces différents contextes de publication, c'est-à-dire des règles qui régissent les

⁷³⁹ John M. Swales, *Genre analysis: English in academic and research settings* (Cambridge GB, Royaume-Uni de Grande-Bretagne et d'Irlande du Nord, 1990), 58.

genres d'articles publiés par ces périodiques quant à la structure schématique, le choix du contenu et le style.

Le diagnostic de Zysberg présenté ci-dessus inspire par ailleurs d'autres questions, qui portent plus spécifiquement sur les deux éléments sur lesquels il insiste dans la citation reproduite ci-dessus : le détail des méthodes et l'expression des difficultés. Quelles sont les fonctions jouées, respectivement, par ces deux types d'énoncés dans la circulation des méthodes informatiques dans notre contexte ? En d'autres termes, pourquoi est-il nécessaire, selon lui, que les acteurs partagent ces éléments dans des publications ?

L'analyse des fonctions dévolues à ces énoncés permettra de dégager les différences qui peuvent exister entre les articles de notre corpus sous un autre jour, non plus essentiellement linguistique, mais liées aux contenus qu'ils exposent. Pour rappel, nous l'avons vu dans le chapitre précédent, *Le Médiéviste et l'ordinateur* n'apparaît pas au sein d'une communauté scientifique préexistante, mais s'articule avec la volonté de créer une communauté, qui passe par le partage de savoirs et de savoir-faire apparus récemment dans la discipline et liés à l'utilisation des ordinateurs. Les fonctions de ces énoncés peuvent donc être utilisées comme des indices pour saisir les problématiques de circulation des connaissances qui traversent cette communauté en cours de développement.

L'analyse qui suit propose de mettre en relief les particularités du *Médiéviste et l'ordinateur*, par contraste avec trois autres corpus construits sur la base de trois revues référentes en histoire dans le contexte académique français. Nous qualifierons deux d'entre elles de revues généralistes, du fait qu'elles publient des articles présentant des études mettant en œuvre tout type de méthodes : 1) les *Annales. Économies, Sociétés, Civilisations*, célèbre revue créée en 1929 par Lucien Febvre (1878-1956) et Marc Bloch (1886-1944) afin de promouvoir une histoire en dialogue avec les sciences sociales ; 2) les *Cahiers de Civilisation Médiévale*, revue créée en 1958 par le Centre d'études supérieures de Civilisation médiévale de Poitiers, dans le but de « donner une vision exacte » de la civilisation médiévale par la publication d'articles de toutes les disciplines s'intéressant à cette période (philologie, archéologie, philosophie, droit, etc.)⁷⁴⁰. Nous

⁷⁴⁰ René Crozet et Edmond-René Labande, « Liminaire », *Cahiers de Civilisation Médiévale* 1, n° 1 (1958): 5-7.

désignerons la troisième comme une revue spécialisée, en raison du fait que ses éditeurs la présentent comme « *un espace d'échange d'expérience, en particulier autour de nouvelles méthodes de travail induite par l'informatique* »⁷⁴¹ : 3) *Histoire & Mesure*, créée en 1986 par la section SHS du CNRS sous l'impulsion de Jean-Philippe Genet (1944-), Jean-Claude Hocquet (1936-), André Straus (1948-) et Alain Guerreau (1948-) ⁷⁴².

Ces trois corpus sont constitués comme suit :

- 1) 60 articles des *Annales. ESC*, publiés entre 1977 et 1990, comportant tous au moins un tableau, un graphique ou une carte avec des indicateurs quantitatifs (le critère de choix repose sur l'idée, somme toute banale, que ces formes sont des marqueurs attestant de l'utilisation de méthodes quantitatives). Ce corpus est constitué de 718 832 mots, pour 75 auteurs différents, et concerne toutes les périodes historiques et tout type de problématiques historiques (culturelle, sociale, économique, etc.) ;
- 2) 99 articles publiés par les *Cahiers de Civilisation Médiévale* entre 1979 et 1989 (soit la moitié des articles publiés sur la période), soit 836 763 mots pour 103 auteurs. Ces articles concernent exclusivement la période médiévale ;
- 3) 104 articles publiés entre 1986 et 1989 dans *Histoire & Mesure* (revue créée en 1986), soit 842 379 mots, pour 80 auteurs. Ces articles peuvent concerner toutes les périodes historiques, et proposent pour la grande majorité d'entre eux des études faisant usage de méthodologies quantitatives, ou de méthodes impliquant de la mesure (dendrochronologie, etc.).

Les raisonnements successifs nous conduiront à constituer des sous-corpus, qui prendront des formes variables (du regroupement de seulement quelques articles au corpus entier). Cette méthode permettra de mieux comprendre comment les auteurs rendent compte de

⁷⁴¹ Voir <https://journals.openedition.org/histoiremesure/1059>

⁷⁴² Le comité de Rédaction, « Trente ans d'*Histoire & Mesure*. Entretiens croisés avec Gérard Béaur, Jean-Philippe Genet et Jean Heffer », *Histoire & mesure* XXXI, n° XXXI-2 (2016): 3-9.

leurs expériences avec les ordinateurs entre 1979 et 1989 et d'observer comment ils s'adaptent à ces différents contextes de publication, c'est-à-dire aux exigences éditoriales et aux différents lectorats.

Le corpus d'articles extraits du *Médiéviste et l'ordinateur* réunit les contributions publiées dans une des rubriques du bulletin, intitulée « *Article de fond* », que les éditeurs définissent dans l'éditorial du premier numéro comme « *des descriptions détaillées d'un projet* »⁷⁴³. Ce corpus est constitué de 77 articles, écrits par environ 80 auteurs. Par comparaison avec les autres rubriques du bulletin (« *Note technique* », « *Information* », « *Voyage au centre de calcul* », « *Courrier des lecteurs* »), les articles issus de la rubrique « *Article de fond* » représentent, en nombre de mots, la part la plus importante des textes du *Médiéviste et l'ordinateur*, soit près de 40% du bulletin.

Nous nous intéresserons dans un premier temps à un corpus constitué autour d'une seule auteure, la médiéviste Élisabeth Carpentier (1933-), qui apparaît dans notre contexte comme l'une des premières utilisatrices des méthodes informatiques en histoire médiévale. Cette étape du raisonnement permettra de soulever de nombreuses questions relatives à l'importance du contexte de publication sur l'exposé des méthodes informatiques en histoire.

Nous analyserons ensuite les fonctions attachées à certaines parties de textes, que nous qualifierons de parties méthodologiques. Pour ce faire, nous nous reposerons sur un sous-corpus constitué uniquement d'études utilisant des méthodes statistiques ou informatiques en histoire médiévale, publiés dans les trois revues académiques cités ci-dessus.

Nous terminerons notre étude en développant, sur la base des conclusions des deux premières parties, une analyse des spécificités des articles du *Médiéviste et*

⁷⁴³ Caroline Bourlet et al., « Éditorial », *Le médiéviste et l'ordinateur* 1, n° 1 (1979): 2.

l'ordinateur, afin de dégager les fonctions propres au genre d'articles qui apparaît dans le bulletin entre 1979 et 1989.

1. Une auteure, quatre contextes de publication.

Quelles différences présentent les articles issus d'études utilisant l'ordinateur en fonction du type de périodique dans lequel ils sont publiés ? Pour répondre à cette question, nous proposons ici de nous intéresser dans un premier temps à une auteure, en comparant quelques-uns de ses articles, publiés dans des revues académiques généralistes, avec ceux qu'elle a fait paraître dans le *Médiéviste et l'ordinateur*. Cette méthodologie nous permettra de repérer des indices qui nous conduiront dans un second temps à caractériser le genre d'articles qui se développent dans les pages du bulletin de liaison.

La médiéviste qui fera l'objet de notre analyse se nomme Carpentier. Née en 1933, elle est l'une des pionnières de l'utilisation de l'informatique en histoire médiévale en France, avec la réalisation d'une thèse d'État sur la Romanie Génoise soutenue en 1975⁷⁴⁴. Nous avons rassemblé ici quatre de ses articles, parus dans quatre périodiques différents, qui dérivent tous d'études utilisant des ordinateurs et qui concernent deux études distinctes : une analyse lexicométrique de biographies royales (XI^e-XIII^e siècle) et son travail sur un cadastre de la ville d'Orvieto de la fin du XIII^e siècle)⁷⁴⁵.

⁷⁴⁴ Cette thèse a été soutenue sous le titre *Orvieto et son « contado »*. *Étude du Cadastre de 1292 par la méthode informatique* en 1975. Les traitements informatiques sont donc, bien entendu, antérieurs. La publication de ce travail n'intervient quant à elle qu'en 1986 : Élisabeth Carpentier, *Orvieto à la fin du XIII^e siècle: ville et campagne dans le cadastre de 1292* (France, 1986).

⁷⁴⁵ Les articles dont il est question ici sont les suivants, cités dans leur ordre d'apparition : Élisabeth Carpentier, « Histoire et informatique. Recherches sur le vocabulaire des biographies royales françaises », *Cahiers de Civilisation Médiévale* 25, n° 97 (1982): 3-30 ; Élisabeth Carpentier, « L'homme, les hommes et la femme. Étude sur le vocabulaire des biographies royales françaises (XI^e-XIII^e siècle) », *Annales* 41, n° 2 (1986): 325-46 ; Élisabeth Carpentier, « Le paysage rural du Contado d'Orvieto à la fin du XIII^e siècle d'après le cadastre de 1292 », *Actes des congrès de la Société des historiens médiévistes de l'enseignement supérieur public* 10, n° 1 (1979): 177-89 ; Élisabeth Carpentier, « Cadastres d'Orvieto », *Le médiéviste et l'ordinateur* 3, n° 1 (1980): 5-7.

1.1. Étude du vocabulaire des biographies royales françaises (1982-1986).

Le premier couple d'articles choisi ici porte sur une recherche menée à partir de quatre biographies royales françaises, rédigées en latin entre le XI^e et le XIII^e siècle., qui sont publiés tous les deux dans ce que Zysberg pourrait appeler des « *revues spécialisées d'histoire* » : les *Cahiers de Civilisations Médiévales* et les *Annales. ESC*. Nous comparerons ici les modalités de présentation des discours portant sur l'utilisation des ordinateurs pour évaluer dans quelle mesure cette auteure développe des considérations différentes en fonction des contextes de publications.

1.1.1. Un article dans les *Cahiers de Civilisation Médiévale*.

Le premier article est publié dans les *Cahiers de Civilisation Médiévale* en 1982⁷⁴⁶. Carpentier y étudie le vocabulaire du corpus de quatre biographies mentionnées ci-dessus, en s'intéressant tout particulièrement au nombre d'occurrences de certains termes. Elle compare alors les fréquences de ces termes en fonction des textes et des périodes, par type de mots (substantif, adjectif et verbe), par cas (nominatif, vocatif, accusatif, génitif, datif, ablatif) et par nombre (pluriel et singulier).

L'étude s'insère dans un contexte historiographique marqué par l'intérêt particulier que les médiévistes portent aux termes rares et originaux⁷⁴⁷. À rebours de cette tendance, l'auteure propose d'utiliser des méthodes informatisées pour se pencher sur les termes les plus fréquemment employés dans ces biographies. Au regard de ses premiers résultats, la deuxième partie de son article porte sur l'étude du terme *rex* (roi) et des mots de la même famille.

On remarque, à la lecture de ce texte, que, si son titre contient le terme informatique, les mentions à l'utilisation des ordinateurs dans le corps de l'article sont

⁷⁴⁶ Carpentier, « Histoire et informatique. Recherches sur le vocabulaire des biographies royales françaises ».

⁷⁴⁷ À ce titre, Carpentier cite notamment, dans le second article dont il sera question ici, les études suivantes sur l'utilisation du mot *miles* : Johann Johrendt, « *Milites* » und " *Militia* " im elften Jahrhundert: *Untersuchung zur Frühgeschichte des Rittertums in Frankreich und Deutschland* (na, 1971); Pierre Van Luyn, « Les milites dans la France du XI^e siècle: examen des sources narratives », *Le Moyen-Âge*, 1971; Jean Flori, *La notion de chevalerie dans les chansons de geste du XII^e siècle: étude historique du vocabulaire*, 1975; Georges Duby, « La diffusion du titre chevaleresque sur le versant méditerranéen de la chrétienté latine », *La Noblesse au Moyen-Âge*, 1976, 42-52.

extrêmement rares. À deux reprises, elle signale l'intérêt de ce nouvel outil pour sa recherche, mais sans jamais en préciser le rôle dans l'analyse. Elle affirme seulement, dans l'introduction :

« Les possibilités offertes par l'ordinateur permettent maintenant, à partir de ces instruments simples que sont les index et les concordances, de procéder à la fois à l'analyse globale du vocabulaire d'une œuvre et à l'analyse fine et approfondie de champs sémantiques déterminés. »⁷⁴⁸.

Cette première occurrence de la nature des documents de travail utilisés pour mener cette étude (index et concordances) ne dit cependant rien ni sur les procédés qui ont conduit à leur élaboration, ni sur les méthodes de traitement des données qui sont mises en œuvre pour étudier le vocabulaire et les champs sémantiques.

Plus loin, dans le bilan de la première partie de son analyse, on trouve une autre remarque qui éclaire le rôle de l'informatique dans l'étude :

« Pour résumer la première partie de cet article, il nous semble avoir montré que l'étude statistique du vocabulaire d'un texte historique — non pas la recherche des mots rares, mais, au contraire, celle du vocabulaire de haute fréquence — , désormais facilitée par l'existence d'index établis automatiquement, donne de ces textes une image fidèle, conforme à la fois aux intentions de l'auteur et à l'univers dont il fait partie. »⁷⁴⁹.

Le lien entre les index et les traitements statistiques mis en œuvre se distingue ici un peu plus précisément. Toutefois, Carpentier reste très loin d'une explication qui doterait le lecteur des moyens lui permettant de comprendre comment ces outils de travail, clefs pour l'analyse rappelons-le, sont fabriqués « *automatiquement* ». Elle laisse par ailleurs dans l'ombre l'utilisation qu'elle développe des concordances et le rôle des ordinateurs dans leur élaboration.

⁷⁴⁸ Carpentier, « Histoire et informatique. Recherches sur le vocabulaire des biographies royales françaises », 4.

⁷⁴⁹ Carpentier, 19.

À la lumière de ces deux citations, qui constituent les seules mentions relatives aux usages des ordinateurs dans le corps de l'article, on observe clairement que les aspects informatiques de l'étude ne s'intègrent que très discrètement dans les propos de l'auteure. L'accent est plutôt mis sur les apports de la méthodologie statistique, « facilitée » par l'utilisation des moyens informatiques, qui est, quant à elle, explicitée.

Pour prendre un exemple, le choix des mots pris en compte dans les décomptes est expliqué en ces termes :

« Pour ce texte, comme pour les suivants, nous avons procédé à une distinction entre les mots dits « mots-outils » et les « mots significatifs » ou « mots pleins », et nous n'avons retenu pour étude que ces derniers : les substantifs, les adjectifs, les verbes (à l'exception du verbe « être ») et éventuellement, mais ils sont très peu nombreux dans nos textes, les adverbes composés à partir d'adjectifs. »⁷⁵⁰.

De la même manière, l'auteure décrit dans la suite de son texte les opérations nécessaires à la construction de catégories permettant l'étude des termes précédant les occurrences du mot *rex* :

*« Il peut être intéressant de rechercher quels sont les antécédents les plus fréquents du mot *rex*. Mais la diversité est ici la règle, la plupart de ces antécédents n'apparaissant qu'une seule fois ; ce qui nous impose la nécessité de regroupements autour de quelques thèmes. Trois peuvent être représentés ici : la personne physique du roi et ses vertus ; la famille du roi ; le service et les hommes du roi. »⁷⁵¹.*

L'exposé de la méthode n'est donc pas absent du texte, loin de là, mais les considérations qui regardent l'équipement mobilisé pour mettre en œuvre cette méthode - les ordinateurs - sont quant à elles exclues du corps du texte.

Deux indices nous montrent cependant que cette absence n'est pas insignifiante pour notre recherche. En effet, Carpentier cite, dans les notes de bas de pages, deux

⁷⁵⁰ Carpentier, 4.

⁷⁵¹ Carpentier, 24.

références qui concernent plus directement ces problématiques informatiques. La première renvoie à l'une des annexes d'un livre de deux médiévistes français, Jean-Pierre Poly (1941-) et Éric Bournazel (1924-), intitulé *La mutation féodale (X^e-XII^e siècle)*. Dans cette annexe, intitulée « Lexicographie, typologie, informatique »⁷⁵², les auteurs reviennent sur les problèmes de vocabulaire qui intéressent les spécialistes du bas Moyen-Âge, et sur les approches informatiques qui ont été développées pour les examiner.

La place de ce discours dans le livre de Poly et Bournazel nous donne un premier indice sur les contextes de publication des développements portant sur les méthodes informatiques : le cœur du livre ne s'attarde pas sur ce propos, qui est rejeté dans les annexes. Carpentier cite donc, en marge de son article, un texte lui-même séparé du corps du livre dans lequel il apparaît. Tout semble donc indiquer que les problèmes pratiques liés à l'usage des ordinateurs se discutent en des endroits distincts des résultats des études.

À ce premier indice, il faut en ajouter un autre, identifié lui aussi dans l'article de Carpentier. Au début de la seconde partie de ce texte, Carpentier propose deux tableaux, qui permettent de comparer le nombre d'occurrences de certains termes dans les quatre textes de l'étude. En note de bas de page, elle indique :

*« Il ne s'agit ici que d'une ébauche. Le travail complet devrait porter sur l'ensemble du vocabulaire lemmatisé des quatre textes, qui ferait ensuite l'objet de différents types d'analyse factorielle (cf. les différents art. parus dans « Le médiéviste et l'ordinateur », n° 5, printemps 1981). »*⁷⁵³

Cette référence est, elle-aussi, hautement significative pour les questions qui nous intéressent ici, puisque Carpentier guide encore une fois le lecteur vers une autre publication, dans laquelle les problèmes techniques peuvent être discutés et précisés. Ce nouvel indice suggère que le contexte de publication de cet article de 1982 ne paraît pas, pour l'auteure, adapté aux développements des problématiques proprement informatiques, au contraire d'un autre périodique : *Le Médiéviste et l'ordinateur*.

⁷⁵² Cette partie de texte, figure aux pages 19 à 23 du livre suivant : Jean-Pierre Poly et Éric Bournazel, *La mutation féodale: X^e-XII^e siècles* (Paris, France: Presses universitaires de France, 1980).

⁷⁵³ Carpentier, « Histoire et informatique. Recherches sur le vocabulaire des biographies royales françaises », 16.

1.1.2. Un article dans les *Annales. ESC*.

Le second article qui nous intéresse est publié dans la revue *Les Annales ESC* quatre ans plus tard, en 1986⁷⁵⁴. Il porte sur le même corpus, et propose la même méthodologie. Cependant, à la différence du précédent, Carpentier s'intéresse non plus au terme *rex*, mais au terme *vir*, qui désigne en latin aussi bien l'homme adulte que l'individu socialement supérieur⁷⁵⁵. Cette fois, l'informatique a disparu du titre, et les mentions à cet aspect du travail se font toujours aussi rares.

Dans l'introduction, Carpentier affirme :

*« Le développement des études de vocabulaire accompagné du dépouillement systématique de corpus constitués par informatique est un des traits marquants de la recherche actuelle en histoire médiévale. S'agit-il d'une sorte de retour au passé qui permet aux médiévistes - sous le couvert de nouvelles techniques et à l'abri de quelques résultats chiffrés - de revenir inlassablement sur leurs sources traditionnelles, chartes et chroniques... ? Ou d'une évolution porteuse d'avenir, susceptible de fournir des bases sûres à une histoire sociale et à une histoire des mentalités en pleine transformation ? Pour répondre ce problème de fond, il ne suffit pas de constituer des corpus, il faut apprendre à les utiliser, à les interpréter : là nous n'en sommes encore, me semble-t-il, qu'à l'aube de la recherche. »*⁷⁵⁶

Contrairement au premier article, Carpentier détaille ici la place de l'informatique au sein de sa méthodologie : ses liens avec les opérations de constitution du corpus et d'enregistrement des données. L'énoncé reste néanmoins très général, quoique légèrement plus développé quant à l'intérêt de l'informatique pour les questions historiographiques qui l'intéressent.

On retrouve le même type de déclaration générale dans la conclusion :

⁷⁵⁴ Carpentier, « L'homme, les hommes et la femme. Étude sur le vocabulaire des biographies royales françaises (XI^e-XIII^e siècle) ».

⁷⁵⁵ Carpentier, 326.

⁷⁵⁶ Carpentier, 325.

« Mais puisque nous en sommes réduits à percevoir ces mondes inférieurs au deuxième degré, en filigrane, les études de vocabulaire, facilitées par la puissance et la finesse de nouvelles méthodes d'analyse, peuvent constituer des instruments d'approche privilégiés »⁷⁵⁷.

Quatre ans après le premier article, les mentions portant sur les pratiques de l'informatique dans le corps de l'article sont toujours circonscrites à l'introduction ou aux conclusions. Elles sont cependant moins timides, et affirment clairement l'intérêt de l'informatique pour l'histoire sociale et l'histoire des mentalités. L'auteur y appelle même au développement plus large de ces méthodes : la revue académique devient dans ce cas le lieu d'une promotion de ces outils.

Par contraste, il est possible de constater que les précisions sur la mise en œuvre de ces méthodes sont une fois de plus rejetées dans les notes de bas de page⁷⁵⁸. Une première référence renvoie ainsi, dès l'introduction, non seulement au colloque de Rome de 1975, mais aussi, comme dans l'article précédent, au *Médiéviste et l'ordinateur*. Une seconde note de bas de page a plus particulièrement retenu notre attention. À l'inverse du premier article, dans lequel il n'était nulle part fait mention ni du programme informatique utilisé, ni du centre de calcul fréquenté, Carpentier renvoie, dans la note 4, à un autre article, portant précisément sur le programme utilisé :

« Index et concordances ont été établis à Poitiers par le programme appelé PECEI (Programme d'Établissement de Concordances et d'Index) cf. PELLEN et CL SIREDEY « PECEI un système de dépouillement de textes », dans *La Licorne (Poitiers)*, 1980, n°4, pp 121-141 »⁷⁵⁹.

Notons d'emblée que, là encore, l'article provient d'une publication extérieure à la discipline historique, à savoir, une revue de langue et de littérature française.

Une fois de plus, l'espace approprié à la discussion sur les aspects purement techniques se situe en dehors du corps du texte, et même, dans ce cas, en dehors de la

⁷⁵⁷ Carpentier, 343.

⁷⁵⁸ Notons cependant ici qu'à la différence d'autres disciplines comme la linguistique, les articles d'histoire font souvent état d'une utilisation massive et fréquente des notes de bas de page.

⁷⁵⁹ Carpentier, 344.

discipline. Pour autant, Carpentier propose ici une référence qui permet aux lecteurs de se documenter plus précisément sur sa méthode, et, pourquoi pas, de s'engager à leur tour dans une analyse automatique de vocabulaire à l'aide du programme PEVI. Au contraire du premier article, l'auteure donne au lecteur la possibilité de critiquer comme de répliquer l'expérience.

1.2. Étude du cadastre d'Orvieto de 1292.

Tous les traits que nous avons mis en évidence se retrouvent à l'extrême dans le second couple d'articles de notre corpus, qui concernent tous les deux une étude menée antérieurement par Carpentier : l'exploitation, assistée par ordinateur, d'un cadastre de la ville d'Orvieto (Ombrie, Italie), daté de l'année 1292. Cette source, constituée de deux registres, recense des informations sur 6 305 propriétaires qui possèdent au total 52 131 parcelles de la ville et de son *contado*⁷⁶⁰.

1.2.1. Quelques points communs entre un article dans la *Revue du Nord* et un article dans le *Médiéviste et l'ordinateur*.

Le premier article est publié en 1979 dans les actes du 10^e congrès de la Société des historiens médiévistes de l'enseignement supérieur public, consacré aux réalités et aux représentations du paysage rural⁷⁶¹. Il sera republié l'année suivante, sous la même forme, dans la *Revue du Nord*, une revue académique d'histoire et d'archéologie⁷⁶². Dans ce premier article, on ne retrouve tout simplement aucune mention relative à l'aspect informatique de l'étude, ni dans le corps du texte, ni dans les notes de bas de page. L'auteur présente pourtant bien sa méthodologie, en développant d'un côté l'aspect linguistique de son étude, puis l'exploitation des données portant sur les propriétaires et les parcelles.

⁷⁶⁰ Carpentier, « Le paysage rural du Contado d'Orvieto à la fin du XIII^e siècle d'après le cadastre de 1292 », 177.

⁷⁶¹ Carpentier, « Le paysage rural du Contado d'Orvieto à la fin du XIII^e siècle d'après le cadastre de 1292 ». Les actes du colloque sont les suivants : *Actes des congrès de la Société des historiens médiévistes de l'enseignement supérieur public, 10^e congrès, Lille, 1979. Le paysage rural : réalités et représentations.*, vol. 10, 1979.

⁷⁶² Elisabeth Carpentier, « Le paysage rural du Contado d'Orvieto à la fin du XIII^e siècle d'après le cadastre de 1292 », *Revue du Nord* 62, n° 244 (1980): 177-89.

Le deuxième article est publié l'année suivante dans *Le Médiéviste et l'ordinateur*. Au contraire du premier, il est très court (seulement trois pages, par opposition avec 15 pour le premier). Il faut préciser qu'il ne comporte aucune note de bas de page. À ces premiers indices formels sur la différence de nature entre les deux textes, il faut encore ajouter que le contenu de ce second article change du tout au tout par rapport au premier. En effet, aucun des résultats de l'étude présentée dans le premier article n'apparaît dans le second. Seules les premières phrases sont similaires dans leurs contenus, puisqu'elles se composent exactement des mêmes éléments visant à décrire la source : le nombre de folio, ou registres, le nombre de propriétaires pour la cité et pour le *contado*, et enfin, le nombre total de parcelles. Cette description est cependant plus précise dans l'article du *Médiéviste et l'ordinateur*, puisque Carpentier détaille, en plus, la distribution des informations dans chacun des registres.

1.2.2. Dimensions techniques versus résultats.

Après cette courte introduction, les différences deviennent flagrantes, puisque ce deuxième article traite exclusivement de la méthode informatique. Le deuxième paragraphe se présente par exemple comme la justification du choix d'utiliser des ordinateurs pour mener cette étude. Cette explication se structure en deux éléments : Carpentier met en avant 1) la masse des données qu'offrent ce corpus, ainsi que 2) l'homogénéité de l'ensemble documentaire auquel elle s'intéresse, qui permettait un codage « *aisé* ». Dans ce cadre, elle insiste toutefois, quoiqu'en seulement une phrase, sur les obstacles qu'elle a rencontrés quant aux noms de lieux et de personnes, dont elle a dû harmoniser les dénominations pour en permettre un traitement informatique.

L'auteure se tourne ensuite vers la description de sa méthode d'enregistrement des données, qui constitue l'essentiel de ce court article. Elle y présente la structure du fichier qu'elle a élaboré, constitué de trois types de cartes perforées IBM à 80 colonnes : deux cartes par parcelles, et une carte par propriétaires de parcelle. Pour détailler le contenu et la fonction de chacune de ces cartes, elle présente tout d'abord, colonne par colonne, les informations que doivent accueillir chacune de ces 80 colonnes. Pour le premier type de cartes, elle écrit par exemple :

« - col. 1 à 12 : système d'identification. Soit le numéro du registre (col. 1), du castrum ou "piviere" (col. 2-3), de la villa (col. 4-5), du propriétaire (col. 6-8), de la parcelle (col. 9-11), dans l'ordre rigoureux de l'apparition au Cadastre. En col. 12 : I (cartes de type I).

- col. 13 à 21 : les propriétaires de la parcelle. Soit la nature du feu propriétaire : simple, multiple, masculin, féminin... (col. 16) ; les étrangers (col. 18) ; les mentions particulières dominus, magister ... (col. 19-20).

- col. 22 à 80 : la parcelle. Soit la nature, localisation, toponyme, nombre de confins, superficie, valeur, mentions particulières... Toutes ces données, sauf les toponymes, sont codées numériquement. »⁷⁶³

Carpentier associe ensuite à cette description un développement sur les programmes informatiques mis en œuvre sur la base de chacun de ces types de cartes, en insistant sur les résultats qu'ils ont permis d'obtenir. Pour prendre seulement un exemple, elle affirme pour le deuxième type de cartes :

« Ces cartes ont permis, par un système de tris, de dresser des listes de propriétaires et de confins, et, surtout, par ces dernières, de reconstituer partiellement et de localiser la propriété ecclésiastique, absente du Cadastre. »⁷⁶⁴

Le contenu de ce court article se distingue ainsi totalement des trois articles précédemment évoqués. En effet, ce dernier texte renseigne les lecteurs autant sur la structure du fichier qu'elle a élaboré que sur les données qu'elle a sélectionnées pour mener son étude. Cette présentation s'appuie non pas, comme dans les articles précédents, sur les résultats des traitements informatiques et statistiques, mais au contraire sur l'inscription matérielle de ces opérations, qui reposent sur l'élaboration de cartes perforées. Dans l'article du *Médiéviste et l'ordinateur*, les artefacts de la pratique des

⁷⁶³ Carpentier, « Cadastres d'Orvieto », 6.

⁷⁶⁴ Carpentier, 6.

ordinateurs sont ainsi mis au cœur de l'exposé. En ce sens, et nous y reviendrons longuement, elle permet aux lecteurs de mieux comprendre les opérations proprement informatiques de l'enregistrement, mais aussi d'en avoir une idée concrète.

La conclusion de ce court article de Carpentier mérite également l'attention. Au contraire des trois autres textes considérés, qui n'abordaient que très rapidement la question des pratiques de l'informatique, la conclusion de celui-ci est entièrement tournée vers ce problème :

« Au total, dans un cas comme celui-ci, seul l'ordinateur permettait d'obtenir à la fois des conclusions d'ensemble et surtout des analyses fines. Il n'en reste pas moins que l'utilisation de l'informatique demande un travail considérable, d'abord au niveau du choix très délicat et de l'organisation des données qui sont retenues, harmonisées, codées en vue de leur traitement informatique, ensuite au niveau de l'élaboration des programmes qui vont permettre la constitution d'un nouveau corpus de données. Ce corpus représente l'apport irremplaçable de l'ordinateur ; si les programmes ont été ambitieux, il peut se révéler très lourd à manier et par sa masse même, difficile à interpréter. Alors commence véritablement le travail de l'historien. »⁷⁶⁵

De la même manière que pour le couple d'articles précédents, l'auteure met en avant l'intérêt des ordinateurs pour la recherche en histoire médiévale.

Ici, elle va cependant plus loin, abordant les difficultés liées à l'utilisation de ces nouveaux instruments. Elle indique notamment aux lecteurs que leurs usages demandent « *un travail considérable* », que les choix qu'impliquent ces méthodes sont « *très délicats* », que les corpus ainsi constitués peuvent « *se révéler très lourd à manier* » ou que les résultats issus des programmes peuvent être « *difficile[s] à interpréter* ». L'emploi de cette série d'adjectifs, qui permet à l'auteure de qualifier les difficultés que pose l'utilisation de l'informatique pour les médiévistes à cette période, caractérise la spécificité de l'article du *Médiéviste et l'ordinateur* par contraste avec les trois autres étudiés plus haut.

⁷⁶⁵ Carpentier, 7.

1.3. Premières conclusions : esquisse d'une formation discursive.

L'analyse de ce corpus constitué autour d'une seule auteure permet d'esquisser les contours de la formation discursive constituée des discours portant sur l'utilisation des ordinateurs en histoire médiévale.

1.3.1. Genre et sous-genre d'articles en histoire quantitative.

Un premier constat s'impose : la mention des aspects proprement informatiques des études se fait en dehors des *revues académiques généralistes* (*Cahiers de Civilisation Médiévale*, *Annales. ESC*, *Revue du Nord*). En effet, si ces dernières permettent dans certains cas de partager l'intérêt ou la nécessité du recours à l'informatique en histoire, voire d'en faire la promotion, les développements méthodologiques et pratiques semblent réservés à d'autres contextes de publication, et, notamment, au bulletin de liaison *Le Médiéviste et l'ordinateur*.

Notons toutefois, comme on a pu le voir, que les précisions sur la méthodologie statistique ne sont pas absentes de ces revues. Dans le cas de ce premier corpus, l'auteure développe par exemple les raisonnements mis en œuvre, du choix des sources à la construction des catégories. Elle publie par ailleurs dans ces articles des tableaux et des graphiques qui présentent des résultats de tris croisés ou d'autres opérations statistiques.

Une comparaison non formalisée avec d'autres articles d'histoire, qui présentent, eux aussi des études mobilisant des méthodes statistiques, nous autorise à faire le constat d'une forme de régularité dans la présentation des raisonnements, mais aussi des résultats. En schématisant, il est possible d'affirmer que, dans ce type de texte, les auteurs formulent des hypothèses qu'ils se proposent d'explorer par des analyses statistiques (tris, classement, etc.). Les résultats sont présentés sous la forme d'un tableau ou d'un graphique, et conduisent à formuler de nouvelles questions, qui conduisent à l'expression de nouvelles hypothèses. Le raisonnement se poursuit étape par étape.

Le plan-type que nous dégagons ici semble indiquer qu'il serait possible d'identifier un genre d'article d'histoire propre à l'histoire dite quantitative⁷⁶⁶. Les articles qui présentent des études mettant en œuvre des ordinateurs pourraient ainsi être inclus dans ce genre d'article « histoire quantitative », en se distinguant du fait que les auteurs ajouteraient parfois, en introduction ou en conclusion, de brèves mentions à l'usage des ordinateurs.

1.3.2. Définition et application du critère de répliquabilité.

Intéressons-nous maintenant aux contenus proprement informatiques des articles de notre étude de cas. Quelles autres pistes d'analyse peut-on dégager ? Plusieurs points apparaissent dans les deux sous-parties précédentes. On constate d'abord que les parties de textes qui portent sur l'utilisation des ordinateurs sont plus ou moins développées en fonction des contextes de publication. De l'absence pure et simple de mention relative à l'utilisation de l'informatique dans la *Revue du Nord*, à l'article entièrement dédié aux opérations de saisie des données et à leur exploitation dans *Le Médiéviste et l'ordinateur*, il existe certainement un éventail de formes de présentations de ces méthodes qui, comme nous l'avons vu, dépendent du contexte de publication. Dès lors, comment préciser l'analyse sur la base des remarques précédentes ?

Nous proposons ici une autre approche pour saisir autrement ces différences. Elle repose sur l'examen d'un critère que nous appellerons de répliquabilité de ces études. Pour en comprendre la logique, il faut examiner ce qui n'est pas dit dans les articles académiques, mais, au contraire, ce qui est formulé dans le *Médiéviste et l'ordinateur* ? Dans notre cas, supposons que nous voulions reproduire point par point la méthode de Carpentier, pour en produire une autre, sur d'autres biographies en latin par exemple. Sur la base des deux articles publiés dans les *Cahiers de Civilisation Médiévale* et *Les Annales. ESC*, on ne peut que faire le constat d'une impossible répliquabilité de la méthode de cette médiéviste. En effet, trop d'éléments restent dans l'ombre pour le lecteur.

Par exemple, le lecteur pourrait essayer de reproduire les opérations permettant d'obtenir une liste des substantifs en fonction de leur nombre (singulier ou pluriel). Pour

⁷⁶⁶ Cette affirmation nécessiterait, pour être plus solide, une étude complète, portant sur un corpus d'article plus large. Cependant, ce n'est ni l'objet de ce chapitre, ni de cette thèse. On se contentera donc ici de soulever la question, et d'annoncer que ce travail est en cours.

cela, il devrait savoir comment Carpentier a indexé son texte, mais aussi si celui-ci a été lemmatisé. Il lui faudrait alors connaître, entre autres, le dictionnaire utilisé pour la lemmatisation. Or, nous l'avons vu, les réponses à ces questions sont rejetées, au mieux, hors du texte, dans une référence en notes de bas de page. À la lecture des seuls articles des revues académiques généralistes, le lecteur ne peut donc pas répliquer l'entreprise. Il n'a d'ailleurs pas non plus appris grand-chose sur les conditions d'utilisation ou les pratiques de l'informatique en histoire médiévale.

Pour l'étude du cadastre d'Orvieto, nous l'avons vu, la situation est encore plus complexe, puisque l'utilisation de l'ordinateur n'est même pas mentionnée dans l'article de la *Revue du Nord*. Au contraire, dans l'article du *Médiéviste et l'ordinateur*, le propos est tourné entièrement vers le développement des opérations matérielles (utilisation de cartes perforées, répartition des informations dans les différentes colonnes, codages). Il fournit à Carpentier l'occasion de partager avec le lecteur la succession d'opérations matérielles attachées à une partie centrale de la méthode informatique : la saisie et le codage des données extraites du cadastre. Cet article peut ainsi être caractérisé comme le complément essentiel du premier. De ce fait, on peut affirmer que l'objectif visé par ce texte est tout autre : il permet aux lecteurs de saisir les opérations matérielles qui caractérisent l'utilisation des ordinateurs pour l'étude de cette source en donnant la « recette » de cette étude.

Tirant les leçons de cette application du critère de répliquabilité que nous avons défini ici, nous distinguerons désormais schématiquement, d'une part, les parties des articles (ou articles entiers) portant sur les résultats et, d'autre part, les parties des articles (ou articles entiers) portant sur les méthodes. Cette distinction permettra, comme nous le verrons, de nous donner les outils pour constituer des corpus d'articles à même de saisir les fonctions particulières des « articles de fond » du *Médiéviste et l'ordinateur* pour les membres de cette communauté discursive.

2. Parties méthodologiques.

Ces premières conclusions formulées sur la base d'un corpus constitué autour d'une médiéviste et de deux enquêtes nous invitent à étudier un corpus plus large. Cette opération permettra de poser de manière plus générale les questions suivantes : les parties

méthodologiques portant sur l'usage des ordinateurs sont-elles toujours rejetées en dehors du corps des articles publiés dans les revues académiques que nous avons qualifiées de généralistes ? Si non, de quels éléments sont-elles composées ? Et qu'en est-il dans un contexte de publication spécialisée comme *Histoire & Mesure* ? Quelles sont les *finalités* de ces parties méthodologiques, c'est-à-dire quelles modifications de la situation dans laquelle ils s'inscrivent ces discours visent-ils ? Et la réponse à cette question dépend-elle des différents types de contextes de publication susmentionnés ? Nous verrons enfin, dans la troisième partie de ce chapitre, quelles sont, à la lumière des réponses à ces questions, les particularités des articles du *Médiéviste et l'ordinateur*.

Il faut cependant faire remarquer dès à présent que de nombreuses difficultés apparaissent lorsque l'on s'intéresse aux parties méthodologiques d'un grand ensemble d'articles d'histoire. La première d'entre elles tient au fait que les articles de notre corpus ne présentent pas, comme c'est le cas dans d'autres disciplines, et particulièrement la recherche en médecine, une structure normalisée de type IMRAD (Introduction, Méthode, Résultats, Analyse, Discussion)⁷⁶⁷. Ce constat rejoint celui fait par la linguiste norvégienne Kjersti Fløttum qui a montré que les articles en sciences humaines ne partagent pas de standards quant au plan selon lesquels ils s'organisent⁷⁶⁸. De fait, on retrouve dans ces articles une dispersion des discussions méthodologiques dans différentes parties des textes, ce qui enrichit, autant que complique, les opérations de comparaison.

La seconde difficulté que nous rencontrons pour mener ce type d'analyse repose sur la nature polymorphe de ces parties méthodologiques. Elles peuvent en effet être plus ou moins développées, mais, surtout, elles présentent des contenus très différents d'un article à l'autre. Pour le cas de notre corpus, nous l'avons vu avec l'exemple de Carpentier, l'exposé des méthodes peut se concentrer sur ses aspects statistiques ou mathématiques, ou au contraire sur des aspects purement informatiques. D'autres types de développement, que nous n'avons pas encore rencontrés, peuvent encore être envisagés. Par suite, nous considérerons comme indice de la fonction des articles, aussi

⁷⁶⁷ Voir sur ce sujet Luciana B. Sollaci et Mauricio G. Pereira, « The introduction, methods, results, and discussion (IMRAD) structure: a fifty-year survey », *Journal of the Medical Library Association* 92, n° 3 (juillet 2004): 364-71.

⁷⁶⁸ Sur ce point voir notamment Kjersti|Dahl Fløttum, Trine Dahl, et Torodd Kinn, *Academic Voices* (John Benjamins Publishing Company, 2006).

bien la dispersion des discussions méthodologiques que le type de contenus partagés dans ces parties.

Afin de poursuivre l'analyse, nous avons constitué un second corpus, composé de tous les articles d'histoire médiévale faisant explicitement mention de l'usage de l'informatique⁷⁶⁹ dans les trois revues académiques présentées en introduction. Nous appellerons désormais ce corpus CAHM 1 (pour *Cahiers de Civilisation Médiévale*, *Annales ESC*, *Histoire & Mesure*)⁷⁷⁰. Il n'est composé que de douze articles (4 articles des *Annales ESC*, 3 articles des *Cahiers de Civilisation Médiévale* et 5 articles d'*Histoire & Mesure*). Il faut ainsi d'emblée faire le constat que la mention de l'utilisation des ordinateurs reste un fait linguistique très marginal dans le champ discursif qui nous intéresse. Il est en fait hautement probable, comme le suggère fortement l'analyse menée plus haut sur les articles de Carpentier, qu'un certain nombre d'études, qui utilisent effectivement des méthodes informatiques, n'en fassent pas explicitement état dans les articles qui en présentent les conclusions.

⁷⁶⁹ L'identification de ces articles a été menée via une recherche par mots-clefs dans le corpus (informatique, ordinateur, moyen électronique, mécanographie, traitement). La lecture des articles a ensuite permis de dégager ceux qui concernaient effectivement l'utilisation d'ordinateur ou d'autres outils de calculs (41 articles), puis ceux qui concernaient effectivement des études d'histoires médiévales (12 articles).

⁷⁷⁰ Le corpus CAHM1 est constitué ainsi : 1) 4 articles des *Annales ESC* : Hélène Millet, « La composition du chapitre cathédral de Laon au XIV^e siècle : résultats d'une analyse factorielle », *Annales* 36, n° 1 (1981): 117-38 ; Alain Guerreau, « Analyse factorielle et analyses statistiques classiques : le cas des ordres mendiants dans la France médiévale », *Annales* 36, n° 5 (1981): 869-912 ; Marie-Jeanne Tits-Dieuaide, « Les campagnes flamandes du XIII^e siècle au XVIII^e siècle, ou les succès d'une agriculture traditionnelle », *Annales* 39, n° 3 (1984): 590-610 ; Carpentier, « L'homme, les hommes et la femme. Étude sur le vocabulaire des biographies royales françaises (XI^e-XIII^e siècle) ». 2) 3 articles des *Cahiers de Civilisation Médiévale* : Joachim Wollasch, « Les obituaires, témoins de la vie clunisienne », *Cahiers de Civilisation Médiévale* 22, n° 86 (1979): 139-71 ; Carpentier, « Histoire et informatique. Recherches sur le vocabulaire des biographies royales françaises »; Merritt Blakeslee, « Lo dous jocx sotils : la partie d'échecs amoureuse dans la poésie des troubadours », *Cahiers de Civilisation Médiévale* 28, n° 110 (1985): 213-22. 3) 5 articles d'*Histoire & Mesure* : Marie-Anne Polo de Beaulieu, « Étude statistique de la structure lexicale de l'exemplum médiéval (d'après la Scala Coeli de Jean Gobi) », *Histoire & Mesure* 1, n° 3 (1986): 47-80 ; Dominique Coq et Ezio Ornato, « Les séquences de composition du texte dans la typographie du XV^e siècle. Une méthode quantitative d'identification », *Histoire & Mesure* 2, n° 1 (1987): 87-136 ; Hervé Leblond, « Recherches métrologiques sur des plans de bastides médiévales », *Histoire & Mesure* 2, n° 3 (1987): 55-87 ; Alain Marie, « Histoire de la famille et outil informatique : la contribution de l'Église de Jésus-Christ des Saints des Derniers Jours », *Histoire & Mesure* 4, n° 1 (1989): 63-79 ; Marion Selz, « Procédures d'aide à la reconstitution des généalogies », *Histoire & Mesure* 4, n° 1 (1989): 3-19.

2.1. Présentation étape par étape, matérialité des opérations et expression des difficultés.

Une première lecture des douze articles du corpus CAHM 1 permet de constater que six d'entre eux font mention de l'utilisation des ordinateurs, soit très brièvement, soit dans les notes de bas de page. Au contraire, les six autres articles présentent, *a minima*, une partie de texte consacrée à la méthodologie informatique. C'est à ce second groupe d'articles que nous nous intéresserons en particulier, puisqu'eux seuls permettent de répondre aux questions soulevées ci-dessus.

Le premier de ces six articles est publié dans les *Annales ESC* en 1981. Hélène Millet (1944), membre de l'Équipe de Recherche Associée 713 (ERA 713) dirigée par Bernard Guenée (1927-2010) à l'université Paris 1 Panthéon-Sorbonne⁷⁷¹, y présente, des résultats d'analyses factorielles des correspondances en lien avec son étude, menée dans le cadre de son doctorat de 3^e cycle, de données prosopographiques d'une communauté de chanoines ayant vécu à Laon entre 1272 et 1412. Elle vise notamment à dégager la structure et la composition de cette communauté d'une manière originale, c'est-à-dire autrement que par l'établissement d'un « *portrait-robot du chanoine moyen* » ou encore par l'évocation de « *quelques chanoines soigneusement choisis parmi les plus représentatifs* »⁷⁷². Une partie entière de l'article, sobrement intitulée « Les traitements informatiques », expose explicitement les aspects informatiques de l'étude. Quatre traits caractéristiques se dégagent de l'analyse de cette partie de texte.

2.1.1. Présentation étape par étape.

Le premier concerne la structure de l'exposé de la méthode. L'auteure détaille, étape par étape, chacune des opérations de mise en œuvre de l'analyse factorielle : élaboration du questionnaire prosopographique, modalités de choix des informations à extraire des sources au moment de la saisie, opérations de recodage et, enfin, traitement statistique à l'aide d'un programme informatique. Cet énoncé sous forme de procédure peut notamment être repéré à l'aide de marqueurs de l'enchaînement des opérations, comme dans l'exemple qui suit :

⁷⁷¹ Nous rappelons ici que le chapitre 2 de cette thèse est consacré à l'un des projets de ce collectif.

⁷⁷² Millet, « La composition du chapitre cathédral de Laon au XIV^e siècle », 118.

« Dans un second temps, pour affiner les résultats, on utilise un programme de tris à croisements multiples »⁷⁷³.

On peut encore identifier dans cette partie du texte d'autres marqueurs de ce type. Les opérations de codage et de recodage sont par exemple présentées comme deux étapes distinctes du travail :

« Les questions qui se posent alors permettent d'établir les règles d'un nouveau codage, que l'on pourrait appeler codage au second degré »⁷⁷⁴.

Dans le même registre, Millet découpe par ailleurs, schématiquement, ces différentes étapes du travail en deux parties distinctes, qui s'articulent autour de l'intervention de l'ordinateur :

« Avec le remplissage du questionnaire et la transcription des réponses en langage codé sur les bordereaux s'achève la première partie du travail de l'historien : la machine entre alors en jeu »⁷⁷⁵.

Cette façon de présenter les opérations en une succession d'étapes du travail ne saurait cependant être considérée comme corrélée uniquement à l'utilisation des ordinateurs pour au moins deux raisons.

La première est que ce type de présentation se confond avec celle de l'exposé des opérations propres à l'emploi de méthodologies statistiques, et que ces méthodologies n'impliquent pas nécessairement l'emploi des ordinateurs. La seconde repose sur le fait que cette façon de présenter demeure un fait linguistique, dont il nous est permis de douter qu'il corresponde à la réalité de la pratique. Ces deux remarques nous invitent donc, pour la suite de l'analyse, à tenter d'identifier d'autres éléments à même de caractériser ces parties de textes.

⁷⁷³ Millet, 120.

⁷⁷⁴ Millet, 120.

⁷⁷⁵ Millet, 120.

2.1.2. Matérialité des opérations.

Le deuxième trait caractéristique repéré dans cette partie du texte est la présence d'énoncés qui détaillent, par opposition avec les aspects intellectuels des raisonnements, ce que nous appellerons la matérialité des opérations. Dans son article, Millet rend compte au lecteur non seulement du modèle d'ordinateur utilisé, des supports de saisie et de stockage, des formats de sorties, de la taille précise des tableaux mais également de ce que nous appellerons la dimension temporelle du travail. Les citations suivantes illustrent plusieurs de ces éléments :

« Pour mon étude je pouvais compter sur les services d'un IBM 370 avec laquelle il fallait communiquer par cartes perforées. Celles-ci étant réalisées, l'ordinateur a mémorisé sur disque (9) l'ensemble des informations collectées puis il les a retranscrites sur un listing, ce qui permet une consultation rapide du fichier pour d'éventuelles corrections ou modifications : la matière documentaire était alors prête pour subir des traitements informatiques »⁷⁷⁶.

Elle ajoute par ailleurs, en note (9) :

« Cette manipulation supplémentaire permet de gagner du temps pour la suite des opérations. »⁷⁷⁷

À ces deux exemples il est possible d'en ajouter d'autres du même type : Millet précise notamment chacune des interventions de l'ordinateur dans la mise en œuvre de la méthode d'analyse factorielle :

« Toutes ces informations ont été comptabilisées par un ordinateur qui a fourni ensuite tous les pourcentages désirés et effectué de nombreux classements »⁷⁷⁸.

⁷⁷⁶ Millet, 120.

⁷⁷⁷ Millet, 137.

⁷⁷⁸ Millet, 118.

« La répartition de chacun des chanoines dans ces différents groupes a été faite par ordinateur à partir du fichier informatique »⁷⁷⁹.

« Une fois terminée la répartition des chanoines dans ces différents groupes, l'ordinateur a fourni un listing à 850 lignes et cinq colonnes »⁷⁸⁰.

Ces multiples indications s'intègrent elles aussi dans la structure de l'exposé mis en lumière ci-dessus. Elles conduisent l'auteure à détailler le lien qui existe entre l'utilisation des moyens informatiques et la méthodologie statistique mise en œuvre. Millet détaille l'enchaînement des opérations statistiques mais éclaire aussi les éléments de leur déroulement matériel, en insistant 1) sur les opérations automatisées (calcul de « pourcentages », « classements », « répartition ») ; 2) sur les formes matérielles des documents d'entrée et de sortie (« cartes perforées », « disque », « listing ») et 3) sur la dimension temporelle de ces opérations (« rapide », « gain de temps »).

Cependant, en dépit de la forme de cet exposé et des quelques précisions matérielles qu'il éclaire, certains aspects du travail au contact des ordinateurs restent dans l'ombre. Tout d'abord, on ne retrouve dans cet article aucun détail permettant de comprendre certains aspects proprement informatiques de l'étude, et en particulier les opérations de programmation : le langage de programmation n'est pas précisé, et l'auteur ne mentionne ni le programme utilisé ni son processus d'élaboration. A-t-elle utilisé un programme préexistant ? Ou, au contraire, a-t-elle elle-même élaboré un programme permettant la mise en œuvre des analyses factorielles des correspondances ?

Ensuite, l'article ne donne aucune précision sur l'un des aspects mathématiques pourtant essentiel pour comprendre son étude : la procédure d'analyse factorielle des correspondances. Sur ce point, Millet renvoie le lecteur, dans les notes de bas de page, à des ouvrages spécialisés détaillant les modalités des opérations de calcul. Dans le corps de l'article, l'auteure évoque toutefois, au début de la deuxième sous-partie, les règles qui régissent la lecture des graphiques et l'interprétation des axes.

La nature de ces précisions, combinée à l'absence de détails sur les procédures de calcul, nous permettent de faire deux hypothèses de deux natures différentes. La première

⁷⁷⁹ Millet, 121.

⁷⁸⁰ Millet, 123.

concerne le type de lecteur visé par Millet dans cet article et donc, la finalité de cette partie du texte. L'auteure vise ici, selon nous, l'historien désireux de pouvoir lire et comprendre les résultats qu'elle présente dans son article, sans pour autant s'encombrer d'une présentation mathématique. En ce sens, l'article ne poursuit donc pas l'objectif de rendre cette méthode répliquable par d'autres, mais de communiquer les résultats de ce travail et de promouvoir ainsi l'intérêt épistémologique des analyses factorielles pour la discipline : éviter, pour les études prosopographiques, à la fois l'étude de quelques individus « *parmi les plus représentatifs* » et l'établissement d'un « *portrait-robot* » d'un individu idéal-typique.

La seconde hypothèse concerne la relation entre l'expérience vécue par Millet et la façon qu'elle a de la raconter. Dans son article, la médiéviste précise qu'elle a collaboré avec un informaticien, Jean-Paul Millet, pour réaliser mettre en œuvre les analyses factorielles des correspondances. Elle insiste en particulier sur le fait que cette aide lui a permis de sortir de « *l'impasse* » dans laquelle elle se trouvait, qui résultait des difficultés qu'elle avait pour comprendre les aspects mathématiques de cette procédure d'analyse des données. L'expérience que relate Millet dans cet article est ainsi celle d'une utilisatrice assistée d'un technicien, ce qui pourrait expliquer l'absence de présentation technique de la méthode.

2.1.3. Expression des difficultés et ressentis personnels.

Le troisième trait qui permet de caractériser cette partie de texte va dans le sens de cette dernière hypothèse. Il concerne les nombreuses références aux obstacles, aux pièges et aux difficultés que Millet a rencontrés au cours de cette expérience :

« *Une fois les règles de codage établies, il n'est plus possible d'y revenir* »⁷⁸¹.

« *Ventiler les chanoines dans ces différents groupes n'a pas été chose aisée* »⁷⁸².

⁷⁸¹ Millet, 119.

⁷⁸² Millet, 119.

« La dernière composante, celle des fonctions occupées, est assurément celle qui a causé le plus d'embarras, et le titre qu'elle porte ne reflète que très imparfaitement la réalité qu'elle recouvre »⁷⁸³.

Ces derniers exemples rendent compte d'une implication de l'auteure dans le texte qui nous apparaît inhabituelle par rapport à d'autres articles publiés dans des revues d'histoire.

Millet convoque en effet un ensemble de termes que nous pourrions rattacher au champ lexical du ressenti personnel (« *aisée* », « *embarras* »), qui lui permet notamment de justifier de l'imperfection de certaines solutions choisies. Ce type de séquences linguistiques se retrouve d'ailleurs à un autre endroit du texte de façon frappante :

« Je me suis donc efforcée de regrouper les diocèses, solution qui peut donner lieu à toutes sortes de critiques dont je n'ai que trop conscience mais qui est apparue comme étant la moins mauvaise »⁷⁸⁴.

La mention des difficultés se combine ici au quatrième trait que l'on peut repérer dans ce texte : l'utilisation presque systématique du *je* plutôt que du *nous*, qui se retrouve non pas seulement dans la partie méthodologique mais bien dans l'ensemble de l'article.

La combinaison de ces deux éléments se traduit ici par une auto-évaluation explicite de ses décisions par l'auteure, à laquelle sont associés des commentaires personnels portant sur ces dernières. Cet ensemble de faits linguistiques renforce par ailleurs l'impression d'avoir à faire à un discours personnel qui tend, nous y reviendrons, à « *raconter comment cela se passe vraiment* »⁷⁸⁵. Les deux observations que nous formulons ici, comme celles présentées plus haut, ne reposent cependant que sur un exemple, et il nous faut donc les confronter à d'autres pour en saisir la pertinence à l'échelle de notre corpus.

⁷⁸³ Millet, 123.

⁷⁸⁴ Millet, 122.

⁷⁸⁵ Nous reprenons ici l'expression choisie par Shirley Carter-Thomas et Elizabeth Rowley-Jolivet comme titre de leur intervention : Elizabeth Rowley-Jolivet et Shirley Carter-Thomas, « Raconter comment cela se passe vraiment. Comparaison entre la narration de la recherche dans les cahiers de laboratoire et dans les articles de recherche », in *Journées d'études "La mise en récit dans les discours spécialisés"* (Paris, France, 2015).

2.2. Variations dans un contexte de publication généraliste : les *Annales ESC*.

Regardons à présent les cinq articles restants et comparons-les avec celui de Millet. Le premier qui nous intéressera ici est un article du médiéviste Alain Guerreau, dont il a été plus d'une fois question depuis le début de cette thèse. Ce texte est publié lui aussi dans les *Annales ESC* en 1981, la même année que l'article de Millet. L'auteur y présente les conclusions d'une étude qui porte sur la répartition géographique et chronologique des couvents des ordres mendiants en France, qui « *fournit un assez bon calque de la carte des villes de l'époque* »⁷⁸⁶. La méthode mise en œuvre par l'auteur repose sur des analyses statistiques probabilistes et des analyses factorielles des correspondances menées sur la base d'une recension des ordres mendiants établis en France aux XIII^e et XIV^e siècle⁷⁸⁷.

La partie du texte qui nous intéresse ici est intitulée « L'analyse factorielle ». Elle comporte deux sous-parties. La première, « Évolution et problèmes d'emplois », est plus générale, et porte sur les problèmes intellectuels et matériels que pose l'utilisation de cette méthode d'analyse de données en histoire médiévale. Guerreau y justifie ses choix méthodologiques, et notamment celui de l'outil de calcul qu'il utilise : une calculatrice programmable de poche Texas Instrument. La seconde, intitulée « Mise en œuvre de l'analyse factorielle des fondations de couvents », présente rapidement les limites de la calculatrice programmable, et les logiques de codage qu'elles rendent nécessaires. L'auteur y rappelle par ailleurs l'objectif des analyses statistiques qu'il met en œuvre.

Une comparaison avec les quatre traits dégagés à partir de l'article de Millet nous permet de repérer les points communs et les différences qui existent entre les parties méthodologiques de ces deux articles.

⁷⁸⁶ Guerreau, « Analyse factorielle et analyses statistiques classiques », 869.

⁷⁸⁷ Guerreau, « Analyse factorielle et analyses statistiques classiques ». Pour comprendre l'intérêt historiographique de l'article, on pourra notamment regarder les quelques publications issues d'une étude antérieure sur le même sujet, et menée à partir de la même source, par une équipe du Centre de Recherches Historiques dirigée par Jacques Le Goff, « Avant-propos », *Publications de l'École Française de Rome* 30, n° 1 (1977): 5-5. Précisons encore ici que nous préparons un article visant à comparer ces deux études.

2.2.1. Difficultés personnelles et difficultés générales.

La première sous-partie de l'article de Guerreau s'apparente à une suite d'avertissements sur les difficultés et les pièges rencontrés par l'auteur au cours de ses expériences. Toutefois, à la différence de Millet, Guerreau n'y expose pas les difficultés qu'il a lui-même rencontrées dans la mise en œuvre de cette procédure, mais celles auxquelles se heurte plus largement le développement des méthodes d'analyses statistiques multidimensionnelles en histoire⁷⁸⁸. Il identifie notamment « *quatre groupes de problèmes* », qu'il présente en ces termes :

*« 1. les outils de calcul ; 2. l'organisation sociale des calculs ; 3. les formes et techniques de calcul ; 4. l'imposition d'une orientation méthodologique par une tendance dominante. »*⁷⁸⁹.

Sans revenir ici sur chacun de ces points, il faut cependant insister pour la suite sur le lien qu'il établit entre les outils de calcul et l'organisation sociale des calculs.

Après avoir rappelé que la puissance de calcul des ordinateurs « *défie l'imagination* », et que les manipulations de ces nouveaux outils « *restent assez ésotériques* »⁷⁹⁰, il précise :

*« Cet aspect technique est bien sûr étroitement lié à un aspect social : le coût de ces instruments, leur fonctionnement délicat ont accentué la division du travail en faisant naître en dépendance de certains pôles de pouvoir (universitaire ou administratif) un spécialiste nouveau chargé des calculs et des manipulations des outils de calcul. L'effet pervers de ce processus réside moins dans la division en soi que dans la dynamique même de la spécialisation qui entraîne le nouveau spécialiste à renforcer son autonomie en monopolisant le mieux possible le savoir-faire dont il est détenteur. »*⁷⁹¹

⁷⁸⁸ Le chapitre 6 de cette thèse est dédié à l'examen de ce sujet, à travers l'analyse d'une controverse entre Guerreau et le médiéviste Pierre Bonnassie (1932-2005), qui prend forme dans les pages du *Médiéviste et l'ordinateur*.

⁷⁸⁹ Guerreau, « Analyse factorielle et analyses statistiques classiques », 871.

⁷⁹⁰ Guerreau, 870.

⁷⁹¹ Guerreau, 871.

Par contraste avec l'article de Millet, Guerreau propose donc dans cette première sous-partie un diagnostic de la situation générale de la pratique des ordinateurs en histoire. Il y apporte par ailleurs une réponse pratique : l'utilisation d'une calculatrice programmable de poche Texas Instrument, qui lui permet de développer une pratique individuelle, et donc plus maîtrisée, d'un certain nombre de procédures statistiques très utiles pour le médiéviste.

La seconde sous-partie évoquée plus haut nous intéresse cependant plus directement pour notre analyse puisque Guerreau y revient plus directement sur sa propre application des analyses factorielles des correspondances. Il présente notamment les limites de la mémoire de sa calculatrice, qui restreint à sept le nombre de colonnes des matrices qu'il lui était possible de traiter à l'aide des analyses factorielles. Quant à cette limite technique, il affirme :

« Les diverses analyses que nous avons effectuées nous ont montré au contraire que cette limitation n'avait rien de rédhibitoire et n'entravait réellement ni le fonctionnement ni l'efficacité de l'analyse des correspondances »⁷⁹².

Ici, Guerreau mentionne donc également les difficultés et les obstacles qu'il a rencontrés, mais sous un jour différent de celui de Millet puisqu'il n'utilise pas le registre lexical du ressenti personnel. Il procède cependant lui aussi à une auto-évaluation explicite de sa démarche, qui le conduit à justifier l'intérêt d'utiliser des analyses factorielles des correspondances en dépit des limites de sa machine.

Pour résumer, l'article de Guerreau manifeste, sur ce premier point, certains éléments communs avec celui de Millet (expression des difficultés, auto-évaluation explicite). À la différence de cette dernière, qui se positionnait en utilisatrice pionnière de l'analyse factorielle des correspondances en histoire médiévale, donnant à voir l'arrière cuisine d'une pratique nouvelle, Guerreau se présente comme un spécialiste et un théoricien de ces méthodes, conscient des difficultés matérielles, sociales, techniques et intellectuelles que pose l'introduction de cette procédure d'analyse des données en histoire. Toutefois, les finalités de ces parties du texte semblent bien similaires : il s'agit

⁷⁹² Guerreau, 872.

1) de présenter aux lecteurs ce qu'est une analyse factorielle de correspondances, sans pour autant rentrer dans le détail des calculs et 2) de promouvoir l'intérêt de cette méthode pour la recherche historique.

2.2.2. Considérations matérielles et instruments de calcul.

Le deuxième élément de comparaison entre les deux articles concerne les marqueurs explicites de la matérialité des usages des moyens électroniques de calcul. Comme nous l'avons vu, Guerreau consacre quelques paragraphes aux problèmes « matériels » que pose l'utilisation des outils de calcul en sciences humaines dans sa première sous-partie. Il insiste notamment sur l'accessibilité des calculatrices programmables de poche en raison de leur prix, et des possibilités qu'elles offrent du point de vue du calcul :

« On voit facilement le rôle central dans cette perspective de la calculatrice programmable de poche. La première du genre est apparue en 1974. Depuis les prix ont sensiblement baissé tandis que les capacités de calcul augmentaient (7) »⁷⁹³.

Il ajoute :

« Dès à présent, une multitude d'analyses statistiques, qui voici quelques années encore relevaient surtout de la théorie, sont exécutables chez soi, avec rapidité et précision, on peut même parfaitement réaliser de petites analyses factorielles (8) »⁷⁹⁴.

Ces indications sur ce que nous avons appelé la matérialité des opérations se différencient cependant de celles données par Millet dans son article sur deux points. Tout d'abord, si ces deux auteurs précisent aux lecteurs l'outil de calcul qu'ils utilisent, et mettent l'accent sur le temps que permet de gagner leur utilisation, Guerreau est le seul à inclure dans son article des considérations sur les aspects économiques ces pratiques :

⁷⁹³ Guerreau, 871.

⁷⁹⁴ Guerreau, 871.

prix des outils de calcul, possibilité de travailler chez soi, et donc d'éviter de payer des heures de calcul. Ensuite, là où Millet présentait les supports matériels du travail (cartes perforées, listings, disque), Guerreau ne mentionne jamais les nombreux documents qui participent pourtant à la mise en œuvre d'analyse factorielle à l'aide d'une calculatrice programmable (feuille de programmation, papier fonctionnel, règle à calcul, etc.).

Un autre point commun entre les deux articles peut encore être mis en lumière à propos de ces précisions sur la matérialité des opérations. Comme dans l'article de Millet, les détails sur la programmation sont rejetés dans les notes de bas de page (7) et (8). Celles-ci renvoient à d'autres types de publications : un manuel intitulé *Méthodes numériques pour calculateur de poche*, de Jon N. Smith, publié en version française en 1976⁷⁹⁵, et un autre article de Guerreau, qui présente comment programmer une analyse factorielle sur une calculatrice programmable de poche, publié dans *Le Médiéviste et l'ordinateur* en 1979⁷⁹⁶. Par ailleurs, on ne retrouve pas non plus de précisions dans le corps du texte sur les opérations constitutives de la méthode d'analyse factorielle.

Plus généralement, on constate ainsi que l'article de Guerreau présente, dans sa partie méthodologique, bien moins d'indications sur les dimensions matérielles de sa pratique des outils de calculs électroniques. Sur la base de ces deux seuls articles, et à la lumière des grandes différences que présentent les modalités d'utilisation des ordinateurs de ces deux médiévistes (travail dans un centre de calcul assisté d'un informaticien *versus* travail seul chez soi à l'aide d'une calculatrice programmable de poche), il est possible d'émettre l'hypothèse que plus les conditions matérielles de la mise en œuvre des calculs à l'aide de moyens électroniques sont lourds, plus les auteurs développent dans leurs articles des considérations sur la matérialité de leurs expériences et sur les ressentis personnels que suscitent ces expériences. Il nous appartiendra dans la suite de notre analyse de confirmer ou d'infirmer cette hypothèse.

⁷⁹⁵ Jon Michael Smith, Christian Saguez, et Jean-Marc Kerilis, *Méthodes numériques pour calculateur de poche* (Paris, France: Eyrolles, 1976, 1976). Nous profitons de cette note pour insister sur le fait que ce type de sources est capital pour comprendre les logiques de la circulation des savoirs liés à la programmation en histoire. Nous avons ainsi, au fil de notre doctorat, constitué un premier corpus composé d'une vingtaine de manuels de ce type, qui feront l'objet d'une étude une fois cette thèse soutenue.

⁷⁹⁶ Alain Guerreau, « Analyse factorielle au moyen d'une calculatrice de poche programmable », *Le médiéviste et l'ordinateur* 2, n° 1 (1979): 10-12.

2.2.3. Des marqueurs de l'enchaînement des opérations discrets.

Un troisième ensemble de considérations doit être pris en compte pour conclure cette comparaison : il concerne la structure de l'exposé de la méthode. Sur ce point, le premier élément qui se dégage concerne les dimensions des développements sur les opérations de codage et de saisie, qui sont de tailles inégales : plus de cinq pages chez Millet contre moins d'une chez Guerreau. De plus, par contraste avec l'article de Millet, Guerreau ne précise pas à quels moments de l'analyse intervient véritablement la calculatrice, renvoyant au contraire le lecteur à ses « *diverses analyses* ». La présentation de la méthode informatique ne suit donc pas une présentation étape par étape comme dans l'article de Millet.

Seules deux mentions de l'enchaînement des opérations apparaissent dans le texte. Or, celles-ci restent assez vagues :

*« Nous avons également décidé [...] de considérer exclusivement les couvents des quatre ordres principaux : Dominicains, Franciscains, Carmes et Augustins. Nous obtenons ainsi une « population » de 609 couvents que nous avons divisée en groupes chronologiques correspondant à un nombre à peu près similaire de fondations. »*⁷⁹⁷

Seul l'adverbe « *ainsi* » fait ici office de connecteur logique entre les deux opérations de constitution de la population statistique et d'établissement des groupes chronologiques. Par ailleurs, si Guerreau détaille dans ce cas les principes généraux de ces deux opérations, il n'en précise toutefois ni les détails ni si celles-ci sont automatisées à l'aide de la calculatrice.

La seconde mention de l'enchaînement des opérations est la suivante :

*« Il nous fallait dès lors pour obtenir 7 colonnes procéder à un découpage chronologique de la période »*⁷⁹⁸.

⁷⁹⁷ Guerreau, « Analyse factorielle et analyses statistiques classiques », 873.

⁷⁹⁸ Guerreau, 873.

Dans ce second exemple, c'est la locution adverbiale « *dès lors* » qui fait office de marqueur permettant de saisir l'enchaînement des opérations. Toutefois, comme dans la citation précédente, cette tournure linguistique est si discrète qu'elle ne nous permet pas d'affirmer qu'il existe, comme dans l'article de Millet, une structure spécifique à l'exposé de la méthode. Le partage de l'expérience de l'auteur se fait ainsi de manière moins linéaire que pour l'article de Millet, sans être cependant complètement absent de la présentation de la méthode.

Là encore, il est possible de se demander quelles raisons sous-tendent ces différences ? Sont-elles liées au fait que la rigidité attachée aux usages des ordinateurs dans les centres de calculs conduit les médiévistes qui y travaillent à développer des procédures de travail plus linéaire ? Ou tout simplement au fait que la finalité de ce deuxième article n'est pas de présenter précisément au lecteur le détail du déroulement des opérations mises en œuvre pour mener cette enquête ? Tiennent-elles encore à d'autres critères que nous n'envisageons pas ? Les pistes dégagées seront poursuivies dans la suite de ce chapitre.

2.2.4. *Nous* ou *je* : positionnement des auteurs vis-à-vis du lectorat.

Un dernier élément doit encore être mis en avant pour mettre en lumière les différences entre les parties méthodologiques de ces deux articles. Il concerne les marques de la présence de l'auteur dans le texte. À la différence de Millet, Guerreau utilise le *nous*, pour prendre position dans le texte. Assurément, en fonction des cas, ces pronoms ne désignent pas tous les mêmes sujets. La linguiste Fløttum, dans un ouvrage collectif de 2006 déjà cité⁷⁹⁹, affirme que les pronoms *nous* et *on* peuvent assumer quatre types de valeurs dans les discours scientifiques en fonction des verbes qui leurs sont associés et du contexte dans lequel ils sont employés :

1. *nous/on* correspondant à un *je*,

⁷⁹⁹ Voir en particulier le quatrième chapitre de Fløttum, Dahl, et Kinn, *Academic Voices*. Voir encore Anje Müller Gjesdal, « The Influence of Genre Constraints on Author Representation in Medical Research Articles. The French Indefinite Pronoun On in IMRAD Research Articles », *Discours. Revue de Linguistique, Psycholinguistique et Informatique. A Journal of Linguistics, Psycholinguistics and Computational Linguistics*, n° 12 (10 juillet 2013).

2. *nous/on* correspondant à un *je + vous* (lecteur),
3. *nous/on* correspondant à un *je + vous* (communauté de recherche pertinente)
4. *nous/on* correspondant à un *je + tout le monde*

Dans le cas de la partie méthodologique de l'article de Guerreau, on compte 14 occurrences de *nous*, dont 13 ont valeur de *je* et 1 de *je + lecteur*. Sur les 16 occurrences de *on*, que l'on retrouve surtout dans la première sous-partie du texte, 12 ont valeur de *je + la communauté des utilisateurs des outils de calculs en histoire*, 3 ont valeur de *je*.

Cette rapide analyse montre que, comme Millet, Guerreau se fait lui aussi fortement présent dans le texte, en utilisant cependant une formulation plus conventionnelle, avec l'utilisation du *nous* plutôt que du *je*, que privilégiait Millet. Ce trait linguistique semble ici imputable au style des auteurs plutôt qu'à la situation de communication ou au contexte de publication.

Par contraste, les nombreux emplois du *on* à valeur de *je + communauté* ne se retrouvent pas dans l'article de Millet. Ceci s'explique par le fait que Guerreau rend compte des problématiques qu'il soulève comme étant celles de l'ensemble de la communauté des chercheurs utilisant des instruments de calculs, tandis que Millet prend seule la responsabilité du récit de son expérience. Cette différence entre les deux articles rejoint celles, identifiées plus haut, sur les formes de l'expression des difficultés et sur l'absence d'occurrence du champ lexical du ressenti personnel dans le texte de Guerreau.

La conclusion que l'on peut formuler à la lumière de ces différents éléments est que ces deux auteurs ne se positionnent pas de la même manière vis-à-vis du lectorat des *Annales*. Guerreau se présente comme un expert de l'analyse factorielle des correspondances, qu'il met en œuvre seul, et dont il revendique l'intérêt pour la discipline et présente les problèmes généraux qui se posent aux historiens désireux de l'utiliser. Millet, *a contrario*, décrit une expérience d'utilisatrice, aidée par un informaticien. En dépit des difficultés qu'elle a rencontrées, elle insiste sur le fait qu'elle a su tirer profit des analyses factorielles des correspondances pour sa recherche.

Ainsi, si les deux articles présentent des aspects linguistiques communs, leurs différences esquissent déjà une première idée de certains éléments qui pourrait nous aider à dégager plus précisément la structure de la formation discursive attachée à l'utilisation des outils de calculs programmables en histoire médiévale. Avant d'en proposer une

nouvelle ébauche, nous proposons de nous intéresser à quatre des dix articles restants du corpus CAHM 1, tous publiés dans la revue *Histoire & Mesure*.

2.3. Variations dans un contexte de publication spécialisé : *Histoire & Mesure*.

Le troisième article de notre corpus est publié dans *Histoire & Mesure* en 1987. Les deux auteurs, Dominique Coq (1948-), conservateur à la Bibliothèque Nationale, et Ezio Ornato, médiéviste au sein de l'Unité Propre de Recherche 52 du CNRS, présentent une méthode quantitative de reconstitution des séquences de composition typographique d'un corpus d'incunables⁸⁰⁰. L'article comporte 11 parties. Dans les neuf premières, Coq et Ornato développent des modèles théoriques qui doivent permettre de saisir l'ordre selon lequel les pages des premiers livres imprimés ont été composés. Pour ce faire, la fréquence des abréviations par pages constitue l'un des indices primordiaux pour les historiens, puisqu'elle permet de repérer les problèmes de jonction que rencontraient les imprimeurs.

La dixième sous-partie de l'article, intitulée « La mise à l'épreuve des modèles théoriques sur 18 éditions », nous intéresse plus particulièrement, puisqu'elle consiste en une présentation des modalités techniques de l'évaluation de l'un de ces modèles théoriques sur un corpus de 18 incunables.

2.3.1. Étude des séquences de composition des incunables (1987).

Le premier point commun entre cette sous-partie et l'article de Millet repose sur la structure de l'exposé de l'enchaînement des opérations. Tout d'abord, les logiques de sélection des éléments du corpus, ainsi que ses limites, sont précisées. Ensuite, Coq et Ornato exposent une partie de ces opérations sous la forme d'une liste en quatre points,

⁸⁰⁰ Coq et Ornato, « Les séquences de composition du texte dans la typographie du XV^e siècle. Une méthode quantitative d'identification ».

présentée par les auteurs comme « *la marche à suivre pour obtenir des résultats chiffrés* »⁸⁰¹.

C'est ici le quatrième point de cette liste qui nous intéresse particulièrement, puisque lui seul a trait à l'utilisation des ordinateurs. En effet, les trois premiers concernent 1) l'examen de la pertinence de cette méthode en fonction du type de corpus étudié, 2) la nécessité de définir fermement ce qui sera pris en compte comme une abréviation dans le dénombrement de ces éléments et 3) la méthode manuelle de comptage de ces abréviations, qui doit aboutir à la construction de deux matrices (« *abPg1 et abPg2* »), qui correspondent chacune, pour des raisons sur lesquelles nous ne reviendrons pas, à une moitié de page.

La présentation du quatrième élément se présente comme une description étape par étape du traitement informatique de ces matrices. L'exposé de cette procédure commence par la justification de la nécessité d'utiliser un micro-ordinateur :

*« Le maniement des matrices d'abréviations est trop lourd pour être effectué manuellement ou sur une calculatrice de poche. Les relevés doivent donc être saisis et traités sur micro-ordinateur »*⁸⁰².

Coq et Ornato poursuivent alors leur exposé en présentant trois ensembles d'opérations à l'aide d'une liste :

- « - Lors de la saisie, le programme crée automatiquement les deux autres matrices abPg et abPg1/abPg2.*
- Pour chacune des matrices, on doit ensuite calculer la variance du nombre d'abréviations dans les 16 sous-ensembles secondaires. Point n'est besoin de recommencer l'opération 16 fois : on se souviendra que chaque page appartient à plusieurs sous-ensembles (cf. plus haut, p. 101) ; il suffira alors de dresser une table d'appartenance à laquelle le programme fera appel à chaque changement de page.*

⁸⁰¹ Coq et Ornato, 113.

⁸⁰² Coq et Ornato, 112

- *On procèdera ensuite au calcul des 24 rapports de variance discriminant, qui sera répété pour chacune des matrices.* »⁸⁰³

À la lecture de cet énoncé, plusieurs remarques s'imposent.

Tout d'abord, on remarque que, de la même manière que dans l'article de Millet, les auteurs présentent l'enchaînement des opérations à l'aide de marqueurs de temporalité (« *ensuite* », « *alors* »). Ce trait commun, sur lequel nous reviendrons longuement dans l'étude du corpus des articles du *Médiéviste et l'ordinateur*, nous apparaît comme caractéristique de la présentation d'une procédure⁸⁰⁴. Le fait de le retrouver ici renforce l'idée que les articles présentant des études qui mettent en œuvre des ordinateurs proposent des présentations étape par étape des opérations, et ce, même si les auteurs ne travaillent pas dans un centre de calcul, contrairement à ce que nous avons pu suggérer au 2.2.3.

Toutefois, à la différence de l'article de Millet, on remarque ici que l'exposé de l'enchaînement des opérations n'est pas présenté en rapport avec l'expérience matérielle qu'il suppose. Au contraire, Coq et Ornato indiquent la succession des opérations de traitement des données qu'ils réalisent (« *calculer la variance* », « *dresser une table d'appartenance* », etc.) sans la mettre en relation avec la matérialité. Cette différence nous révèle que la fonction de la présentation de cette procédure n'est pas la même que dans l'article de Millet. Ici, les auteurs donnent aux lecteurs les clés pour reproduire la méthode proposée sans pour autant que cette répliquabilité n'implique d'utiliser le même instrument de calcul. Sur ce point, nous y reviendrons, la différence entre l'environnement technique de l'enquête de Millet menée à la fin des années 1970 et celui dans lequel évoluent Coq et Ornato en 1987 est certainement déterminante. Nous alimenterons cette hypothèse d'un lien entre la forme des exposés et le type d'environnement de travail au fil de la comparaison avec les autres éléments précédemment repérés.

Un deuxième point commun avec l'article de Millet ressort de cette présentation de l'enchaînement des opérations. On y retrouve de nombreuses mentions explicites de

⁸⁰³ Coq et Ornato, 114.

⁸⁰⁴ Voir par exemple sur ce sujet, mais sur la base de sources tout à fait différentes : Karine Chemla, *Literacy and the History of Science : Reflections Based on Chinese and Other Sources*, *The Cambridge Handbook of Literacy* (Cambridge University Press, 2009).

la matérialité de ces opérations. Les indications sont cependant bien plus détaillées que dans les deux articles analysés précédemment, comme l'illustre la citation suivante :

« La saisie de abPg1 et abPg2 au clavier ne prend que quelques minutes. Quant à l'exécution du programme, elle demande à titre d'exemple, pour chaque édition, 4 à 5 minutes en Basic interprété sur IBM-PC ou compatible, mais ne requiert que 12 secondes en Basic compilé. »⁸⁰⁵

Au contraire de Millet, Coq et Ornato précisent non seulement le temps de saisie des matrices sur le micro-ordinateur, mais ils donnent encore aux lecteurs le temps d'exécution du programme en fonction de l'ordinateur (IBM-PC) et du langage utilisé (Basic ou Basic compilé).

Plus loin, ils ajoutent encore une indication sur le temps nécessaire à la programmation :

« Si l'on se contente de ces calculs, le programme de traitement peut être écrit en très peu de temps par tout amateur ayant quelques rudiments d'un langage de programmation quelconque. »⁸⁰⁶

Les différents éléments dégagés ici vont dans le sens du fait que les auteurs cherchent à rendre possible la réplique de leur méthode, en donnant aux lecteurs des indications précises sur la faisabilité de ce type de procédure, à la lumière du temps, du matériel et des compétences nécessaires.

Cette conclusion nous apparaît d'autant plus justifiée du fait du lectorat visé par cette partie du texte, que Coq et Ornato introduisent en ces termes :

« Nous pouvons maintenant indiquer au lecteur - ou plutôt au spécialiste de l'incunable qui voudrait connaître la séquence de composition de l'édition qu'il étudie - la marche à suivre pour obtenir des résultats chiffrés »⁸⁰⁷.

⁸⁰⁵ Coq et Ornato, 115.

⁸⁰⁶ Coq et Ornato, 114.

⁸⁰⁷ Coq et Ornato, 114.

Pour résumer, les deux auteurs proposent donc, dans cette sous-partie de leur article, une procédure permettant aux spécialistes ayant le même objectif qu'eux d'employer la même méthode, en leur donnant les moyens de s'en emparer.

Pour renforcer cette possibilité, cette sous-partie est par ailleurs complétée par une reproduction de deux tableaux présentant les résultats des opérations statistiques mises en œuvre pour analyser la fréquence des abréviations. La publication de ce type de documents, absents des articles précédemment analysés, peut être comparée aux pratiques éditoriales analysées dans la troisième partie du chapitre précédent. De la même manière que dans *Le Médiéviste et l'ordinateur*, et nous y reviendrons, la reproduction des documents de travail contribue à préciser la finalité originale de cet article dans notre corpus : ce texte doit permettre aux lecteurs de ré-utiliser cette méthode.

Le dernier point de comparaison qui nous intéresse regarde la présence de l'auteur dans le texte. Contrairement aux articles de Millet et de Guerreau, l'article qui nous intéresse ici est co-écrit par deux auteurs. Le *je* est donc logiquement absent du texte. Dans la partie que nous analysons ici, il faut cependant faire remarquer que la totalité des occurrences de *nous* prend pour valeur un *je* + *je*, qui correspond donc, dans le cas de la co-écriture, au *je* de l'article de Millet.

Cette utilisation du *nous* à valeur de *je* + *je*, permet notamment aux auteurs de rappeler le caractère expérimental de leur étude, et d'insister sur ses limites. À propos du choix qu'ils effectuent de ne s'intéresser qu'à des incunables composés de « *feuilles de papier* « *reçutes* » ou « *mezzane* », et non pas « *royales* » » - ces dernières pouvant accueillir « *deux fois plus de texte* »⁸⁰⁸ - ils affirment par exemple :

*« Le choix de cette solution de paresse est bien révélateur du fait que le comptage des abréviations dans un grand nombre de pages est une tâche fastidieuse, que nous serions tentés de recommander aux insomniaques ayant perdu tout espoir de guérison. »*⁸⁰⁹

⁸⁰⁸ Coq et Ornato, 113.

⁸⁰⁹ Coq et Ornato, 114.

Quant à la décision qu'ils prennent de dénombrer les abréviations de tous les cahiers des ouvrages de leur corpus, en dépit de la possibilité de travailler sur un échantillon, ils précisent :

« *Cependant, une telle réduction ne nous paraît pas souhaitable tant que nos connaissances sur les séquences de composition demeurent encore très embryonnaires.* »⁸¹⁰

Comme dans les articles comportant une partie méthodologique précédemment analysés, les auteurs procèdent donc à une auto-évaluation de leurs méthodes, en illustrant les limites de leur travail.

Pour conclure, on remarque que, comme dans le couple d'articles précédents, on retrouve dans la partie méthodologique de l'article de Coq et Ornato plusieurs ingrédients : 1) une présentation structurée selon l'enchaînement des opérations, 2) des marqueurs des aspects matériels de ces opérations, 3) une forte présence des auteurs dans le texte et 4) une auto-évaluation des méthodes liée à leurs limites ou aux difficultés rencontrées. Une différence fondamentale semble cependant devoir être précisée : à la différence de l'article de Millet, et de manière similaire à l'article de Guerreau, le champ lexical du ressenti personnel n'est pas mobilisé par ces auteurs.

2.3.2. Étude du vocabulaire des *exempla* (1986).

Le quatrième article de ce corpus CAHM1 a été publié en 1986 dans *Histoire & Mesure*. Marie-Anne Polo de Beaulieu (1950-), qui vient d'intégrer le Centre de Recherches Historiques en tant que chargée de recherche deuxième classe du CNRS, y présente une étude lexicométrique d'un corpus d'*exempla*, type de récit bref qui vise à donner un modèle de comportement moral⁸¹¹. Dans la première partie de l'article,

⁸¹⁰ Coq et Ornato, 114.

⁸¹¹ Polo de Beaulieu, « Étude statistique de la structure lexicale de l'exemplum médiéval (d'après la Scala Coeli de Jean Gobi) ».

intitulée « *L'exemplum : définition, évolution* », Polo de Beaulieu présente la méthodologie informatique et statistique de son étude.

Elle y détaille, dans un premier temps, la structure du recueil d'*exempla* qui constitue son corpus, puis expose le pré-codage des parties de textes qu'elle propose en ces termes :

- « 1. *La leçon morale ou théologique placée en introduction (=L).*
2. *Le canal d'information chargé d'authentifier le récit (=I).*
3. *Le récit en place centrale (=R).*
4. *L'exégèse allégorique ou moralisation qui est une nouveauté du XIV^e siècle (=E).* »⁸¹²

Elle poursuit en présentant la grille de codage des catégories grammaticales et sémantiques des textes auxquelles elle s'attache :

- « - *aucun codage pour les mots outils*
- *0 pour les entités théologiques*
- *1 pour les verbes*
- *2 pour les adjectifs*
- *3 pour les substantifs désignant des êtres inanimés*
- *4 pour les substantifs désignant des êtres animés*
- *5 pour les noms de lieu*
- *6 pour les noms de personnages historiques*
- *7 pour les adverbes*
- *8 pour les noms d'auteurs*
- *9 pour les noms d'œuvres.* »⁸¹³

Comme dans les articles de Millet d'une part, et de Coq et Ornato d'autre part, cette présentation de la méthode se fait étape par étape. L'auteure commence par définir les logiques de constitution de son corpus et poursuit dans un second temps vers la

⁸¹² Polo de Beaulieu, 48.

⁸¹³ Polo de Beaulieu, 48-49.

description de l'utilisation des documents produits à l'aide de l'ordinateur. Pour exposer sa méthode, elle opte toutefois pour une présentation différente de ce que l'on a vu précédemment, puisqu'elle choisit de présenter son codage par un « *exemple concret* »⁸¹⁴ sur la base de l'un des textes de son corpus.

Pour ce qui concerne la matérialité des opérations mettant en œuvre des ordinateurs, l'article présente lui aussi des points communs avec les trois précédents. Polo De Beaulieu indique notamment l'équipement nécessaire à la saisie des données, le logiciel employé (JEUDEMO), ainsi que les documents de sorties produits à l'aide de ce logiciel :

*« Ce texte précodé a été saisi sur un clavier d'ordinateur pour en constituer un fichier qui a été soumis au logiciel de lexicométrie JEUDEMO (8). Celui-ci nous a fourni des index alphabétiques et hiérarchiques (par ordre de fréquence) pour l'ensemble du recueil et pour chacune des quatre parties de l'exemplum. »*⁸¹⁵

En note, elle renvoie aux deux concepteurs du logiciel JEUDEMO, développé au Canada début des années 1970 par le médiéviste Serge Lusignan (1943-) et l'informaticien Paul Bratley (1940-), tout en précisant que le logiciel est disponible au Laboratoire d'Informatique pour les Sciences de l'Homme (LISH), à Paris.

En annexe, l'auteure propose une bibliographie, qui comprend les principaux périodiques qui portent sur l'utilisation des ordinateurs pour l'étude des textes : *Computer and the Humanities*, *MOTS*, *Le Médiéviste et l'ordinateur*, mais aussi les travaux d'auteurs que nous avons déjà rencontrés au fil de cette thèse : Paul Tombeur (1936-), Léopold Génicot (1914-1995), Michel Pêcheux (1938-1983) ou encore Genet par exemple. Ces éléments présents dans les notes de bas de page constituent autant d'indices de la finalité de l'article, qui apparaît tourné vers le partage des savoirs statistiques et linguistiques utilisés par l'auteure, plutôt que des procédures informatiques effectivement mises en œuvre dans le cadre de l'étude.

⁸¹⁴ Polo de Beaulieu, 48.

⁸¹⁵ Polo de Beaulieu, 49.

Un autre point commun avec les articles de Millet et de Coq et Ornato doit être noté, puisque Polo de Beaulieu distingue les opérations qui ont été effectuées à la main de celles réalisées automatiquement par l'ordinateur :

« C'est essentiellement à partir de ces derniers index que nous avons calculé à la main deux indicateurs statistiques »⁸¹⁶.

Toutefois, contrairement aux autres articles, l'auteure détaille le calcul des indicateurs statistiques (des écarts réduits), à partir d'un exemple extrait de son corpus. Comme pour la présentation de son codage, elle choisit de s'appuyer, pour l'exposition des éléments techniques, sur un exemple. À la différence de l'article de Coq et Ornato, cette stratégie de présentation ne porte pas sur la programmation de la méthode sur ordinateur, mais sur les calculs faits à la main par la médiéviste. Polo de Beaulieu fait figurer ainsi une partie de texte absente des autres articles, qui, comme nous allons le voir, nous renseigne sur la particularité du contexte de publication comme sur le positionnement de l'auteure au sein de la formation discursive qui nous intéresse ici.

Un autre indice linguistique nous permet en effet de mieux comprendre la particularité de ce développement mathématique. Dans cette partie, on constate que les marqueurs de l'implication de l'auteur dans le texte changent. Au contraire du reste de l'article, dans lequel Polo de Beaulieu utilise un *nous* à valeur de *je*, la médiéviste utilise ici le *on* :

« On calcule d'abord un effectif théorique qui nous donne le nombre d'adjectifs que l'on devrait trouver dans les 4 registres [...]. De là, on calcule l'écart qui n'est rien d'autre que le résultat de la soustraction suivante :

*effectif réel (nombre observé d'adjectifs dans la leçon) : 597 - effectif théorique :
241 = 356*

Ensuite on calcule l'écart type ainsi : [...] »⁸¹⁷.

⁸¹⁶ Polo de Beaulieu, 49.

⁸¹⁷ Polo de Beaulieu, 52.

Comment interpréter ce passage du *nous* au *on* ? L'une des explications possibles serait de considérer que ce *on* a valeur, selon la classification de Fløttum, de *je* + communauté pertinente.

Pour étayer cette hypothèse, il faut ajouter que l'auteure précise que le choix de cet indicateur statistique repose sur une méthode développée par Charles Muller (1909-1915), linguiste spécialiste des statistiques lexicales, dont Polo de Beaulieu cite le manuel à deux reprises dans cette partie de l'article⁸¹⁸. À la lumière de cette référence, le passage au *on* permettrait ainsi à l'auteure de ne pas passer sous silence les origines de son savoir statistique, tout en le légitimant. Mais il est aussi tout à fait possible qu'il ait valeur de *je* + lecteur. En ce sens, l'auteure chercherait à impliquer le lecteur dans le calcul, ajoutant ainsi une dimension didactique à son article, du type de celle que nous avons observée dans la dernière partie du chapitre 4.

Parmi les articles faisant explicitement mention de l'usage des ordinateurs publiés dans *Histoire & mesure*, il faut encore ajouter qu'il existe deux autres textes entièrement dédiés à ce type de discussions. Ils partagent comme sujets la présentation d'un programme informatique : le premier vise à faciliter la recherche dans une base de données généalogiques ; le second permet la reconstitution semi-automatique de généalogies⁸¹⁹. Ces deux articles sont aussi les seuls à proposer des outils utilisables par les historiens de toutes les périodes. Sans revenir en détail sur ces deux textes, il est possible d'y repérer les quatre traits caractéristiques que nous avons ici mis en lumière.

Les traits relevés à partir de l'article de Millet s'avèrent donc présenter un certain degré de généralité puisque les quatre articles d'*Histoire & Mesure* regroupent les mêmes caractéristiques : 1) une structure basée sur l'enchaînement des opérations, 2) marqueurs de la matérialité des opérations et 3) forte présence des auteurs dans le texte. La mention des difficultés n'apparaît cependant pas, pour ce corpus, comme un élément généralisable, puisque les articles de Coq et Ornato et de Polo de Beaulieu n'en comportent pas. On remarque encore que les variations autour de ces trois caractéristiques de base permettent de saisir, non sans difficultés, le positionnement des auteurs au regard des lecteurs, ainsi que le poids de leurs environnements de travail sur les formes de

⁸¹⁸ Charles Muller, *Initiation à la statistique linguistique* (Paris, France: Librairie Larousse, 1968).

⁸¹⁹ Il s'agit respectivement de : Marie, « Histoire de la famille et outil informatique » et Selz, « Procédures d'aide à la reconstitution des généalogies ».

l'exposé des méthodes. Ces différents éléments permettent d'esquisser les contours d'une formation discursive attachée aux usages des ordinateurs en histoire médiévale.

2.4. Deuxièmes conclusions : quelques caractères des parties méthodologiques.

2.4.1. Quelques remarques générales.

Quelles conclusions peut-on tirer de l'analyse de ce corpus de douze articles CAHM1 ? On constate d'abord qu'une proportion importante d'articles (6 sur 12) n'aborde que très brièvement l'utilisation des ordinateurs, ou renvoie entièrement ces questions dans les notes de bas de page. Il faut encore insister sur le fait qu'il existe certainement, nous en avons vu un exemple dans la première partie de ce chapitre, des articles présentant des études ayant recours aux moyens électroniques qui ne mentionnent pas explicitement leur utilisation.

Pourtant, et contrairement à ce que percevait Zysberg en 1980, on observe que certains articles présentent au moins une partie explicitement méthodologique. Un bref retour sur la période antérieure aux années 1980 nous permet de constater que les développements méthodologiques dans ce type d'articles n'apparaissent pas soudainement à cette période. On trouve même, dans certains articles du début des années 1970, des textes littéralement coupés en deux, avec, dans une première partie, un exposé méthodologique signé par un informaticien ou un mathématicien et, dans une seconde, une analyse historique rédigée par un historien⁸²⁰. À ce titre, le début des années 1980 semble marquer un tournant puisqu'aucun des articles du corpus CAHM1 n'est co-écrit par un binôme historien/informaticien.

La seconde remarque d'ordre général porte sur la place accordée aux énoncés portant sur les méthodes informatiques dans chacune des revues et donc, sur la formation

⁸²⁰ Un exemple particulièrement frappant de séparation entre partie informatique et partie historique a d'ailleurs conduit les équipes de la plateforme *persée.com* à proposer une version numérique d'un même article en deux parties distinctes : Yves Auffray, « I. Traitement informatique », *Mélanges de la Casa de Velázquez* 12, n° 1 (1976): 141-45 ; Jacqueline Guiral, « II. Informatique et péages du royaume de Valence (1494) », *Mélanges de la Casa de Velázquez* 12, n° 1 (1976): 146-63. Dans les années 1970, les articles sont d'ailleurs le plus souvent co-signés par l'informaticien et l'historien. Voir par exemple : Marion Créhange et Lucie Fossier, « Essai d'exploitation sur ordinateur des sources diplomatiques médiévales », *Annales* 25, n° 1 (1970): 249-84 et Yvette Daubèze et Jean-Claude Perrot, « Un programme d'étude démographique sur ordinateur », *Annales* 27, n° 4 (1972): 1047-70.

discursive qui semble se dessiner sur la base de l'analyse de ce corpus. Dans les *Cahiers de Civilisation Médiévale*, aucun des trois articles repérés ne comporte de partie explicitement méthodologique. Les développements sur l'utilisation de l'informatique sont brefs et se situent dans les introductions ou les conclusions de ces articles. Ils peuvent encore être rejetés en notes. Dans ces deux types de cas, ils ne portent jamais sur les aspects matériels de ces entreprises tels que nous les avons rencontrés ailleurs (équipement, support matériels, dimension temporelle du travail, analyse coût-bénéfice, organisation sociale des calculs, articulation opérations manuelles et automatisées).

Il faut par ailleurs ajouter, comme nous l'avons vu avec l'exemple de Carpentier, qu'il est tout à fait possible que certaines études ayant eu recours aux moyens électroniques n'en fassent pas mention. Partant, il est possible d'affirmer que les articles présentant des études ayant eu recours aux ordinateurs dans cette revue sont exclusivement tournés vers la présentation des résultats de ces traitements, et non de la méthodologie informatique elle-même.

Dans le journal les *Annales ESC*, on remarque par contre que deux des quatre articles d'histoire médiévale que nous avons repérés comportent une partie explicitement méthodologique. Si l'on étend cette analyse à l'ensemble des articles des *Annales ESC* mentionnant explicitement l'utilisation des ordinateurs, et plus seulement ceux qui concernent des études portant sur la période médiévale - soit 18 articles -, on peut noter que 9 articles comportent des parties méthodologiques présentant les aspects informatiques de ces études. Par contraste avec les *Cahiers de Civilisation Médiévale*, la revue les *Annales ESC* constitue donc un espace dans lequel il est possible de partager non seulement les résultats des traitements informatiques, mais également certains éléments méthodologiques. Cela s'explique certainement par une « attraction » plus forte de cette revue pour les méthodes quantitatives et leur automatisation, que traduit la publication de plusieurs articles sur ce sujet depuis la fin des années 1950⁸²¹.

⁸²¹ Les éditeurs des *Annales ESC* peuvent en effet se targuer d'avoir été parmi les premiers à publier les résultats, même provisoires, d'études utilisant des machines mécanographiques ou électroniques, mais également des présentations plus théoriques des possibilités offertes par ces nouveaux outils. On pourra notamment regarder, dans l'ordre chronologique : Adeline Daumard et François Furet, « Méthodes de l'Histoire sociale : les Archives notariales et la Mécanographie », *Annales* 14, n° 4 (1959): 676-93 ; Jean-Claude Gardin et Paul Garelli, « Étude par ordinateurs des établissements assyriens en Cappadoce », *Annales* 16, n° 5 (1961): 837-76 ; Jean Delumeau, « Méthode mécanographique et trafic maritime : les terre-neuviens malouins à la fin du XVII^e siècle », *Annales* 16, n° 4 (1961): 665-85 ; Valentin A. Ustinov, « Les calculateurs électroniques appliqués à la science historique », *Annales* 18, n° 2 (1963): 263-94 ;

Plus conformément aux attentes, on retrouve, sur les cinq articles identifiés dans *Histoire & Mesure* 1) trois articles entièrement consacrés à des questions de méthodes (Coq et Ornato, Selz, Marie), 2) un autre qui comporte une partie méthodologique (Polo de Beaulieu), et 3) seulement un article qui ne fait que brièvement mention des aspects informatiques de l'étude qu'il concerne (Leblond). Si l'on étend cette analyse aux 16 articles mentionnant explicitement l'utilisation des ordinateurs dans *Histoire & Mesure*, et plus seulement ayant trait aux études médiévales, on peut cette fois identifier 14 articles présentant une partie méthodologique. Soulignons par ailleurs que le choix d'exclure les articles après 1989 fait de ce sous-corpus le plus petit des trois. La proportion d'articles présentant les aspects informatiques des méthodes mises en œuvre est, de ce fait, assurément beaucoup plus élevée que pour les deux autres contextes de publications.

2.4.2. Un genre propre aux parties méthodologiques ?

Une fois établies ces différences entre ces trois contextes de publication, que peut-on dire sur la teneur de ces parties méthodologiques ? Comme nous l'avons vu, les articles analysés partagent quatre traits de façon plus ou moins marquée :

- 1) Une présentation des opérations étape par étape
- 2) Des mentions des aspects matériels de la mise en œuvre des moyens électroniques
- 3) Une implication personnelle des auteurs dans les textes
- 4) Des indications sur les obstacles et les difficultés rencontrées dans l'application de ces méthodes statistiques automatisées.

La combinaison de certains de ces quatre traits, et notamment des points 1-2 et 2-4 observée dans les articles du corpus CAHM1 permet, selon nous, de formuler l'hypothèse qu'il existerait un genre d'article corrélé à la présentation des méthodes informatiques en histoire médiévale.

Tout d'abord, cinq des six textes que nous avons analysés proposent une présentation de la méthode que nous qualifierons de procédurale, puisqu'elle expose, étape par étape, la succession des opérations intellectuelles et matérielles menées dans le

Marcel Couturier, « Vers une nouvelle méthodologie mécanographique. La préparation des données », *Annales* 21, n° 4 (1966): 769-78.

but de réaliser une recherche au contact des calculateurs électroniques. Il ne faudrait cependant pas considérer cette structure comme le reflet d'une réalité pratique, mais plutôt comme une reconstitution logique, faite *a posteriori*, dans le but de faire passer plus aisément la démarche aux lecteurs. Certains auteurs du corpus CAHM1 se refusent d'ailleurs à utiliser cette structure artificielle de présentation, comme Guerreau, qui ne précise pas l'ordre des opérations, certainement attentif au caractère artificiel de cette forme de présentation.

Ce premier trait commun ne saurait cependant être corrélé à l'utilisation des ordinateurs s'il ne se combinait pas avec le second, que l'on retrouve dans ces cinq articles : les nombreuses mentions à la matérialité des opérations informatiques (saisie, codage des données, utilisation des programmes). Les articles font tous références à l'outil de calcul utilisé (IBM 360, micro-ordinateur, calculatrice électronique de poche). Toutefois, ces précisions sont de natures différentes en fonction des textes. Les articles plus anciens, ceux de Guerreau et de Millet, ne mentionnent pas précisément le fonctionnement des programmes et ne présentent que les résultats des opérations qu'ils permettent. Millet présente cependant la méthode de stockage des données (cartes perforées, puis disques).

Au contraire, les articles les plus récents, d'une part celui de Ornato et Coq, et d'autre part celui de Polo de Beaulieu, font état du programme utilisé (JEUDEMO), ou même du langage informatique choisi (BASIC). Ornato et Coq renseignent par ailleurs le lecteur sur le temps nécessaire pour chaque opération de calcul ou de saisie. C'est ici à l'éclairage d'une évolution technologique qu'il faut comprendre ces différences. Les premiers utilisent des outils de calculs pratiquement inconnus des lecteurs (informatique lourde des centres de calcul ou calculatrice programmable de poche). Les programmes sont le plus souvent artisanaux et adaptés à un corpus particulier. Au contraire, Ornato, Coq et Polo de Beaulieu présentent des études menées à l'aide de micro-ordinateurs, dont les usages sont, en 1986, déjà répandus chez de nombreux médiévistes, comme nous l'avons vu au chapitre précédent.

Pourtant, à la lecture de ces articles, de nombreuses zones d'ombres subsistent pour celui qui voudrait reproduire la démarche présentée par les auteurs. L'article de Guerreau est, à ce titre le plus obscur. Celui d'Ornato et Coq, au contraire, laisse, selon nous, bien moins d'éléments dans l'ombre, si ce n'est le plus important peut-être, à

savoir : le logiciel employé. Trois éléments particuliers de la démarche informatique permettent de qualifier plus précisément ces différences entre ces quatre articles : le codage, la présentation des procédures de calculs et la description des programmes informatiques. Sur ces six textes, aucun ne fait par exemple l'économie de la description du codage. Mais cette partie de texte peut être plus ou moins longue, et plus ou moins précise. Nous y reviendrons par contraste avec l'analyse des « articles de fond » du *Médiéviste et l'ordinateur*.

Pour ce qui concerne l'exposé des procédures de calcul, seuls les deux articles publiés dans *Histoire & Mesure* les présentent dans le corps du texte. Polo de Beaulieu illustre par exemple, formules à l'appui, comment calculer les écarts réduits entre les occurrences de différentes formes de mots. Ornato et Coq ne développent quant à eux aucune formule, mais indiquent point par point l'enchaînement des calculs à effectuer avec le micro-ordinateur. Il faut ajouter à ces deux cas un troisième, issu du même corpus CAHM1. Si l'auteur n'y fait pas explicitement mention de l'usage de l'informatique en dehors de notes de bas de pages, il présente cependant la méthode de calcul qu'il utilise pour son étude, là encore, formules à l'appui⁸²².

Au contraire, les deux articles publiés dans les *Annales ESC* excluent les développements sur les calculs nécessaires à la mise en œuvre des analyses factorielles de correspondances, et renvoie le lecteur à des ouvrages spécialisés dans les notes de bas de page. Bien entendu, la complexité de la procédure de calcul de l'analyse factorielle est à prendre en compte pour expliquer cette absence, mais rappelons que Millet et Guerreau en font le sujet central de leurs articles. La présentation est axée sur les graphiques des plans factoriels, et sur leur lecture ou leur utilisation. Aucun élément permettant aux lecteurs de saisir les opérations qui conduisent à l'élaboration des programmes ne transparaît dans le texte. Millet décrit les autres programmes utilisés de la même manière, c'est-à-dire par les opérations qu'ils permettent de réaliser (tris, comptage). Guerreau renvoie quant à lui, en note, à un autre article présentant la programmation d'une calculatrice programmable pour effectuer une analyse factorielle, article qui est justement, et nous y reviendrons, publié dans *Le Médiéviste et l'ordinateur*.

De la même manière, mais dans *Histoire & Mesure*, Polo de Beaulieu renvoie dans les notes aux programmes qu'elle utilise. Cependant, dans ce cas, le contexte de

⁸²² Leblond, « Recherches métrologiques sur des plans de bastides médiévales ».

publication est différent. Nous l'avons dit, *Histoire & Mesure* s'adresse aux historiens intéressés par les méthodes statistiques et informatiques. La présentation du logiciel JEUEMO, bien identifié au sein de cette communauté à la fin des années 1980⁸²³, pourrait apparaître comme une précision non nécessaire pour l'auteur. De plus, dans les notes, Polo de Beaulieu renvoie, non pas à d'autres articles, mais à un contact au Laboratoire d'Informatique pour les Sciences de l'Homme (LISH). L'indication qu'elle donne nous paraît donc, en un sens, bien plus précise que celles faites par Millet ou Guerreau.

Un autre article du corpus CAHM1, brièvement évoqué ci-dessus, va cependant à l'encontre de cette conclusion. Il s'agit de l'article publié par la statisticienne Marion Selz-Laurière en 1989 dans *Histoire & Mesure*⁸²⁴. L'auteur y mentionne non seulement les règles de codage qu'elle met en œuvre, les procédures de classement et de calcul qu'elle applique à ses données, le modèle de l'ordinateur qu'elle utilise et même le système d'exploitation. Elle présente également une liste de chacune des neuf fonctions du programme, dont voici un très bref extrait :

« *Programme* :

<i>patro(*x) <- readat(*x, entree).</i>	<i>un patronyme x est lu</i>
<i>Squelet <- patro(*x)</i>	<i>un prédicat « squelet » est créé</i>
	<i>pour chaque patronyme x</i> » ⁸²⁵

La suite du texte présente les neuf autres fonctions de ce programme sous une forme similaire. Cet exemple est le seul du corpus CAHM1 dans lequel un langage de programmation, ici PROLOG, fait irruption dans le texte, et marque ainsi plus fortement le caractère technique de l'article.

⁸²³ Le logiciel JEUEMO, développé au Canada début des années 1970 par Serge Lusignan (1943-) et Paul Bratley (1940-), a déjà été cité plus d'une dizaine de fois dans le contexte francophone avant 1986. Les lecteurs attentifs aux études médiévales assistée par ordinateur, ou aux statistiques lexicales, ont ainsi pu en entendre parler dans *Médiévales*, dans *MOTS*, dans les actes du colloque « Histoire et informatique » de 1975, déjà plusieurs fois cités dans cette thèse, mais aussi dans la *Revue Philosophique de Louvain*, dans la *Revue d'Histoire de l'Amérique française* et bien entendu dans *Le Médiéviste et l'ordinateur*.

⁸²⁴ Selz, « Procédures d'aide à la reconstitution des généalogies ».

⁸²⁵ Selz, 9.

La deuxième combinaison de traits qui nous intéresse concerne 3) les nombreuses occurrences de la présence de l'auteur dans le texte et 4) les indications des difficultés et des obstacles rencontrés. L'utilisation du *je* (ou du *nous* à valeur de *je*), la mention des difficultés, et pour l'un des articles, le registre lexical du ressenti personnel, rapprochent ces articles du genre des récits d'expériences vécues, étudié notamment par Carter-Thomas et Rowley-Jolivet⁸²⁶. Nous reviendrons dans la troisième partie de ce chapitre, chiffres à l'appui, sur cet élément. Pour le moment, il faut toutefois noter que des différences notables apparaissent entre les articles concernés. Elles mettent en lumière les façons par lesquelles les auteurs proposent des narrations différentes de leurs expériences, par ailleurs toutes singulières.

En effet, le public visé, la place des auteurs dans les communautés de chercheurs, mais aussi la finalité des articles joue pleinement sur les manières de raconter l'expérience informatique. Dans les *Annales*, et en particulier du fait que les articles de notre corpus sont publiés au début la période que nous étudions, ces stratégies de communication révèlent avant tout un objectif de promotion des méthodes informatiques, et, certainement, plus largement, d'une ambition de promotion individuelle dans le champ plus large de la discipline historique. Au contraire, dans *Histoire & Mesure*, les auteurs semblent plus à l'aise vis-à-vis d'un lectorat plus spécialisé et, *a priori*, mieux formé aux méthodes statistiques et en informatiques.

À la lumière de ces premiers constats, il ressort qu'en plus du contexte de publication, les modalités d'organisation du travail avec les ordinateurs jouent pleinement sur les manières de présenter la méthodologie. En effet, on remarque que les historiens qui disposent d'une autonomie technologique et technique, comme Coq et Ornato par exemple, précisent plus explicitement les modalités d'utilisation des ordinateurs. À ce titre, l'article de Guerreau échappe à cette règle, puisqu'il n'est pas très précis sur la méthodologie, alors même que son auteur dispose d'une indépendance technologique presque totale. Au contraire, Millet, qui travaille dans un contexte matériel très contraignant (informatique lourde, fiches perforées, collaboration avec un informaticien), insiste sur son expérience d'utilisatrice plutôt que sur les aspects techniques du travail.

⁸²⁶ Shirley Carter-Thomas et Elizabeth Rowley-Jolivet, « Open science notebooks: New insights, new affordances », *Journal of Pragmatics* 116 (2017): 64-76; Rowley-Jolivet et Carter-Thomas, « Raconter comment cela se passe vraiment. Comparaison entre la narration de la recherche dans les cahiers de laboratoire et dans les articles de recherche ».

Partant de ces conclusions, comment peut-on alors qualifier à la fois la place des « articles de fond » du *Médiéviste et l'ordinateur* dans cette formation discursive et les particularités de ces articles ? Le genre de textes qui se développe dans ce bulletin de liaison est-il similaire au genre des parties méthodologiques mis en lumière ci-dessus ? Et quelles sont les différences entre ces deux types de contextes de publication ?

3. Les « articles de fond » du *Médiéviste et l'ordinateur*.

Les premières conclusions de ce chapitre tendaient à présenter les « articles de fond » du *Médiéviste et l'ordinateur* comme le complément technique des articles publiés dans d'autres contextes. Les nombreux renvois au bulletin de liaison que nous avons rencontrés dans les articles analysés, notamment dans leurs notes de bas de pages, renforcent cette idée. Cependant, nous avons aussi pu constater que certains articles publiés dans les revues académiques traditionnelles comportaient des parties méthodologiques. Dès lors, comment apprécier la spécificité des articles publiés dans *Le Médiéviste et l'ordinateur* ? Et quelles sont les fonctions particulières attribuées à ces articles par leurs auteurs, si on les compare avec les parties méthodologiques publiées dans d'autres contextes de publication ?

Pour répondre à ces questions, une analyse de l'ensemble des 77 textes de la rubrique « *Article de fond* » publiés entre 1979 et 1989 dans le bulletin de liaison s'impose. Mais, pour mieux saisir les différences entre ces articles publiés dans le *Médiéviste et l'ordinateur* et les autres, il est nécessaire dans un premier temps de les comparer aux articles du corpus CAHM1. Il sera alors possible d'observer s'il existe ou non des spécificités dans les discours des acteurs, en fonction du contexte de publication. Cette comparaison est facilitée par le fait que six des 11 auteurs du corpus CAHM1 ont publié également dans le *Médiéviste et l'ordinateur*, et que certains de ces articles concernent les mêmes études. Dans un premier temps, nous opérerons une analyse de deux articles publiés dans les pages du bulletin de liaison, d'une part par Millet, et d'autre part par Guerreau.

Dans un second temps, et à la lumière des conclusions dégagées, il sera alors possible d'étendre l'analyse et de comparer le corpus constitué de l'ensemble des

« articles de fond » du *Médiéviste et l'ordinateur* avec le corpus des 41 articles mentionnant explicitement l'utilisation des ordinateurs dans les revues les *Annales ESC* et d'*Histoire & Mesure*. Pour étudier autant d'articles, une approche quantitative s'impose. Comme nous le verrons, la méthode choisie propose de s'attacher à l'examen des quatre critères que nous avons dégagés jusqu'ici à partir du corpus CAHM 1. Il sera alors possible de dégager la finalité et les fonctions de ce type d'articles du bulletin, et de définir certaines caractéristiques concernant le genre et les fonctions des « articles de fond » du bulletin de liaison qui nous intéressent depuis le début de ce chapitre.

3.1. Expériences pratiques des auteurs et contextes de publications.

Comme nous l'avons dit, sur les onze auteurs du corpus CAHM1, six publient des articles dans le *Médiéviste et l'ordinateur* entre 1979 et 1992. Trois d'entre eux proposent des articles directement liés aux études présentées plus haut.

Par exemple, Millet publie deux articles qui portent directement sur son étude des chanoines de Laon. Le premier se présente, dans la lignée de sa publication dans les *Annales ESC*, comme un témoignage de l'intérêt des analyses factorielles des correspondances pour l'historien médiéviste⁸²⁷. Le second, qui nous intéressera plus particulièrement, concerne une « expérience » de classification automatique des chanoines réalisée à partir du même corpus que celui qu'elle avait mobilisé dans l'enquête parue dans les *Annales ESC*⁸²⁸.

Guerreau publie lui aussi dans *Le Médiéviste et l'ordinateur* une contribution en lien avec son article étudié au 2.2., qu'il cite d'ailleurs dans son étude parue dans les *Annales ESC* en 1981⁸²⁹. Il y détaille les modalités selon lesquelles il programme une analyse factorielle de correspondances sur une calculatrice de poche.

Enfin, comme nous l'avons vu en première partie, Carpentier décrit dans un article publié dans *Le Médiéviste et l'ordinateur* en 1980 la saisie et le codage de son corpus de données établi sur la base du cadastre d'Orvieto, dont elle a par ailleurs publié les résultats

⁸²⁷ Hélène Millet, « Pour l'analyse factorielle », *Le médiéviste et l'ordinateur* 5, n° 1 (1981): 18-19.

⁸²⁸ Hélène Millet, « Une expérience : essai de classification des chanoines de Laon », *Le médiéviste et l'ordinateur* 7, n° 1 (1982): 14-17.

⁸²⁹ Guerreau, « Analyse factorielle au moyen d'une calculatrice de poche programmable ».

dans la *Revue du Nord* en 1980⁸³⁰. Les trois autres auteurs que nous avons repérés publient, quant à eux, des articles plus généralistes, sur lesquels nous reviendrons dans un second temps.

Cette présentation rapide des contenus des articles nous suffit à mettre en lumière le fait que ces textes ne présentent pas exhaustivement les méthodes employées, mais seulement certains éléments de ces méthodes. Dans leurs articles publiés dans le bulletin de liaison, ni Millet ni Guerreau ne reprennent par exemple les calculs qui conduisent à la mise en œuvre de l'analyse factorielle. Dès lors, quels sont les points particuliers sur lesquels les auteurs insistent ? Et comment ces articles se différencient-ils de ceux du corpus CAHM1 ?

3.1.1. Finalités et stratégies rhétoriques.

Les deux articles qui concernent l'enquête de Millet semble, à première vue, présentés de nombreuses similitudes. Plusieurs différences notables apparaissent cependant. La première concerne la finalité de ce texte. Comme dans son article des *Annales ESC*, une première partie présente, étape par étape, l'enchaînement de certaines des opérations qui aboutissent à la construction de l'analyse factorielle. Quatre tirets permettent à l'auteure d'établir une liste de quatre moments du travail, constitués chacun d'un ensemble d'opérations. Mais, à la différence de son article paru dans les *Annales ESC*, ce texte se présente comme un résumé très court de sa méthode.

La finalité de cet article se révèle ensuite dans la seconde partie du texte, qui est introduite en ces termes :

*« Pour ceux que tente cette méthode et qui se demandent comment l'aborder, je voudrais ici présenter quelques conclusions pratiques tirées de mon expérience. »*⁸³¹

Par cette phrase, Millet précise le type de lecteurs auxquels son article s'adresse. Les « *conclusions pratiques* » qu'elle formule à leur attention peuvent être résumées ainsi :

⁸³⁰ Elisabeth Carpentier, « Le paysage rural du Contado d'Orvieto à la fin du XIII^{ème} siècle d'après le cadastre de 1292 », *Revue du Nord* 62, n° 244 (1980): 177-89.

⁸³¹ Millet, « Pour l'analyse factorielle », 19.

1) une remarque sur l'accessibilité de cette procédure pour les historiens, dont le principe est « *compréhensible intuitivement* » et les programmes « *utilisables par ceux qui ont compris le principe* »⁸³² ; 2) des conseils sur la « *meilleure initiation* » possible à cette méthode ; 3) un avertissement sur la nécessité de « *comprendre le principe de la méthode* »⁸³³ pour interpréter les graphiques ; 4) une préconisation sur l'intérêt de collaborer avec un informaticien ; 5) sur l'importance d'être le plus indépendant possible du point de vue informatique et 6) des considérations sur le profit de cette méthode malgré « *l'investissement en temps et en moyens* »⁸³⁴ qu'elle nécessite.

À la différence de l'article publié dans les *Annales ESC*, on remarque que la finalité de l'article n'est pas de présenter l'intérêt de la méthode pour l'avancée de la recherche en histoire médiévale, en insistant, rappelons-le, sur la possibilité qu'elle offre d'éviter une approche prosopographique fondée sur l'étude de certains cas marginaux ou par l'établissement d'un profil sociologique moyen. Il s'agit au contraire d'éclairer les lecteurs sur les aspects pratiques de cette démarche.

Cette finalité se révèle encore à la lumière du type de difficultés partagées par Millet dans cet article. Au contraire de l'article des *Annales ESC*, dans lequel ce type d'énoncés portait sur des opérations statistiques, comme le codage ou la saisie, les remarques conclusives présentées par Millet ici concernent des obstacles tout à fait différents (difficultés de la méthode, organisation sociale des calculs, analyse coût-bénéfice).

Un autre élément clef permet de saisir la finalité de cet article. Il réside dans l'existence de plusieurs passages qui ont vocation à convaincre le néophyte de l'intérêt de cette nouvelle méthode :

*« le principe même de l'analyse factorielle est compréhensible intuitivement et les programmes disponibles sont aisément utilisables par ceux qui ont compris ce principe. »*⁸³⁵

Ou encore :

⁸³² Millet, 19.

⁸³³ Millet, 19.

⁸³⁴ Millet, 19.

⁸³⁵ Millet, 19.

« L'investissement en temps et en moyens que représente cette méthode me paraît grandement justifié par l'aide qu'elle apporte : les graphiques d'analyse factorielle sont d'incomparables stimulants pour l'esprit »⁸³⁶.

Dans ces deux citations, l'utilisation des adverbes (« intuitivement », « aisément », « grandement ») permet à l'auteure d'insister sur l'analyse coût-bénéfice de cette méthode pour les médiévistes, en l'éclairant à l'aune de sa propre expérience. Cette stratégie rhétorique permet à Millet de poursuivre l'objectif qu'elle poursuivait avec cet article : persuader le lecteur du *Médiéviste et l'ordinateur*, qui serait tenté par cette méthode d'analyse des données, de sauter le pas.

La seconde différence, plus ténue, entre cet article et celui paru dans la revue les *Annales ESC*, va dans le sens de la finalité que nous venons de repérer. Elle concerne la présence de l'auteur dans le texte. Comme dans le premier article, l'utilisation du *je* est privilégié et les difficultés rencontrées par l'auteure sont mises en avant. Cependant, les deux articles présentent un contraste. Tout d'abord, Millet utilise dans ce texte un style beaucoup plus informel, voire même familier :

« Il avait fallu toute la persuasion de mon informaticien de mari pour que j'accepte de me prêter, moi et mes chanoines, à cette expérience. Or, voilà que j'avais sous les yeux comme un résumé graphique de l'ensemble des correspondances que l'on pouvait établir entre mes diverses séries de chiffres »⁸³⁷.

L'informaticien, qu'elle présentait sans plus de précision comme « Jean-Paul Millet » dans le premier article, est ici explicitement désigné comme son mari.

De plus, pour faire connaître les motifs qui l'ont conduite à accepter cette « expérience », l'auteure mobilise une série de termes que nous qualifierons de champ lexical du ressenti personnel, et développe ainsi une rhétorique de la persuasion plutôt que la conviction. On retrouve ce même type de fait linguistique plus loin dans le texte :

⁸³⁶ Millet, 19.

⁸³⁷ Millet, 18.

*« ce résumé, non seulement était cohérent, mais encore m'invitait à tirer des conclusions qui m'avaient jusque-là échappé : le résultat dépassait de beaucoup les espérances ! »*⁸³⁸

Dans ce second exemple, l'auteure multiplie également l'emploi de verbes pronominaux (« *m'invitait* », « *m'avaient [...] échappé* »), ce qui renforce son implication dans le texte et la sensation pour le lecteur de suivre l'« *expérience* » vécue par la médiéviste. La dernière phrase de l'article, dans laquelle l'auteure fait référence aux « *espérances* » qu'elles avaient placées dans cette méthode, renforce cette impression qu'il serait possible de saisir la recherche dans ses dimensions plus affectives, par contraste avec l'article publié dans les *Annales ESC*, qui présentait les résultats d'une recherche de façon plus désincarnée.

3.1.2. Description d'une procédure de calcul et répliquabilité.

L'analyse du couple d'articles de Guerreau permet de constater elle aussi, quoique d'une façon tout à fait différente, la complémentarité entre le texte du *Médiéviste et l'ordinateur* et celui des *Annales ESC*. Dans ce dernier, il affirmait manquer de place pour développer les modalités de calcul et de programmation d'une analyse factorielle des correspondances faite à l'aide d'une calculatrice programmable de poche. Il renvoyait alors, dans les notes de bas de page, à l'article dont il sera question ici, publié dans le bulletin de liaison deux ans auparavant, en 1979⁸³⁹. Dans cet article, il détaille la « *méthode générale* » de l'analyse factorielle, en insistant sur trois ensembles de calculs qu'il qualifie d'« *étapes essentielles* » :

- « 1. constitution d'une matrice carrée, symétrique, qui représente, selon certains critères, les distances entre les colonnes prises deux à deux.*
- 2. recherche des valeurs propres et vecteurs propres de cette matrice.*

⁸³⁸ Millet, 18.

⁸³⁹ Guerreau, « Analyse factorielle au moyen d'une calculatrice de poche programmable ».

3. *calcul des positions, sur les divers axes factoriels, des points-lignes, puis des points-colonnes.* »⁸⁴⁰

L'annonce du plan de ce court article de trois pages est suivie d'une série de remarques qui porte, entre autres, sur les lacunes de cet exposé.

Tout d'abord, Guerreau précise qu'il ne peut, faute de place, développer l'ensemble des calculs ni *a fortiori* les commenter précisément. Il faudrait selon lui « *une trentaine de page pour donner tous les programmes avec un commentaire vraiment complet* »⁸⁴¹. Il insiste ensuite sur le fait que les « *segments de programme* » qu'il présente dans ce texte sont ceux qu'il considère

« *nécessaires à une analyse factorielle des correspondances, en fonction de notre [son] expérience personnelle sur une calculatrice Texas-Instrument 58* »⁸⁴².

Comme dans l'article de Millet précédemment analysé, l'accent mis sur l'aspect expérimental, au sens de l'expérience vécue, et non de l'expérience scientifique, apparaît comme un marqueur essentiel de ce type d'article.

Guerreau développe ensuite ce qu'il qualifie d'« *observations pratiques* », en présentant pour chacune des étapes de la programmation la suite des calculs à mettre en œuvre. Pour la deuxième « *étape essentielle* », il présente par exemple l'enchaînement des calculs en ces termes :

« *Le calcul des valeurs propres s'effectue en diagonalisant la matrice par itérations, selon la méthode des "rotations de Jacobi". À chaque itération, on ne modifie que deux lignes et deux colonnes, de manière symétrique. On peut donc se contenter de travailler avec la partie supérieure de la matrice, ce qui diminue d'autant le nombre de mémoires nécessaires, mais interdit d'utiliser le programme préenregistré de multiplication de matrices. Comme la numérotation des mémoires est fixe, et qu'un programme capable de repérer automatiquement les mémoires concernées demande lui aussi de la place, on peut, avec la T.-I. 58,*

⁸⁴⁰ Guerreau, 11.

⁸⁴¹ Guerreau, 11.

⁸⁴² Guerreau, 11.

réaliser au mieux la diagonalisation automatique d'une matrice de rang 5. En procédant de manière semi-automatique, c'est-à-dire en introduisant au clavier à chaque itération le numéro des mémoires concernées, on peut diagonaliser une matrice de rang 7. Dans ce cas, une précision de l'ordre de 10^{-40} est obtenue en moyenne avec une cinquantaine d'itérations, soit environ une heure de travail. Pour calculer le $n^{\text{ième}}$ vecteur propre, on retranche de la diagonale de la matrice carrée obtenue à la fin de la première étape la $n^{\text{ième}}$ valeur propre et l'on considère les valeurs obtenues comme les coefficients d'un système d'équations linéaires à 6 lignes et 6 inconnues. Le programme préenregistré de résolution de systèmes d'équations linéaires permet d'opérer ces calculs très facilement. On ajoute aux 6 nombres obtenus un 1 comme septième valeur, et l'on normalise le vecteur. Tous les vecteurs propres devant être orthogonaux par construction, il est facile de vérifier si leurs produits scalaires sont bien nuls. »⁸⁴³

Cette longue citation nous permet de mettre en lumière plusieurs éléments absents de l'article des *Annales ESC*, qui sont pourtant au cœur de la méthode choisie par Guerreau dans ce dernier article.

Le premier élément frappant, que nous avons déjà pu remarquer dans l'analyse de l'article de 1981, concerne les limites de l'outil de calcul qu'il utilise. Cependant, par contraste, Guerreau ne se contente pas ici uniquement de mentionner les limites de la taille des matrices exploitables (5 ou 7 colonnes en fonction des cas), mais expose également les raisons qui expliquent cet état de fait. Elles reposent sur la mémoire interne de son outil de calcul. La capacité de cette mémoire oblige ainsi le médiéviste à prendre des décisions quant aux fonctions préprogrammées dans la calculatrice de poche qu'il pourra ou non mettre en œuvre. Comme nous l'avons vu au fil de l'analyse de l'article de Millet paru dans les *Annales ESC*, la distinction entre les tâches faites « à la main » et les tâches automatisées permet au lecteur de saisir la fonction de l'outil de calcul dans la réalisation des opérations. Ce type de précision nous montre comment Guerreau inclut

⁸⁴³ Guerreau, 12.

dans cet article des considérations qui permettent la réplication de la procédure à l'aide du même type de matériel⁸⁴⁴.

Le deuxième élément qui ressort de cet exposé de la procédure concerne la dimension temporelle du travail, qui est bien plus détaillée que dans l'article des *Annales ESC*. Guerreau indique notamment le rythme de travail qui est le sien (50 itérations par heure), précision qui apparaît pertinente pour quiconque aimerait se lancer dans ce type d'expérience. Ce type d'indication va dans le sens de l'analyse de l'élément précédent : ce texte doit permettre de donner les clefs au lecteur pour se lancer dans la programmation d'une analyse factorielle des correspondances à l'aide d'une calculatrice Texas Instrument, en profitant de l'expérience acquise et transmise ici par Guerreau. Le fait que l'auteur insiste, dans ce développement, sur la facilité de certains calculs et de certaines vérifications, nous semble être un autre indice de cette finalité de l'article.

3.1.3. Une première fonction des articles du *Médiéviste et l'ordinateur*.

Ces différents éléments nous permettent de dresser une première comparaison avec les autres articles analysés jusqu'ici. Tout d'abord, plusieurs points communs peuvent être identifiés parmi lesquels le fait que les articles publiés dans *Le Médiéviste et l'ordinateur* proposent des exposés d'une toute autre nature que ceux publiés dans les contextes de publication que nous avons qualifiés de généralistes (*Annales ESC*, *Cahiers de Civilisation Médiévale*). On retrouve toutefois quelques ressemblances avec l'article de Coq et Ornato étudié dans la deuxième partie de ce chapitre, notamment à travers la volonté de décrire une procédure clef en main pour les lecteurs. Sur ce point, il est hautement probable que le fait que les outils de calculs soient, dans ces deux cas, des instruments personnels (micro-ordinateur, calculatrice de poche), conduit les auteurs à partager les informations concernant la programmation, plutôt que de renvoyer au concours des informaticiens, comme avait pu le faire Millet dans ses deux articles par exemple.

Ensuite, si l'on se concentre sur le contexte de publication qu'offre *Le Médiéviste*

⁸⁴⁴ Il serait ici intéressant de connaître les utilisations qui ont pu être faites par d'autres médiévistes de ces indications. Toutefois, à notre connaissance, aucun autre médiéviste en France n'a programmé des analyses factorielles des correspondances sur ce type d'instrument de calcul.

et l'ordinateur, on remarque que de nombreuses différences apparaissent entre l'article de Guerreau et celui de Millet. Les contenus partagés par ces deux auteurs dans ces deux articles, quoique portant sur le même type de sujet (la mise en œuvre d'analyses factorielles des correspondances), ne sont pas du tout du même ordre. La finalité de ces deux articles est même tout à fait différente.

Cependant, il n'en demeure pas moins évident que le contexte de publication offerts aux auteurs par le *Médiéviste et l'ordinateur* les autorise à s'étendre plus précisément - mais pas plus longuement - sur les aspects informatiques de leurs méthodes, et, notamment, sur leurs aspects techniques et pratiques. Comme Millet, Guerreau offre au lecteur les éléments qui doivent lui permettre de se lancer par lui-même dans la mise en œuvre de la méthode. Là encore, la présentation méthodologique est toute tournée vers la possibilité d'une répliquabilité de la méthode par de futurs utilisateurs.

L'analyse de ces deux articles du *Médiéviste et l'ordinateur*, à laquelle il faut ajouter celle de l'article de Carpentier présentée dans la première partie de ce chapitre, fait apparaître plusieurs éléments communs, qui permettent de distinguer clairement *Le Médiéviste et l'ordinateur* de deux des trois autres contextes de publication qui nous intéressent dans ce chapitre. Dans les *Annales ESC* ou les *Cahiers de Civilisation Médiévale*, les détails de la méthode informatique sont rejetés entièrement dans les notes de bas de page. Au contraire, dans le *Médiéviste et l'ordinateur*, les aspects pratiques et techniques associées aux méthodes informatiques sont mis en avant, à partir de l'expérience personnelle des auteurs en tant qu'utilisateurs pionniers de ces méthodes.

Une première conclusion sur la particularité des articles du *Médiéviste et l'ordinateur* peut ainsi être formulée. Par contraste avec les textes publiés dans les autres revues académiques, les articles que nous observons ici ont pour fonction première de donner un aperçu de l'arrière-cuisine des historiens. Il participe donc d'un dévoilement des pratiques informatiques en histoire, comme il n'en existe pas ailleurs à notre connaissance.

Pour les auteurs, deux bénéfices scientifiques apparaissent. Le premier, c'est la possibilité de valoriser auprès des pairs une expérience d'un autre type que celle d'une pratique de l'histoire plus « traditionnelle ». Le second, c'est la chance offerte, par ce contexte de publication, de partager ses méthodes avec d'autres historiens, en donnant aux lecteurs la possibilité de répéter ce type d'expérience, que ce soit en s'inspirant des

conseils pratiques, comme ceux prodigués par l'article de Millet, ou en reproduisant des procédures de calcul, comme celles présentées par Guerreau.

3.2. Comparaison entre les parties méthodologiques et les « articles de fonds ».

L'examen de ces deux articles au prisme des quatre critères que nous avons dégagés dans la deuxième partie de ce chapitre nous permet de faire une nouvelle hypothèse : il existerait, dans les pages du bulletin, un genre d'article proche du genre des parties méthodologiques défini pour les trois autres contextes de publication. Pour confirmer cette nouvelle hypothèse, l'étude de deux articles ne sauraient toutefois suffire. Nous allons donc élargir l'étude, et prendre en compte un ensemble plus vaste de textes, afin de confirmer l'existence de ce genre. Nous nous appuierons sur les quatre critères dégagés jusqu'ici pour révéler les différences entre les « articles de fond » du *Médiéviste et l'ordinateur* et les articles publiés dans les autres journaux étudiés jusqu'ici.

3.2.1. Présentation étape par étape des opérations.

Le premier élément qui nous intéressera ici concerne la présentation des opérations étape par étape dans ces articles, ou, ce que nous avons pu qualifier de présentation procédurale. Le tableau ci-dessous représente la proportion du nombre d'articles présentant ce type de structure d'énoncé dans les quatre corpus qui nous intéressent, soit 114 articles :

Figure 27. Proportions d'articles comportant des éléments de présentation étape par étape.

	articles	avec présentation procédurale	proportion
<i>Cahiers de Civilisation Médiévale</i>	3	0	0
<i>Annales ESC</i>	18	5	0,28
<i>Le Médiéviste et l'ordinateur</i>	77	40	0,52
<i>Histoire & Mesure</i>	16	12	0,75
TOTAL	114	57	0,5

Si l'on compare les valeurs de la troisième colonne (« proportion »), on remarque que les articles du *Médiéviste et l'ordinateur* se différencie clairement des corpus d'articles des *Cahiers de Civilisation Médiévale* et des *Annales ESC*, avec plus de la moitié des articles qui comportent une présentation procédurale. Cependant, du point de vue de ce critère, on note que *Histoire & Mesure* se distingue encore plus fortement des deux autres corpus, avec trois quart des articles exposant les étapes du travail informatique (0,75).

3.2.2. Mentions de la matérialité des opérations.

Cette première observation ne saurait cependant suffire pour affirmer fermement quoi que ce soit. Comme nous l'avons vu, il est nécessaire pour qualifier la spécificité des articles d'examiner comment ce premier critère s'articule avec le second : les mentions de la matérialité des opérations au contact des ordinateurs. Nous l'avons vu, les articles du corpus CAHM1 comportant une partie méthodologique présentent tous des mentions explicites à la matérialité des opérations (ordinateurs utilisés, instruments de saisie ou méthode de stockage, etc.). Il n'est pas surprenant de retrouver des occurrences de ce type dans les « articles de fond » du *Médiéviste et l'ordinateur*. Ces éléments s'y présentent cependant plus systématiquement, puisqu'on y retrouve de très nombreuses mentions, par exemple, aux cartes perforées (68 occurrences), aux disques et disquettes (66 occurrences), aux écrans (56 occurrences), aux claviers (22 occurrences) ou aux bordereaux (15 occurrences).

Pour en permettre la comparaison, nous avons caractérisé les articles des quatre corpus en fonction du fait qu'ils faisaient ou non mention à la matérialité des opérations utilisant des ordinateurs. Le tableau ci-dessous permet de comparer la proportion de ces articles dans les différents corpus :

Figure 28. Proportions d'articles comportant des mentions à la matérialité des opérations.

	articles	avec mention de la matérialité	proportion
<i>Cahiers de Civilisation Médiévale</i>	3	0	0
<i>Annales ESC</i>	18	6	0,33
<i>Histoire et Mesure</i>	16	6	0,375
<i>Le Médiéviste et l'ordinateur</i>	77	61	0,79
TOTAL	114	73	0,64

Au contraire du tableau précédent, les proportions observées ici montrent que *Le Médiéviste et l'ordinateur* se distingue, sur la base de ce critère, beaucoup plus clairement des trois autres corpus. Sur la base de ce tableau, il est possible d'affirmer que l'une des particularités des textes de la rubrique « *Article de fond* » du bulletin de liaison est de faire mention, presque systématiquement, des aspects matériels des opérations, selon des modalités que nous avons analysées dans ce chapitre.

Il faut par ailleurs ajouter, comme nous l'avons vu dans le chapitre précédent, que dans le bulletin de liaison les auteurs publient dans de très nombreux cas des reproductions de documents de sorties des traitements informatiques dans le corps de leurs articles (tableau, listes, graphiques, segments de programme, etc.). Certains articles sont même exclusivement constitués de ce type de reproduction, comme c'est notamment le cas de deux d'entre eux publiés dans le numéro 11 de 1984 dédié aux traitements informatiques des testaments⁸⁴⁵.

3.2.3. Présence des auteurs dans les textes.

Troisièmement, la particularité de ces récits d'expérience se manifeste d'autant plus fortement à travers l'utilisation de la première personne du singulier dans les articles.

⁸⁴⁵ Voir Pierre-Clément Timbal, « Une grille de départ : celle du juriste », *Le médiéviste et l'ordinateur* 11, n° 1 (1984): 5-5 et Inge Shoups, « Une grille plus connue : celle de l'historien », *Le médiéviste et l'ordinateur* 11, n° 1 (1984): 5-7.

Une comparaison quantitative de l'utilisation du *je*, du *nous* et du *on* dans les quatre publications qui nous intéressent permet d'établir ce fait :

Figure 29. Distribution de quelques pronoms personnels (*je*, *moi*, *m'*, *nous*, *on*).

	<i>MO</i> (‰)	<i>Annales</i> (‰)	<i>H et M</i> (‰)
<i>je, moi, m'</i>	1.36	1.08	0.25
<i>nous</i>	1.67	3.08	2.49
<i>on</i>	4.38	3.56	4.61

La prévalence de l'utilisation de la première personne du singulier dans le corpus fait état, dans notre contexte, du fait que les auteurs rendent compte de leurs pratiques des ordinateurs sous l'angle de leur expérience personnelle. Cette analyse permet, comme dans le cas précédent, de distinguer les articles du fond du *Médiéviste et l'ordinateur* des deux autres corpus qui paraissent pourtant les plus proches.

Cet aspect des « articles de fond » du *Médiéviste et l'ordinateur* se combine avec de nombreuses occurrences des impressions et des ressentis des auteurs : « *ingratitude* »⁸⁴⁶, « *désespoir* »⁸⁴⁷, « *méfiance* »⁸⁴⁸, « *interminable et ingrate période de travail* »⁸⁴⁹, « *expérience ô combien éprouvante !* »⁸⁵⁰. Ce type de mentions est absent des deux autres corpus étudiés ici, et il nous apparaît même, intuitivement, qu'elles sont le plus souvent absentes des publications académiques traditionnelles. La finalité des « articles de fond » a ainsi, pour corrélat, l'usage d'un style personnel qui va dans le sens d'une révélation des aspects pas forcément positif de ce type d'expérience.

⁸⁴⁶ Carpentier, « Cadastres d'Orvieto », 7.

⁸⁴⁷ Paul Bratley, « Quelques réflexions et beaucoup d'information sur les problèmes de « sorties » », *Le médiéviste et l'ordinateur* 12, n° 1 (1984): 17.

⁸⁴⁸ Claudine Billot et Guy Lobrichon, « À propos des thesaurus », *Le médiéviste et l'ordinateur* 7, n° 1 (1982): 22.

⁸⁴⁹ Louis Stouff, « Un livre terrier arlésien complété de documents divers », *Le médiéviste et l'ordinateur* 3, n° 1 (1980): 13.

⁸⁵⁰ Yves Chartier, « Édition critique et micro-ordinateur », *Le médiéviste et l'ordinateur* 16, n° 1 (1986): 5.

3.2.4. Mentions des difficultés.

Enfin, un dernier point de comparaison permet encore de différencier ces quatre contextes de publications. Il concerne la mention des difficultés et des limites par les auteurs. Nous l'avons vu, ces éléments sont explicites dans presque tous les articles du corpus CAHMI qui possèdent une partie méthodologique. On retrouve aussi dans les articles du *Médiéviste et l'ordinateur* de nombreuses occurrences de ce type. Elles se manifestent, notamment, par un usage fréquent de la négation appliquée au verbe pouvoir :

« je n'ai pas pu retenir »⁸⁵¹

« je n'ai pu [...] dépasser »⁸⁵²

« des propriétés qui n'ont pu être exploitées »⁸⁵³

« l'analyse factorielle n'a pu prendre en compte »⁸⁵⁴

Ces formules linguistiques font transparaître les limites pratiques que rencontrent les auteurs au cours de telles entreprises.

Comme pour les trois critères précédents, nous avons établi pour chaque article du corpus si les auteurs faisaient ou non mention des difficultés dans leur texte. Le tableau suivant permet de comparer sur ce point les quatre corpus que nous avons définis :

⁸⁵¹ Monique Zerner, « Cadastres du Comtat Venaissin », *Le médiéviste et l'ordinateur* 3, n° 1 (1980): 8.

⁸⁵² Chartier, 6

⁸⁵³ Jacques Rouault et George Montpied, « Reconstitution et cartographie d'un parcellaire à partir de cadastres textuels du XV^e siècle », *Le médiéviste et l'ordinateur* 8, n° 1 (1982): 13.

⁸⁵⁴ Zerner, 9.

Figure 30. Proportion des articles comportant des mentions des difficultés ressenties par les auteurs.

	articles	avec mention des difficultés	proportion
<i>Cahiers de Civilisation Médiévale</i>	3	0	0
<i>Annales ESC</i>	18	6	0,33
<i>Le Médiéviste et l'ordinateur</i>	77	65	0,84
<i>Histoire et Mesure</i>	16	15	0,94
TOTAL	114	86	0,75

La lecture du tableau nous permet de mettre en lumière la spécificité des articles du *Médiéviste et l'ordinateur* et de *Histoire & Mesure* par rapport aux deux autres revues. Elle nous renseigne sur la fonction de ces articles, qui ne visent pas uniquement à donner des résultats, mais également à présenter aux lecteurs les difficultés que les auteurs ont pu rencontrer dans la mise en œuvre des méthodes informatiques.

Deux éléments permettent cependant de différencier les articles du bulletin de liaison et ceux de *Histoire & Mesure*. Le premier concerne le fait que les articles de ce second périodique présentent à la fois des résultats et des difficultés, par contraste avec les articles du *Médiéviste et l'ordinateur* qui ne comportent jamais de discours portant sur les résultats historiques des études menées par leurs auteurs.

Par ailleurs, les articles du *Médiéviste et l'ordinateur* se différencient par la mention explicite des échecs. Dans un cas, une erreur est même reconnue explicitement :

« [cela] nous a amené à commettre une erreur dans le principe du fichier »⁸⁵⁵.

Et dans un autre, l'auto-évaluation de l'auteur le conduit même à regretter son choix méthodologique :

⁸⁵⁵ Zerner, 8.

« Honnêtement je dois avouer qu'aujourd'hui, pour le type de recherche qui a été le mien, le recours à l'informatique ne me paraît pas rentable. »⁸⁵⁶

Cette reconnaissance explicite des échecs ou des ratés n'a pas été retrouvée dans les articles publiés dans d'autres contextes.

Cette analyse permet de mettre en avant l'existence, dans *Le Médiéviste et l'ordinateur*, d'un genre de texte similaire à celui découvert dans les parties méthodologiques des articles des revues académiques. Plusieurs différences apparaissent cependant : les articles du *Médiéviste et l'ordinateur* sont plus informels, et les auteurs se font beaucoup plus présents dans les textes, en intégrant plus largement leurs doutes, leurs échecs ou leurs ressentis des difficultés, afin d'exposer plus concrètement la nature des expériences liées à l'informatique en histoire médiévale. La présence de l'auteur dans le texte permet ainsi non seulement d'accrocher le lecteur, mais aussi de rendre compte plus concrètement des difficultés matérielles d'une méthodologie ancrée dans un environnement technique très contraignant.

3.3. Plan type des « articles de fond » du *Médiéviste et l'ordinateur*.

3.3.1. Définition du plan type.

Les pages qui précèdent ont permis de montrer que les « articles de fond » du bulletin revêtaient certaines caractéristiques qui les distinguaient des trois autres contextes de publications. Pour autant, nous n'avons pour le moment ni distingué les différents types d'articles qui existaient dans ce corpus, ni analysé les différentes fonctions de ces articles. En effet, si, comme nous l'avons dit, certains articles prennent la forme d'un récit d'expériences personnelles, en combinant certains critères que nous avons dégagés, d'autres doivent être qualifiés autrement, notamment parce qu'ils ne comportent ni présentation procédurale des opérations informatiques ni utilisation du *je* ou du *nous* à valeur de *je*. Pour ce faire, nous repartirons de l'examen des mêmes critères. Nous tenterons, sur la base de cette grille, de distinguer et de qualifier les différentes fonctions des articles publiés dans le bulletin de liaison.

⁸⁵⁶ Stoff, « Un livre terrier arlésien complété de documents divers », 13.

Le premier type d'articles qu'il nous est possible de qualifier est, comme nous l'avons dit, les *récits d'expériences personnelles*, qui combinent les quatre critères que nous avons identifiés. On retrouve dans le corpus des « articles de fond » du *Médiéviste et l'ordinateur*, 18 articles de ce type, soit plus d'un quart de ces textes. Sur la base de cet ensemble, il est possible d'identifier une forme de plan-type, que suivent, plus ou moins clairement les auteurs. Le plan pourrait être résumé ainsi :

1. présentation de la documentation et des données
2. objectifs du projet
3. justification du choix de l'informatique
4. difficultés rencontrées par rapport à ce choix (d'un point de vue technique - ce que l'on a appelé la matérialité -, ou du point de vue de la documentation)
5. présentation des opérations d'enregistrement ou de traitement des données étape par étape (avec parfois une illustration à l'aide d'un exemple)
6. retour sur cette expérience des ordinateurs (avec mentions des difficultés rencontrées et du ressenti des auteurs dans un style personnel)

Sur la base de ce plan, il est alors possible de qualifier plus précisément les écarts à ce premier type d'articles.

3.3.2. Variations autour du plan type et diversité des sous-genres d'articles.

Le premier de ces écarts consiste en l'utilisation d'un style impersonnel, c'est-à-dire en l'absence 1) d'utilisation du *je* ou du *nous* à valeur de *je + je* et 2) d'occurrence des ressentis personnels des auteurs. Ce nouvel ensemble n'est composé que de 7 articles. Ils partagent selon nous la même finalité que les 18 premiers : communiquer au lecteur sur une expérience de l'utilisation des ordinateurs en histoire médiévale, en précisant les étapes du travail, leur inscription dans un environnement matériel spécifique et les difficultés rencontrées dans la mise en œuvre de ces technologies. Le choix des auteurs de développer un style plus personnel rend compte d'une stratégie rhétorique qui permet de distinguer le contexte de publication que constitue le bulletin de liaison des trois autres. Cette option rhétorique ne modifie cependant en aucun cas la finalité des articles. Seule

la conclusion de ces textes apparaît réellement différente, puisque l'auteur ne fait pas apparaître son ressenti de l'expérience.

Un deuxième type de variation à ce plan-type peut être observé dans les articles qui ne mentionnent pas les difficultés rencontrées. Dans ce cas, deux aspects ne sont pas conformes au plan-type décrit ci-dessus (points 4 et 6). On compte dans ce cas 2 articles qui emploient un style personnel sans pour autant mentionner les difficultés rencontrées, et 6 articles rédigés dans un style impersonnel qui laissent eux aussi dans l'ombre ces difficultés.

Un troisième type de variation concerne les articles qui présentent les mêmes caractéristiques que ceux décrits précédemment mais qui ne mentionnent pas les aspects matériels de ces expériences. On trouve dans notre corpus un article de ce type employant le *je* et 6 articles de ce type rédigés dans un style impersonnel. Les articles de ce dernier type doivent cependant être distingués en deux catégories, puisque certains ne se rapportent pas au sous-genre du *récit d'expérience* que nous avons analysé jusqu'ici.

En effet, 3 de ces 7 articles se présentent comme des exposés généraux de certaines méthodes d'enregistrement ou de traitement des données, comme l'indexation documentaire ou le traitement *full-text*. Leurs auteurs sont tous des éditeurs, et ces articles visent à faire une présentation générale de ces méthodes, sans mention à la matérialité des opérations. Les similarités avec les articles que nous avons qualifiés de *récit d'expérience* sont notables, mais elles ne correspondent pas à une expérience spécifique d'application, ce qui explique notamment que le style impersonnel soit privilégié.

Un quatrième type de variation est observable dans les nombreux articles qui ne formulent aucune présentation procédurale des opérations, soit 37 articles sur les 77 que compte le corpus. Ces articles semblent, au contraire des précédents, être plus difficilement qualifiables de *récit d'expérience*, puisque les auteurs ne communiquent pas le déroulement des opérations. Certains d'entre eux font cependant mention des difficultés rencontrées, que celles-ci soient liées à l'équipement mobilisé, à la documentation mise en œuvre ou à l'organisation sociale des calculs, mais sans pour autant rendre compte ni de la matérialité ni d'une présentation procédurale des opérations.

Dans ce cas, le *récit d'expérience* se transforme en *retour d'expérience*. Les articles sont plus brefs, et les auteurs ne présentent que les conclusions pratiques qu'ils ont pu tirer de leur travail avec des moyens électroniques. Ces articles, que nous pouvons

qualifier d'exposé des problèmes, sont au nombre de 8 dans le corpus des articles de fond du *Médiéviste et l'ordinateur*.

Dans ce dernier ensemble de 37 articles, un cinquième type de variations peut encore être repéré : il s'agit des articles dans lesquels les auteurs ne présentent pas les étapes du travail, mais font tout de même référence à la matérialité et aux difficultés rencontrées. On compte dans notre corpus 21 articles de ce type, soit le groupe le plus important. Comme pour les autres articles ne contenant pas d'exposé des étapes du travail, il est difficile de qualifier ce type de texte de récit d'expérience, puisque les auteurs ne présentent pas le déroulement des opérations qu'ils suivent. Par ailleurs, sur ces 21 articles, 13 utilisent un style personnel, ce qui nous permet d'insister sur la difficulté de qualifier cet ensemble.

On y retrouve notamment deux articles qui présentent des *retours d'expérience* qui ne sont pas des expériences de recherche mais qui portent sur l'utilisation de nouveaux équipements techniques. C'est le cas du compte-rendu que Millet fait de sa première utilisation d'un ordinateur portable et de l'article d'Alain Demurger (1939-) qui témoigne de son utilisation d'un lecteur optique UNIVAC. Ces textes ont une finalité similaire aux retours d'expérience définis plus haut, puisqu'ils visent à proposer le point de vue d'un médiéviste utilisateur de nouveaux types d'équipements dans le but de partager une expérience. Par rapport au plan présenté ci-dessus, ce type d'articles se distingue par l'absence de l'élément 5) (présentation procédurale), mais semble cependant tout à fait similaire au premier ensemble de 18 articles de type *récit d'expérience* que nous avons présenté plus haut.

Parmi ces 13 articles employant un style personnel, on constate également l'existence de deux articles aux finalités tout à fait distinctes des articles précédents. Il s'agit de deux *appels à une organisation collective*, lancés d'une part par Millet et d'autre part par Ralph Withney Mathisen (1947-) dans le numéro 10 du bulletin, qui s'adresse aux lecteurs du bulletin dans le but de mettre sur pied des initiatives méta-collectives dans des secteurs bien définis de la recherche (base de données, types de sources). Dans ce cas, le plan des articles est tout à fait différent de ce que l'on a rencontré jusqu'ici. Ces deux articles se présentent donc comme des exceptions par contraste avec les autres textes de notre corpus, en dépit du fait que les auteurs utilisent la première personne et fassent mention des difficultés et de la matérialité des opérations.

Les 9 articles restants de cet ensemble de 13 articles peuvent quant à eux être qualifiés d'articles de présentation. Il faut par ailleurs y ajouter les 8 articles de l'ensemble de 21 articles que nous avons défini ci-dessus, dans lesquels les auteurs utilisent un style impersonnel. Dans ces 17 articles, on retrouve les cinq éléments du plan-type présenté ci-dessus, mais sans l'exposé des opérations étape par étape. Cette absence s'explique par le fait que ces articles ne visent pas à rendre compte d'une expérience vécue, mais à présenter un instrument de travail spécifique (programme informatique, base de données, catalogue, etc.). Dans ces articles, et en dépit de l'utilisation du *je* ou du *nous* à valeur de *je + je*, les auteurs ne se positionne pas comme des utilisateurs des moyens électroniques et informatiques, mais comme les producteurs d'instrument de travail. Le bulletin de liaison leur sert, notamment, à faire circuler des informations sur leurs projets de recherche.

À ces 17 articles de présentation il faut encore ajouter 6 articles qui ne mentionnent ni les étapes de travail, ni les difficultés, tout en faisant néanmoins référence à la matérialité des opérations. Ces articles se présentent sous une forme légèrement différente des 17 précédents, puisqu'en plus de l'absence du cinquième (présentation étape par étape) et du sixième élément (retour sur l'expérience) du plan-type, on ne retrouve pas non plus le quatrième élément (mention des difficultés ressenties). On compte ainsi, dans l'ensemble de notre corpus, 23 articles de *présentation spécifiques* d'instruments de travail. Il est encore possible d'ajouter à ces 23 articles les 3 articles que nous avons qualifiés plus haut de *présentation générale* des méthodes.

Les deux articles restants ne comportent, quant à eux, aucun des quatre critères que nous avons dégagés jusque-là, puisqu'ils sont constitués exclusivement de reproductions de documents de travail, accompagnées d'un court paratexte qui en présentent la fonction. Sur ce point, ces deux items apparaissent comme une version tout à fait édulcorée des types d'articles présentés ci-dessus, puisque seuls apparaissent ici les documents de travail élaborés par les collectifs.

Pour résumer cette longue et complexe analyse, nous avons établi un tableau conclusif, qui reprend la distribution des sous-genres d'articles en fonction des catégories que nous avons définies :

Figure 31. Répartition du nombre d'articles de la rubrique « Article de fond » du *Médiéviste et l'ordinateur* en fonction des sous-genre identifiés.

Sous-genre	Nombre
Récit d'expérience (personnel, impersonnel, technique)	39
- <i>personnel</i>	21
- <i>impersonnel (sans implication de l'auteur)</i>	16
- <i>technique (sans implication de l'auteur et sans mention des difficultés)</i>	2
Retour d'expérience	8
Présentation (générale, spécifique)	26
- <i>générale</i>	3
- <i>spécifique</i>	23
Appels	2
Autres	2
TOTAL	77

Conclusion.

La troisième partie de cette thèse, qu'inaugure ce cinquième chapitre, vise à étudier les modalités des échanges qui prennent forme dans les pages du bulletin de liaison *Le Médiéviste et l'ordinateur*. Dans l'introduction de ce chapitre, nous avons soulevé plusieurs questions qui tenaient à la forme que prenaient les articles publiés dans ce bulletin de liaison, en rapport à la fois avec le fait que certains acteurs proposaient, dans la deuxième moitié des années 1970, la mise en place d'un plan-type permettant de communiquer efficacement sur les éléments techniques et intellectuels des projets utilisant des ordinateurs en histoire médiévale, et sur le constat que faisaient d'autres

acteurs quant à l'impossibilité de publier des discours techniques ou matériels dans les revues académiques traditionnelles.

Il faut toutefois préciser que ces deux problématiques étaient intimement liées. D'une part, la définition d'un modèle de description des projets de ce type posait, de fait, le problème de la circulation de ces informations, du fait de l'absence de périodiques ou de revues qui en permettaient la publication. D'autre part, la création d'un média destiné à assurer ce type de circulation impliquait de se demander sous quelles formes ces informations devaient circuler. À ce titre, nous avons examiné dans le chapitre précédent comment les éditeurs du *Médiéviste et l'ordinateur* avaient défini des objectifs et mis en place des stratégies à même d'assurer la circulation d'informations, de pratiques et de *documents de travail* à partir de 1979. Il s'agissait dans le présent chapitre d'aborder cette problématique par l'autre bout : quel modèle de description des projets utilisant des ordinateurs s'était développé dans les pages de ce bulletin ?

Pour répondre à cette question, nous avons proposé une méthode descriptive inspirée de l'analyse de discours, qui reposait sur la comparaison des articles de plusieurs périodiques qui contenaient des énoncés portant sur ces pratiques des ordinateurs. Pour mettre en œuvre cette méthode, nous avons emprunté plusieurs concepts à la linguistique. L'un d'entre eux est celui de *formation discursive*, familier de certains historiens et historiens des sciences du fait de l'utilisation qu'en proposait Michel Foucault (1926-1984) dans *L'Archéologie du savoir* en 1969⁸⁵⁷. Dans ce livre, il définissait cette notion comme un ensemble d'énoncés présentant des points communs, au-delà de la diversité de leurs objets et de leurs auteurs, constituant un savoir à un moment de l'histoire, et dont il s'agissait d'établir le système de dispersion.

Une autre définition de cette notion avait par la suite été donnée par Claudine Haroche, Paul Henry (1906-1984) et Michel Pêcheux (1938-1963) dans un article de 1971. Ils y définissaient les formations discursives comme déterminant

⁸⁵⁷ Michel Foucault, *L'archéologie du savoir* (Paris, France: Gallimard, 1969).

« ce qui peut et doit être dit (*articulé sous la forme d'une harangue, d'un sermon, d'un pamphlet, d'un exposé, d'un programme, etc.*) à partir d'une position donnée dans une conjoncture donnée. »⁸⁵⁸

La comparaison entre ces deux définitions permet de relever leur complémentarité. L'approche de Foucault conduit à identifier une série d'énoncés portant sur un même objet (la folie, la sexualité ou l'ordinateur). Celle de Pêcheux précise qu'il faut y étudier les règles linguistiques attachées à cet espace discursif à la lumière de deux éléments : la « position » des locuteurs dans une « conjoncture donnée » et le genre dans lequel s'inscrivent ces discours (harangue, sermon, pamphlet, exposé, programme, etc.).

Nous proposons ici, pour conclure notre analyse, de reprendre ces deux points de la définition de Pêcheux, afin de mettre en lumière les évolutions de la formation discursive que nous avons étudiée pour la période 1970-1990.

Le premier concerne la « position » des locuteurs dans une « conjoncture donnée ». Le terme conjoncture - que l'on définira avec le dictionnaire Le Robert comme la « situation qui résulte d'un ensemble de circonstances » - nous apparaît beaucoup trop général et imprécis. Il nous semble toutefois possible d'affirmer que la « conjoncture » dans laquelle s'inscrivent les discours qui nous intéressent dans ce chapitre recoupe l'objet général de notre thèse. Nous avons examiné cette « conjoncture » sous plusieurs angles depuis le premier chapitre de ce travail, en montrant qu'elle résultait de la rencontre d'innovation techniques, d'ambitions historiographiques, de logiques d'organisation collectives, de difficultés matérielles et intellectuelles, etc.

Nous avons encore montré que cette « conjoncture » évoluait rapidement au fil de la période, et, en particulier, dans la deuxième moitié des années 1980 avec l'utilisation de plus en plus massive des micro-ordinateurs, des nouveaux modes de saisie, des nouvelles interfaces entre le chercheur et la machine (clavier, terminal vidéo, etc.), et des nouveaux périphériques comme le lecteur optique ou le lecteur de disquettes.

Dans ce contexte, nous avons montré que *ce que devait dire* les auteurs qui publiaient dans les différents périodiques académiques évoluait au fil de la période. Sous

⁸⁵⁸ Claudine Haroche, Paul Henry, et Michel Pêcheux, « La sémantique et la coupure saussurienne : langue, langage, discours », *Langages* 6, n° 24 (1971): 102.

cet angle, il nous est même possible de présenter une chronologie schématique de l'évolution de la formation discursive que nous avons définie. Les discours portant sur l'utilisation des ordinateurs les plus anciens que nous avons mentionnés dans ce chapitre datent du début des années 1970, et sont pour la plupart publiés dans des *revues généralistes*. Il n'existe en effet alors, à notre connaissance, aucun périodique spécifiquement dédié à la publication de ce type de contenu technique.

Les auteurs de ces articles publiés dans des revues généralistes sont alors des utilisateurs pionniers. Ils y présentent de nouvelles méthodologies, ainsi que les conclusions historiques qui résultent de la mise en œuvre de ces nouvelles méthodologies. Nous avons encore vu que les historiens co-signaient le plus souvent leurs articles avec un informaticien ou un statisticien, et que certains articles distinguaient même explicitement, en leur sein, entre les discours historiques énoncés par les premiers et les discours techniques et mathématiques énoncés par les seconds⁸⁵⁹.

On peut supposer ici que ce type d'articles prend alors cette forme en raison du fait que les historiens n'avaient, à cette période, pas d'autres choix que de travailler dans les centres de calculs, et donc, de collaborer avec un ou plusieurs spécialistes de l'informatique ou des statistiques. Ces derniers devenaient ainsi les garants de la validité de l'étude du point de vue informatique et mathématique, et leurs contributions aux articles constituaient une partie de *ce que devait dire* ces collectifs pour montrer le sérieux et la rigueur de ces études aux lecteurs des revues généralistes d'histoire.

Dans la période suivante, que nous proposons de faire commencer à la toute fin des années 1970, on constate une disparition de ce type d'articles co-signés. On remarque encore, sur la base du corpus mobilisé, que la grande majorité des articles ne mentionnent pas les dimensions méthodologiques liées à l'utilisation des ordinateurs, voire, ne mentionnent pas du tout l'utilisation de l'informatique. Pour rappel, sur les 159 articles que compte le corpus composé sur la base des revues généralistes (*Les Annales ESC*, *Cahiers de Civilisation Médiévale*), seuls sept d'entre eux font explicitement mention de l'utilisation des ordinateurs.

Dans la majorité de ces articles, nous avons montré que la *finalité* de ces discours était de justifier l'utilisation des ordinateurs et de souligner l'intérêt de ces méthodes pour

⁸⁵⁹ Voir par exemple Auffray, « I. Traitement informatique » et Guiral, « II. Informatique et péages du royaume de Valence (1494) ».

la discipline. Dans ces articles, on retrouvait ce type de discours dans les introductions et les conclusions, mais les détails techniques et matériels étaient rejetés en dehors du texte, dans des références en notes de bas de page. Parmi ces sept articles, nous avons cependant identifié deux d'entre eux qui proposaient des parties méthodologiques visant à détailler certaines de ces dimensions techniques et matérielles⁸⁶⁰. Les deux enquêtes qu'ils concernaient présentaient d'ailleurs un caractère commun, puisqu'elles mettaient toutes les deux en œuvre une procédure d'analyse des données originale dans le domaine : les analyses factorielles des correspondances. Ce dernier élément pourrait expliquer l'apparition de ce type de discours dans ce contexte de publication généraliste.

Par contraste, nous avons vu qu'apparaissaient à cette période plusieurs contextes de publication nouveaux, plus spécialement dédiés à l'exposé des dimensions techniques et matérielles de ces projets, dont *Le Médiéviste et l'ordinateur* est l'un des exemples les plus parlants. Dans ce contexte de publication, et au contraire des revues généralistes, nous avons montré que les médiévistes pouvaient développer des discours portant sur les dimensions techniques, matérielles et également des commentaires sur leurs ressentis personnels par rapport à ces expériences. Ces articles apparaissaient ainsi, dans ce contexte, et pour certaines enquêtes, comme le complément méthodologique des articles publiés dans les revues généralistes.

À partir de la deuxième moitié des années 1980, deux éléments viennent bouleverser cette formation discursive. Tout d'abord, la création d'une nouvelle revue académique en 1986, notamment dédiée à la publication d'études mettant en œuvre des moyens électroniques de calcul : *Histoire & Mesure*. Dans ce contexte de publication, nous avons vu que plusieurs articles présentaient des développements méthodologiques, et que ceux-ci contenaient deux éléments absents dans les revues généralistes : des détails sur la programmation et des précisions sur les procédures de calcul mises en œuvre. Par contraste avec les articles de fond du *Médiéviste et l'ordinateur*, nous avons montré que les articles de *Histoire & Mesure* publiaient également les résultats des enquêtes, en plus des dimensions méthodologiques susmentionnées.

⁸⁶⁰ Guerreau, « Analyse factorielle et analyses statistiques classiques » ; Millet, « La composition du chapitre cathédral de Laon au XIV^e siècle ».

Une dernière observation capitale permet encore de distinguer ces deux contextes de publication : les mentions qui y sont faites à la matérialité des opérations. Nous avons en effet montré que le *Médiéviste et l'ordinateur* présentaient une proportion beaucoup plus grande d'articles mentionnant les dimensions matérielles de ces enquêtes. Selon nous, cette différence s'explique en grande partie par le second élément qui vient bouleverser la « conjoncture » dans laquelle se trouve ces auteurs à partir de la deuxième moitié des années 1980 : le développement massif de la micro-informatique.

Pendant cette période, les modalités de l'utilisation des ordinateurs sont en effet absolument différentes de la période précédente, et c'est notamment tous les éléments matériels attachés au centre de calcul et à l'utilisation des cartes perforées qui disparaissent progressivement. L'ordinateur, sous sa forme personnelle, devient un instrument bien plus familier que dans les deux périodes précédentes, et il est donc possible de faire l'hypothèse que la nécessité de détailler ce qu'est, matériellement, un ordinateur disparaît.

De façon analogue, on peut attribuer l'importance attachée à la programmation et aux procédures de calcul dans les pages de *Histoire & Mesure* à cette évolution. Le médiéviste, qui travaille désormais seul sur son micro-ordinateur personnel, se doit d'être en mesure de programmer ses analyses par lui-même et partage ainsi cette dimension de son expérience à ses collègues dans ses articles, non seulement pour en permettre la répliquabilité mais également pour en donner une image fidèle.

Dans les pages du *Médiéviste et l'ordinateur*, cette évolution majeure est cependant difficilement observable sur la base de notre corpus. Pour saisir pleinement l'évolution de *ce qui devait être dit* dans les pages de ce périodique, qui sera publié jusqu'en 2004, il nous faudrait poursuivre l'analyse pour la période 1989-2004. Cette piste de recherche s'articule d'ailleurs avec d'autres, que nous avons déjà formulées au fil de cette thèse, et constitue l'un de nos projets futurs.

Le deuxième élément que soulevait Haroche, Henry et Pêcheux dans leur article de 1971 à propos de l'étude des formations discursives concerne un autre concept central pour notre étude, que nous avons emprunté à l'analyse de discours : le genre. Nous l'avons défini en introduction avec les termes de Swales comme « *a class of communicative events, the members of which share some set of communicative*

purposes »⁸⁶¹. Cette définition était par ailleurs articulée avec la notion de *communauté discursive* que nous avons définie et examinée, pour le cas des éditeurs, des auteurs et des lecteurs du *Médiéviste et l'ordinateur*, dans le chapitre précédent. L'analyse des règles qui régissaient *ce qui pouvait et ce qui devait être dit* dans les pages du bulletin de liaison, objet du présent chapitre, reposait ainsi sur l'articulation qui existait entre les objectifs de cette communauté discursive et la fonction du bulletin de liaison vis-à-vis de ces objectifs.

À ce titre, nous avons montré qu'il était possible de définir un plan-type pour les articles de fond du bulletin de liaison, et que l'étude de la combinaison des différents éléments de ce plan-type aboutissait à l'identification de plusieurs sous-genres : 1) des *récits d'expériences*, qui pouvaient prendre la forme de récit personnel, de récit impersonnel ou de récit technique ; 2) des *retours d'expériences*, qui ne visaient qu'à présenter des conclusions ; 3) des *présentations* de projets ou de méthodes, qui pouvaient être soit spécifiques, c'est-à-dire attachés à un projet précis, soit générales et 4) des *appels à une organisation collective*.

Cette approche descriptive nous permet d'établir de façon précise les trois fonctions qu'assurent les 77 articles du *Médiéviste et l'ordinateur* dans la formation discursive qui nous intéresse : 1) communiquer au lecteur sur une expérience de l'utilisation des ordinateurs en histoire médiévale, en précisant les étapes du travail, leur inscription dans un environnement matériel spécifique et les difficultés rencontrées dans la mise en œuvre de ces technologies ; 2) faire circuler des informations sur des méthodes ou des projets de recherche et 3) mettre sur pied des initiatives méta-collectives dans des secteurs bien définis de la recherche (base de données, types de sources).

Ces conclusions d'ordre linguistique nous permettent de répondre aux questions que nous avons soulevées en introduction à propos de l'intérêt que les acteurs pouvaient avoir à constituer un espace de discussion destiné à faire circuler des discours sur les aspects matériels et techniques des projets utilisant des ordinateurs en histoire médiévale. Deux éléments nous apparaissent tout à fait fondamentaux pour l'historien des sciences. Le premier concerne l'attachement que ces acteurs à la nécessité de partager des pratiques dans toutes leurs dimensions concrètes, ou en d'autres termes, de développer des discours sur la science telle qu'elle est *en train de se faire*, avec tout ce que cela implique de difficultés, d'échecs et de doutes. Le second élément sur lequel nous aimerions insister

⁸⁶¹ Swales, *Genre analysis*, 58.

concerne le fait que nous observons ici une nouvelle conséquence de l'utilisation des ordinateurs sur les cultures écrites. L'apparition de ces sous-genres d'articles dans notre formation discursive tient en effet à une volonté de certains acteurs, et plus particulièrement des éditeurs du *Médiéviste et l'ordinateur*, phénomène qu'il nous semble tout à fait intéressant d'étudier plus largement que dans notre seul contexte, et dont nous espérons que ce chapitre posera certaines bases.

L'étude des modalités des échanges entre les membres de cette communauté discursive ne saurait toutefois être complète sans l'analyse de certains débats qui prennent forme dans les pages de ce bulletin, et c'est justement à l'un d'entre eux que nous consacrons notre chapitre suivant.

Chapitre 6 : Conflits épistémologiques autour d'une utilisation des analyses factorielles des correspondances.

La deuxième partie de cette thèse était consacrée à l'étude d'un large groupe de médiévistes, d'horizons disciplinaires et de sensibilités historiographiques diverses, qui se regroupèrent au milieu des années 1970 afin de mettre un terme à l'isolement des utilisateurs d'ordinateurs en histoire médiévale. Ces premières formes d'organisations méta-collectives avait abouti à la création d'un bulletin de liaison, *Le Médiéviste et l'ordinateur* (chapitre 3). L'analyse des objectifs et des stratégies de ses éditeurs nous a ensuite conduit à conclure que cette publication attestait de l'existence d'une *communauté discursive* originale, qui se donnait pour finalité la constitution d'une *culture épistémique commune* entre les membres de cette communauté (chapitre 4).

Dans le premier chapitre de la troisième partie de cette thèse, nous avons dégagé l'existence, dans les articles de ce bulletin, de plusieurs *sous-genres* de textes adaptés à ces échanges d'une nouvelle nature. *Le Médiéviste et l'ordinateur* offrait, pour les utilisateurs des ordinateurs, un espace d'expression dans lequel il était possible, par contraste avec d'autres contextes de publication, de discuter des dimensions informatiques de ces méthodes (chapitre 5). L'objet de ce sixième chapitre est d'étudier, toujours sur la base du bulletin de liaison *Le Médiéviste et l'ordinateur*, les débats et les tensions qui se développèrent entre les premiers utilisateurs de l'informatique en histoire médiévale.

De quoi débattent les médiévistes dans ce bulletin de liaison ? Sur quels points s'opposent-ils ou s'accordent-ils ? Et quels sont les problèmes qu'il est possible d'identifier comme étant décisifs pour cette communauté ? Pour mieux saisir les nombreuses problématiques qui traversent cette communauté discursive, nous proposons de nous attacher dans ce chapitre à l'étude d'un cas particulier, que d'aucuns pourront percevoir comme une controverse.

Le corpus qui nous intéresse ici constitue l'un des rares exemples d'échanges fournis et contradictoires dont témoignent les pages du *Médiéviste et l'ordinateur* pour la

période étudiée (1979-1989)⁸⁶². Dans le numéro 5 du bulletin, dédié à l'analyse factorielle des correspondances, Alain Guerreau (1948-), propose d'appliquer la procédure d'analyse factorielle des correspondances à un tableau publié en 1980 par un autre médiéviste, Michel Banniard (1945-), dans un livre à usage général de la collection *Que sais-je ?*, consacré au haut Moyen-Âge occidental⁸⁶³.

Guerreau est diplômé de l'École Nationale des Chartes (ENC) en 1971 et agrégé d'histoire en 1973. En 1979, après une brève carrière dans le secondaire, il entre au CNRS et publie, la même année, son premier livre d'histoire médiévale : un ouvrage intitulé *Le féodalisme, un horizon théorique*. Il s'y propose de participer à la construction d'une « théorie du féodalisme »⁸⁶⁴ qui doit se fonder, selon lui, à la fois sur une critique de la « tradition française d'étude du système féodal » et sur la mise en œuvre d'une épistémologie plus proche des sciences sociales, et notamment de la sociologie, de l'anthropologie et de la linguistique.

Dans son article de 1981, dont il sera question ici, Guerreau s'empare d'un tableau de Banniard, publié à la fin d'un chapitre intitulé « Les Institutions ». Cette figure a pour fonction de résumer, de manière synoptique, les développements qui la précèdent. Ces derniers portent sur un ensemble de monographies consacrées aux phénomènes socio-économiques associés au développement du féodalisme dans neuf régions d'Europe occidentale, pour une période allant du VII^e au IX^e siècle⁸⁶⁵.

⁸⁶² On trouve deux cas de ce type pour la période qui nous intéresse. Mis à part celui dont il sera question dans ce chapitre, signalons le second exemple, plus tardif, qu'atteste la rubrique « *Courrier des lecteurs* » du numéro 9 du bulletin paru au printemps 1983. Ce deuxième cas concerne la pertinence d'une méthode de reconstitution semi-automatique de parcelles à partir de cadastres médiévaux. Il semble particulièrement intéressant de noter que ces deux controverses portent toutes les deux sur des applications des analyses factorielles des correspondances, et qu'elles impliquent toutes les deux Guerreau, qui se distingue non seulement par l'originalité des problèmes qu'il soulève que par sa volonté répétée de répondre aux critiques dans les pages du bulletin. Notons par ailleurs que nous préparons actuellement un article qui traite de cette seconde controverse.

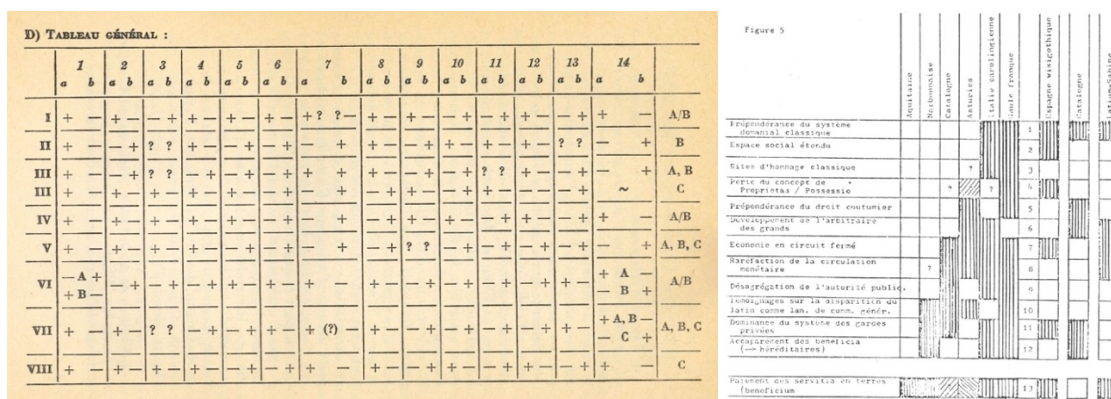
⁸⁶³ Michel Banniard, *Le haut Moyen Âge occidental* (Paris, France: PUF, 1980).

⁸⁶⁴ La référence de la version originale du livre est Alain Guerreau, *Le Féodalisme, un horizon théorique* (Paris, France: le Sycomore, 1980). En raison de l'impossibilité d'accéder aux bibliothèques pendant la rédaction de ce chapitre au printemps 2020, du fait de la crise sanitaire liée à la Covid 19, nous nous appuyerons ici sur la réédition numérique de ce livre mis en ligne par Guerreau lui-même en 2015 : Alain Guerreau, *Le Féodalisme : Un horizon théorique* (réédition numérique, 2015), 29. Les numéros de pages indiqués correspondent à cette dernière version.

⁸⁶⁵ Les monographies utilisées par Banniard dans ce chapitre sont les suivantes (présentées dans un ordre chronologique) : Claudio Sánchez-Albornoz y Menduïña, *Estudios sobre las instituciones medievales españolas* (México, Mexique: Universidad Nacional Autónoma de México Instituto de Investigaciones Históricas, 1965); Pierre Toubert, « Les structures du Latium médiéval: le Latium méridional et la Sabine du IX^e siècle à la fin du XII^e siècle » (Rome, Italie: École française de Rome, 1973); Élisabeth Magnou-Nortier, « La société laïque et l'Église dans la province ecclésiastique de Narbonne: zone cispyrénéenne,

Cependant, ce tableau est jugé illisible par Guerreau. Il y repère toutefois une structure des données qui peut lui permettre d'y appliquer des analyses factorielles des correspondances⁸⁶⁶. Dans un court article de quatre pages, Guerreau présente alors comment « simplifier »⁸⁶⁷ le tableau publié par Banniard à l'aide d' « une procédure de visualisation synthétique »⁸⁶⁸, qui repose sur la mise en œuvre d'analyses factorielles des correspondances. L'objet de cette procédure est alors de produire une nouvelle représentation des données de Banniard, et d'en faire « ressortir les grandes articulations qui ne sont pas lisibles directement »⁸⁶⁹.

Figure 32. Tableaux produits par Banniard (à gauche) et Guerreau (à droite).



Source : Banniard, *Le haut Moyen Âge occidental*, 68 et Guerreau, « Le Haut Moyen Âge factorialisé », 10.

Dans la rubrique « *Courrier des lecteurs* » du numéro suivant du *Médiéviste et l'ordinateur*, Pierre Bonnassie (1932-2005), auteur de l'une des monographies⁸⁷⁰ qu'utilise Banniard pour construire son tableau, remet en cause la pertinence de

de la fin du VIII^e à la fin du XI^e siècle » (Toulouse, France: Association des publications de l'université de Toulouse-Le Mirail, DL 1974); Jean-Pierre Poly, *La Provence et la société féodale: 879-1166* (Paris, France: Bordas, 1976); Pierre Bonnassie, « La Catalogne: du milieu du X^e à la fin du XI^e siècle » (Toulouse, France: Association des publications de l'université de Toulouse-Le Mirail, 1976 1975); Michel Rouche, « L'Aquitaine des Wisigoths aux Arabes (418-781): essai sur le phénomène régional » (Lille, France: Service de reproduction des thèses, 1977).

⁸⁶⁶ Cette structure, sur laquelle nous reviendrons dans la première partie de ce chapitre, correspond selon Guerreau à un tableau d'incidence. Ce type de tableau est défini comme suit par Guerreau dans l'article dont il est question : « ne comportant que des 1 et des 0 ». Alain Guerreau, « Le Haut Moyen Âge factorialisé », *Le Médiéviste et l'ordinateur* 5, n° 1 (1981): 7.

⁸⁶⁷ Guerreau, 11.

⁸⁶⁸ Guerreau, 11.

⁸⁶⁹ Guerreau, 11.

⁸⁷⁰ Bonnassie, « La Catalogne ».

l'ensemble de cette procédure. En élevant le débat au-dessus de l'aspect purement technique que l'article de Guerreau développe, Bonnassie exprime des doutes sur « *le choix des informations* »⁸⁷¹ mais également sur les résultats auxquels aboutit cette analyse. Dans ce même numéro, Guerreau lui répond⁸⁷², en emboîtant le pas de son interlocuteur sur les dimensions épistémologiques que soulève sa critique. Cet échange se déplace alors du terrain technique vers celui, plus général, de l'épistémologie de l'histoire et des sciences sociales.

Cette étude de cas révèle, comme nous le verrons, comment les usages des outils de calcul électroniques s'inscrivent dans des débats de fonds pour la discipline, en raison de l'objet historique discuté, des méthodes mathématiques employées et de la position sociale de ses protagonistes.

Le cadre chronologique et le sujet choisi par Guerreau pour présenter cette méthode synthétique s'inscrivent en effet au cœur de l'un des débats historiographiques fondamentaux pour les médiévistes de la seconde moitié du XX^e siècle : l'apparition et le développement des institutions féodales entre le VII^e et le XI^e siècle en Europe. Depuis des travaux de synthèse comme ceux de Marc Bloch (1886-1944) ou de François-Louis Ganshof (1895-1980) dans l'entre-deux guerres, les questions du rythme et de la chronologie du développement des institutions féodales animent les médiévistes⁸⁷³. Dans *La Société Féodale*, publié en 1939, Bloch présente comme modèle chronologique un développement progressif, et plutôt lent, du féodalisme entre le VII^e et le IX^e siècle⁸⁷⁴. En 1953, la thèse du médiéviste français Georges Duby (1919-1996) sur le Mâconnais⁸⁷⁵ vient complètement bouleverser ce modèle historiographique. Sur la base de ce travail, et en le comparant à d'autres monographies portant sur les régions du Nord de l'Europe, il développe une nouvelle théorie, celle d'une mutation rapide de la société médiévale aux

⁸⁷¹ Pierre Bonnassie et Alain Guerreau, « Échos du numéro 5 : Peut-on factorialisier le Moyen Âge ? », *Le Médiéviste et l'ordinateur* 6, n° 1 (1981): 18. Précisons ici que cette article, présenter dans les pages du *Médiéviste et l'ordinateur* comme un seul item, est en fait composé de deux parties distinctes : la première, sans titre, est écrite par Bonnassie, et la seconde, intitulée « En réponse... », rédigée par Guerreau.

⁸⁷² Bonnassie et Guerreau, « Échos du numéro 5 ».

⁸⁷³ Marc Bloch, *La société féodale*, 2 vol. (Paris, France: Michel, 1939); François-Louis Ganshof, *Qu'est-ce que la féodalité ?* (Bruxelles, Belgique: J. Lebègue, 1944).

⁸⁷⁴ Bloch, *La société féodale*.

⁸⁷⁵ Georges Duby, « La société aux XI^e et XII^e siècles dans la région mâconnaise » (Paris, France: A. Colin, 1953).

alentours de l'an mil. Celle-ci devient alors un véritable paradigme pour les médiévistes français des années 1960 et 1970⁸⁷⁶.

Ce modèle d'interprétation repose notamment sur l'examen de critères de différents types : 1) politiques tout d'abord, avec la dégradation des institutions publiques, la création de nouvelles formes de pouvoir privé et patrimonial et, donc, un affaiblissement de l'ordre public antérieur ; 2) économiques ensuite, avec notamment la dislocation de la paysannerie libre et alleutière, et la naissance d'un nouveau type de servage ; 3) idéologiques enfin, avec des réactions religieuses à la crise féodale, comme l'hérésie, ou la constitution chez les nouvelles élites militaires d'un imaginaire de la chevalerie.

Au début des années 1990, ce modèle est cependant rejeté par un certain nombre d'historiens, plus jeunes, qui défendent quant à eux une thèse moins radicale. Pour les médiévistes Dominique Barthélémy (1953-) ou Claudie Duhamel-Adamo (1937-), deux élèves de Duby, la révolution féodale n'a pas eu lieu, ou du moins pas de manière aussi fulgurante que ce qu'affirme Duby. Dans un article publié en 1992, Barthélémy remet en cause le paradigme de la révolution féodale, en insistant non seulement sur les apports de l'anthropologie et des études anglo-saxonnes, mais aussi sur les lacunes documentaires qui ont conduit à l'émergence de ce modèle. Il dénonce encore les effets réificateurs de la synthèse de Duby et de certaines monographies régionales, et notamment celle de Bonnassie sur la Catalogne, qui ont érigé, sur la base de phénomènes locaux, un modèle structurel du développement du féodalisme en Europe occidentale, sans prendre complètement en compte les particularités régionales de cette transformation⁸⁷⁷.

La procédure proposée par Guerreau, dont l'analyse sert de pivot à notre chapitre, s'inscrit dans ce contexte historiographique. Il propose justement dans son article de produire une visualisation synthétique de la distribution régionale de ces critères, pour une période qui comprend notamment ce moment si controversé qu'est la fin du haut Moyen-Âge. Même si, comme nous le verrons, la structure des données le pousse à exclure les X^e et XI^e siècles de son analyse, et donc, de fait, le moment clef de

⁸⁷⁶ Pour bien comprendre les enjeux de ce débat, lire l'article plus tardif : Dominique Barthélémy, « La mutation féodale a-t-elle eu lieu ? (note critique) », *Annales* 47, n° 3 (1992) : 767-77. Dominique Barthélémy y cite notamment deux ouvrages qu'il présente comme les versions les plus abouties de cette théorie « mutationniste », dont la thèse de Bonnassie, déjà citée. Nous reviendrons plus précisément sur ces questions historiographiques dans la deuxième partie de ce chapitre.

⁸⁷⁷ Barthélémy.

cette mutation féodale aux yeux de Duby, les critiques qui lui sont adressées s'insèrent tout à fait dans ce débat historiographique brûlant au milieu des années 1980.

Nous verrons plus loin comment Bonnassie, grand défenseur et théoricien de cette thèse mutationniste, reproche à la procédure de Guerreau de faire disparaître de l'analyse cette révolution de l'an mil, et, donc, de procéder à un « *aplatissement de l'histoire* »⁸⁷⁸. Les échanges entre les deux médiévistes apparaissent d'autant plus pertinents pour notre sujet qu'ils illustrent, sur la base d'une controverse historiographique majeure du XX^e siècle en histoire médiévale, des désaccords épistémologiques profonds sur les méthodes de synthèse en histoire, que nous développerons dans la deuxième partie de ce chapitre.

Le deuxième élément qui fonde l'intérêt de cette étude de cas repose sur la nature de la procédure statistique utilisée par Guerreau. L'analyse factorielle des correspondances a été développée en France, entre les années 1962 et 1965, par l'équipe du statisticien français Jean-Paul Benzécri (1932-2019)⁸⁷⁹. Ses travaux avaient alors pour objet de s'inspirer des méthodes proches de la *data analysis* utilisées dans le monde anglo-saxon pour en développer d'autres. Les origines de ce type de procédures statistiques sont cependant plus anciennes, puisqu'on les trouve dans les premières analyses à plusieurs variables du mathématicien britannique Karl Pearson (1857-1936), des psychologues britanniques Charles Spearman (1863-1945) et Cyril Burt (1883-1971) au début du XX^e siècle⁸⁸⁰, puis dans les bases théoriques posées entre 1931 et 1936 par le statisticien et économiste américain Harold Hotelling (1895-1973)⁸⁸¹. Toutefois, cette

⁸⁷⁸ Bonnassie et Guerreau, « Échos du numéro 5 ». Bonnassie insiste notamment sur la périodisation choisie par Banniard, qui nierait « *carrément la révolution féodale des années [...] 1020-1060* ». Nous verrons plus bas pourquoi ce reproche manque de pertinence au regard de la méthodologie choisie par Guerreau.

⁸⁷⁹ Pour un point de vue d'acteur sur l'histoire de l'analyse factorielle de correspondances, on regardera les livres de Benzécri lui-même, et notamment Jean-Paul Benzécri, *Histoire et préhistoire de l'analyse des données* (Paris, France: Dunod : Bordas, 1982) mais également Brigitte Escofier, « L'analyse factorielle des correspondances » (France, 1965). Pour un point de vue d'observateur voir par exemple Philippe Bonnet, *Chapitre 2. Pour une histoire sociale de l'analyse des données, La méthodologie de Pierre Bourdieu en action* (Dunod, 2015).

⁸⁸⁰ On pourra notamment regarder Karl Pearson, « On lines and planes of closest fit to systems of points in space », *The London, Edinburgh, and Dublin philosophical magazine and journal of science* 2, n° 11 (1901): 559-72; Charles Spearman, « "General Intelligence" Objectively Determined and Measured. », *American Journal of Psychology* 15 (1904): 201-93; Cyril Burt, « Experimental tests of general intelligence », *British Journal of Psychology* 3, n° 1 (1909): 94. Pour l'étude d'un observateur voir par exemple Olivier Martin, « Aux origines des idées factorielles [Des théories aux méthodes statistiques] », *Histoire & Mesure* 12, n° 3 (1997): 197-249.

⁸⁸¹ Sur ce point voir Harold Hotelling, « Analysis of a complex of statistical variables into principal components. », *Journal of educational psychology* 24, n° 6 (1933): 417; J.-P. Pages, F. Cailliez, et Y. Escoufier, « Analyse factorielle : un peu d'histoire et de géométrie », *Revue de Statistique Appliquée* 27, n° 1 (1979): 5-28.

première « révolution » au sein de la discipline statistique n'allait être pleinement réalisée qu'avec le développement des ordinateurs, qui permettaient enfin une application de ces procédures jusque-là essentiellement théoriques en raison de l'inexistence d'outils de calcul suffisamment puissants⁸⁸².

En France, à partir du début des années 1970, les méthodes de Benzécri s'exportent au-delà de la recherche fondamentale en mathématiques, vers la recherche appliquée en économie, notamment via l'Institut National de la Statistique et des Études Économiques, mais également vers la sociologie et la géographie⁸⁸³. Dans un article de 1981, Philippe Cibois (1941-), sociologue de formation qui œuvre à la circulation de cette procédure mathématique au sein du Laboratoire d'Informatique pour les Sciences de l'Homme (LISH) du CNRS, note que la « mode » de l'analyse des données gagne même des milieux non scientifiques. Cet ensemble de méthodes suscite toutefois des critiques qui tiennent en particulier à l'idéologie que Benzécri y attache et aux stratégies qu'il met en œuvre pour les diffuser dans les milieux académiques⁸⁸⁴.

Comme le signale Cibois, Benzécri rejette par principe la discussion avec ses collègues anglo-saxons spécialistes de la *data analysis* : il organise ses propres conférences et publie principalement dans des revues et bulletins qu'il a lui-même créés. Il forme aussi de nombreux étudiants, qu'il enjoint de collaborer, dans le cadre de stages avec des équipes scientifiques à la recherche de personnes compétentes dans l'utilisation des ordinateurs. Derrière ce prosélytisme, Benzécri développe par ailleurs, en lien avec ces méthodes d'analyse des données, une philosophie néo-thomiste que ne partage que très peu de ses élèves⁸⁸⁵. Le benzécrisme n'en devient pas moins très vite une « école » bien identifiée⁸⁸⁶.

C'est à cette période que certains médiévistes français s'emparent de l'analyse factorielle des correspondances. Dès la fin des années 1960, Christiane Klapisch-Zuber (1936-), ou encore Antoine Prost (1933-), nous l'avons vu dans les deux premiers

⁸⁸² Sur ce question voir l'article référence de Frank Yates, « Computers, the second revolution in statistics », *Biometrics* 22, n° 2 (1966): 233-51.

⁸⁸³ Bonnet, *Chapitre 2. Pour une histoire sociale de l'analyse des données*.

⁸⁸⁴ Dans un article de 1981, Philippe Cibois cite notamment les bureaux d'enquêtes et quelques journaux grand public comme *Le Nouvel Observateur* ou *Le Monde*, voir Philippe Cibois, « Analyse des données et sociologie », éd. Lebart et al., *L'Année sociologique* 31 (1981): 336.

⁸⁸⁵ Jean-Paul Benzécri, « Philosophie thomiste et connaissance mathématique de la Nature », *Aquinas* 11, n° 1 (1968): 37.

⁸⁸⁶ On pourra, ici encore, se référer au très bon article précédemment cité : Cibois, « Analyse des données et sociologie ».

chapitres de cette thèse, travaillent notamment avec l'un des étudiants de Benzécri, Michel Demonet⁸⁸⁷. Dans les années 1970, les médiévistes s'emparent plus largement de cette procédure, et l'on voit les publications présentant des résultats se multiplier au début des années 1980. En 1981, Hélène Millet (1944-) et Guerreau publient par exemple, dans le même numéro des *Annales. ESC*, deux études présentant des résultats d'analyses factorielles⁸⁸⁸. La même année, nous l'avons dit, *Le Médiéviste et l'ordinateur* y consacre d'ailleurs un numéro thématique⁸⁸⁹.

Au contraire des historiens quantitativistes des années 1950 et 1960⁸⁹⁰, les historiens des années 1970 et 1980 disposaient alors non seulement de la possibilité d'automatiser des procédures d'analyse uni et bivariées, mais également de procédures multivariées et des moyens pour les mettre en œuvre : 1) des études théoriques, en français – et notamment les deux tomes de la « bible » de l'analyse des données, publiés par Benzécri et son équipe en 1973⁸⁹¹ - ; 2) des outils de calculs électroniques – via les centres de calculs notamment⁸⁹² - ; mais aussi 3) des programmes informatiques, comme ceux mis à disposition des chercheurs par l'Association pour le Développement et la Diffusion de l'Analyse des Données (ADDAD), dont il a été question au chapitre 2⁸⁹³.

⁸⁸⁷ Michel Demonet est justement, au moment de sa collaboration avec Prost, l'un des stagiaires dont nous avons fait mention plus haut. Pour rappel, Demonet travaille dans un premier temps avec Prost. Ils mènent alors les premières analyses factorielles des correspondances appliquées à des documents historiques, en se concentrant sur l'étude du vocabulaire des proclamations électorales de 1881, 1885 et 1889 (voir chapitre 2). Il collabore ensuite avec l'équipe du laboratoire de lexicométrie de l'École Normale Supérieure de Saint-Cloud, et travaillera en parallèle avec le Centre de Recherches Historiques de la VI^e section de l'École Pratique des Hautes Études, se distinguant notamment par sa collaboration décisive à l'étude du *catasto* florentin de 1427 menée Christiane Klapisch-Zuber et David Herlihy (voir chapitre 1). C'est dans ce cadre qu'il publie avec Klapisch-Zuber les premiers résultats d'analyses factorielles des correspondances liées à des recherches en histoire médiévale : Christiane Klapisch-Zuber et Michel Demonet, « A uno pane e uno vino : la famille rurale toscane au début du XV^e siècle », *Annales* 27, n° 4 (1972): 873-901.

⁸⁸⁸ Ces deux articles sont publiés la même année dans les *Annales ESC* : Alain Guerreau, « Analyse factorielle et analyses statistiques classiques : le cas des ordres mendiants dans la France médiévale », *Annales* 36, n° 5 (1981): 869-912 ; Hélène Millet, « La composition du chapitre cathédral de Laon au XIV^e siècle : résultats d'une analyse factorielle », *Annales* 36, n° 1 (1981): 117-38. Ajoutons par ailleurs que nous avons longuement analysée ces deux articles dans le chapitre 4.

⁸⁸⁹ « Le médiéviste et l'ordinateur, N°5, printemps 1981. Introduction à l'analyse factorielle. » 5, n° 1 (1981).

⁸⁹⁰ En 1969, Emmanuel Le Roy Ladurie appelait déjà les historiens à utiliser plus massivement les analyses qu'il appelait alors « *multivariates* ». Voir à ce titre la préface de la réédition de : Adolphe d'Angeville, *Essai sur la statistique de la population française: considérée sous quelques-uns de ses rapports physiques et moraux* (Paris, France, 1969), 35.

⁸⁹¹ Bonnet, *Chapitre 2. Pour une histoire sociale de l'analyse des données*.

⁸⁹² Nous renvoyons ici à la fois à l'introduction de notre thèse et à la première partie de notre travail.

⁸⁹³ Voir encore Bonnet, *Chapitre 2. Pour une histoire sociale de l'analyse des données*.

La procédure d'analyse factorielle des correspondances relève justement de cet ensemble de nouvelles méthodes d'analyses multivariées rendues possible dans ce nouveau contexte. Cependant, de la structuration des données à la lecture des graphiques, en passant par la programmation des calculs, l'emploi de cette procédure requérait un ensemble de nouvelles connaissances, théoriques et pratiques, dont la compréhension par les autres membres de la discipline conditionnait l'acceptation. L'étude de la réception de cette procédure d'analyse des données en histoire passe donc nécessairement par des études de cas, comme celui que nous avons choisi dans ce chapitre.

Le troisième intérêt de cette étude de cas pour notre sujet tient à l'équipement mobilisé par Guerreau pour mettre en œuvre ces analyses. En effet, à l'inverse des collectifs que l'on a étudiés dans les chapitres 1 et 2, Guerreau préfère travailler en solitaire, en évitant au maximum, non seulement les collaborations avec les informaticiens, mais également le recours aux ordinateurs des centres de calcul. Les raisons de ce choix, que nous avons détaillées dans le chapitre 5, ont notamment pour fondement des convictions sur les méfaits de la division du travail qu'implique le passage dans ces centres.

Pour mener ses études en toute autonomie, Guerreau se forme de manière autodidacte aux statistiques comme à l'analyse de données, et se dote pour effectuer ses calculs d'une calculatrice programmable de poche Texas Instrument 38⁸⁹⁴. Comme nous allons le voir, ces choix politiques se reflètent dans les méthodes de calcul comme dans les méthodes d'analyse défendues par ce médiéviste. Nous découvrirons ainsi de quelle manière les relations entre environnement de travail et épistémologie se nouent dans ce cas particulier et original, par contraste avec d'autres modalités de l'utilisation des moyens électroniques de calculs présentées notamment dans les deux premières parties de cette thèse.

Pour dégager les problématiques qui sont au cœur de cette controverse, l'analyse qui suit porte sur les relations entre les interventions des trois médiévistes présentés plus haut : Guerreau, Banniard et Bonnassie. Sur la base de cet exposé succinct des enjeux de la controverse, la question centrale qui apparaît ne concerne pas seulement la notion de synthèse, mais également celle de simplification, qui lui est corrélée dans le vocabulaire

⁸⁹⁴ Pour un exposé, dans les termes de l'acteur, des raisons de ces choix scientifiques et techniques, le lecteur pourra notamment lire la première section de l'article suivant : Guerreau, « Analyse factorielle et analyses statistiques classiques ».

des auteurs. Nous l'avons dit, cette simplification constitue l'objectif principal que s'assigne Guerreau dans l'article initial de 1981.

Comment l'analyse factorielle des correspondances s'inscrit-elle dans les opérations de synthèse historique dans le cas qui nous intéressent ? Et quels sont les facteurs, pratiques et épistémologiques, qui permettent de comprendre l'opposition de Bonnassie à cette nouvelle manière de pratiquer cette opération de synthèse ?

Pour aborder ces questions de la manière la plus précise possible, et saisir par là ce qui est constitutif du conflit épistémologique qui se joue entre les trois auteurs, nous analyserons dans un premier temps les relations entre les raisonnements développés par Banniard et Guerreau, à partir des technologies intellectuelles utilisées par les deux médiévistes (tableaux, graphiques, matrices). Nous nous intéresserons dans un second temps à la controverse à proprement parler, c'est-à-dire au désaccord qui apparaît entre Guerreau et Bonnassie à propos de l'opération de synthèse historique, de ses limites, comme de son rôle dans l'épistémologie de l'histoire.

1. Le problème ancien du développement du féodalisme à l'aune d'une procédure mathématique nouvelle.

En 1980, Banniard propose dans son *Que sais-je ?* publié en 1980 un chapitre sur le développement du féodalisme dans le Sud de l'Europe de l'Ouest entre le VII^e et le XI^e siècle⁸⁹⁵. Il le conclut par un tableau, le seul de tout le livre, qui a pour objet, selon ses propres mots, de proposer « *un regroupement systématique en fonction des 1) régions ; 2) périodes ; 3) caractères considérés* ». Guerreau, dans son article de 1981, le présente, quant à lui, citant Banniard sans qu'il soit possible de savoir d'où provient cette citation, comme une tentative pour « *donner la clef d'une loi générale sur le développement de la féodalité* »⁸⁹⁶.

Nous analyserons les modalités de construction de ce tableau par Banniard, en cherchant à comprendre la fonction de ce procédé de visualisation dans son argumentation. Nous comparerons ensuite cette manière de faire avec la méthode

⁸⁹⁵ Banniard, *Le haut Moyen Âge occidental*.

⁸⁹⁶ Bonnassie et Guerreau, « Échos du numéro 5 », 17.

proposée par Guerreau dans son article du *Médiéviste et l'ordinateur* de 1981. Nous regarderons enfin quelles sont les principales contributions de ces deux démarches à la compréhension du développement du féodalisme en Europe occidentale pour la période considérée.

1.1. Le tableau de Banniard : un tableau pour opérer une synthèse.

Nous reproduisons ci-dessous les figures publiées en 1980 par Banniard, dans un petit manuel intitulé *Le haut Moyen-Âge occidental*. Quelques mots sur le contexte dans lequel l'auteur insère ce tableau éclaireront le rôle qu'il lui attribue. Cette analyse permettra, dans le même temps, d'illustrer les limites du procédé de visualisation choisi par Banniard, au regard de la fonction que ce dernier lui assigne.

1.1.1. Contexte de publication du tableau.

Banniard est un historien des évolutions des langues latines et romanes du haut Moyen-Âge, spécialiste de la culture orale. En 1980, il est agrégé d'histoire et maître-assistant en latin à l'université de Limoges. Il prépare concomitamment un doctorat d'État, qu'il soutiendra en 1988 à l'université Paris IV, sous la direction du latiniste Jacques Fontaine. Cette thèse sera publiée en 1992 sous le titre *Viva Voce. Communication écrite et communication orale du IV^e au IX^e siècle en Occident Latin*. Au cours de cette période, Banniard suit par ailleurs les séminaires de Jacques Le Goff (1924-2014) à la VI^e section de l'École Pratique des Hautes Études, auxquels participe aussi Guerreau, et ceux de Bonnassie à l'université Toulouse II-Mirail, haut lieu du développement de la thèse de la mutation de l'an mil présentée en introduction.

En 1980, Banniard est donc un jeune spécialiste du haut Moyen-Âge occidental, de seulement 35 ans, qui n'a d'ailleurs pas encore terminé son doctorat. L'exercice auquel il se plie pour ce *Que sais-je ?* constituera sa première publication en dehors des revues académiques. Nous allons le voir, cet exercice s'avèrera particulièrement périlleux. Ce court livre de 128 pages, dans lequel est présenté le tableau qui nous intéresse, est un ouvrage au statut particulier, puisqu'il doit être à la fois concis, pour des raisons qui

tiennent à l'histoire de la collection⁸⁹⁷, et aborder les nombreuses problématiques que soulève la périodisation du haut Moyen-Âge occidental chez les spécialistes. L'auteur doit non seulement aborder l'épineuse question du passage de l'Antiquité au Moyen-Âge, mais aussi exposer les difficultés rencontrées par les médiévistes dans l'étude de ce qui distinguerait la société du haut Moyen-Âge de celle du bas Moyen-Âge en « occident », à travers, notamment, la question du développement des institutions féodales.

Le public auquel cette collection s'adresse renforce cette difficulté, puisque les livres de la collection *Que sais-je ?* sont destinés aussi bien aux étudiants débutant qu'au grand public. Le danger pour l'auteur est donc double : il doit présenter un ensemble de connaissances générales, tout en s'exposant aux critiques de ses pairs quant aux contenus développés. En dépit de ces injonctions contradictoires, la collection *Que sais-je ?* constitue toutefois un ensemble de publications reconnues dans les milieux académiques⁸⁹⁸. *Le haut Moyen-Âge occidental* sera d'ailleurs cité comme référence dans plusieurs travaux académiques postérieurs⁸⁹⁹.

Le chapitre dont il est ici question est consacré aux institutions médiévales du haut Moyen-Âge occidental. Banniard, prisonnier de toutes ces contraintes, y propose une synthèse des conclusions de six monographies régionales dédiées à l'étude des régions du Sud de l'Europe de l'Ouest⁹⁰⁰, articulées autour de la question du passage d'un système non-vassalique à un système vassalique entre le VII^e et le XI^e siècle. Pour résumer les principaux éléments dégagés, et permettre aux lecteurs de les comparer simultanément, il présente sur une double page, en conclusion de ce chapitre, les deux figures qui suivent : une légende (à gauche) et un tableau (à droite).

⁸⁹⁷ Pour des raisons qui tenaient au prix du papier au moment de sa création en 1941 et à la volonté de rendre accessible ces livres au plus grand nombre, les créateurs de cette collection avaient eu l'idée de réaliser chaque volume à partir d'une seule feuille de 1,20 mètres sur 1,60 mètres, pliée en six fois pour obtenir 4 cahiers de 32 pages soit 128 pages. Voir Élisabeth Parinet, *Une histoire de l'édition à l'époque contemporaine, XIXe-XXe siècle* (Paris, France: Éditions du Seuil, 2004).

⁸⁹⁸ La collection *Que sais-je ?* est publiée par les Presses Universitaires de France. Pour découvrir une présentation de la collection dans les termes des années 1970 : Jean Feller, « « Que sais-je ? » : une collection qui a réponse à tout », *Communication & Langages* 28, n° 1 (1975): 71-75.

⁸⁹⁹ Il nous a été possible d'identifier 22 citations de l'ouvrage via Google Scholar, avant tout dans des contextes francophones.

⁹⁰⁰ Sánchez-Albornoz y Menduina, *Estudios sobre las instituciones medievales españolas*; Toubert, « Les structures du Latium médiéval » ; Magnou-Nortier, « La société laïque et l'Église dans la province ecclésiastique de Narbonne » ; Bonnassie, « La Catalogne » ; Poly, *La Provence et la société féodale* ; Rouche et Toubert, « L'Aquitaine des Wisigoths aux Arabes (418-781) ».

Figure 33. Légende (à gauche) et tableau (à droite).

A) PAYS OU RÉGIONS (d'ouest en est) :

I) Espagne (wisigothique) ; II) Asturies ; III) Catalogne ;
 IV) Aquitaine ; V) Narbonnaise ; VI) Gaule franque ; VII) Italie carolingienne ; VIII) Latium, Sabine.

B) ÉPOQUES :

A) VII^e siècle ; B) VIII^e-IX^e siècle ; C) X^e-XI^e siècle.

C) CRITÈRES :

1 a Maintien de l'écrit
 b Effacement de l'écrit
 2 a Maintien du droit écrit
 b Prépondérance du droit coutumier
 3 a Maintien du concept de *Proprietas/Possessio*
 b Perte du concept
 4 a Maintien d'une autorité publique (Etat)
 b Désagrégation de cette autorité
 5 a Maintien dans la discipline des grands
 b Développement de l'arbitraire des grands
 6 a Espace social étendu
 b Espace social réduit
 7 a Prépondérance du système domanial classique
 b Prépondérance de la petite propriété alleutière
 8 a Economie en circuit fermé
 b Economie d'échanges
 9 a Circulation monétaire régulière
 b Rarefaction de la circulation monétaire
 10 a Maintien des gardes « publiques »
 b Dominance du système des gardes privées (royales ou seigneuriales)
 11 a Paiement des *servitia* en argent
 b Paiement des *servitia* en terres (*beneficium*)
 12 a Permanence à titre précaire des *beneficia* (*possessio*)
 b Accaparement des *beneficia* → héréditaires (*proprietas*)
 13 a Rites d'hommage classique
 b Simple serment de fidélité
 14 a Témoignages sur le maintien du latin comme langue de communication générale
 b Témoignages sur la disparition du latin comme langue de communication générale

D) TABLEAU GÉNÉRAL :

	1		2		3	4		5		6		7		8		9		10		11		12		13		14		
	a	b	a	b	a	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b	
I	+	-	+	-	+	+	-	+	-	+	-	+	?	?	-	+	-	+	-	+	-	+	-	+	-	+	-	A/B
II	+	-	-	+	?	?	+	-	-	+	-	-	+	+	+	-	+	+	+	+	-	?	?	-	-	+	B	
III	+	-	-	+	?	?	-	+	-	-	+	+	+	+	+	-	-	+	?	?	+	-	-	+	-	+	A, B	
III	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	-	+	-	~	C	
IV	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	A/B	
V	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	?	?	-	+	-	+	-	+	-	+	A, B, C	
VI	-A	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	A/B	
VI	+	B	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	A/B	
VII	+	-	+	-	?	?	-	+	-	+	-	+	?	(?)	-	+	-	+	-	+	-	+	-	+	-	+	A, B, C	
VIII	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	C	

Source : Banniard, *Le haut Moyen Âge occidental*, 67-68.

1.1.2. La légende comme indice des logiques de construction du tableau.

Sur la gauche de la figure ci-dessus, la légende du tableau permet aux lecteurs de comprendre le codage utilisé par Banniard. Elle constitue aussi, pour l'observateur, un indice incontournable pour saisir les logiques de construction du tableau. Dans un premier temps, notre analyse de la démarche suivie par Banniard pour établir cette synthèse comparative portera sur cette partie de la figure, que nous reproduisons ci-dessous (figure 34).

On constate que le tableau agrège trois types de données, aux statuts différents. Les « *pays ou régions* », au nombre de huit, sont représentés en lignes par des chiffres romains, et classés d'Ouest en Est. Y sont associés des « *époques* », codées A, B et C, correspondant chacune à un ou deux siècles. Enfin, quatorze critères, divisés en deux modalités, l'une retenue comme étant la manifestation la plus avancée d'un phénomène (b), l'autre comme la moins avancée (a), sont disposés en colonnes. Par exemple, si l'on regarde la variable 5, le « *maintien dans la discipline des grands* » est considéré comme la manifestation la moins avancée du phénomène, et le « *développement de l'arbitraire des grands* » comme la manifestation la plus avancée. Ce codage implique que ces modalités s'excluent toujours l'une de l'autre, et répond ainsi à la logique du tiers exclu.

Figure 34. Légende du tableau de Banniard.

- A) PAYS OU RÉGIONS (d'ouest en est) :
- I) Espagne (wisigothique); II) Asturies; III) Catalogne;
IV) Aquitaine; V) Narbonnaise; VI) Gaule franque; VII) Italie carolingienne; VIII) Latium, Sabine.
- B) ÉPOQUES :
- A) VII^e siècle; B) VIII^e-IX^e siècle; C) X^e-XI^e siècle.
- C) CRITÈRES :
- 1 a Maintien de l'écrit
b Effacement de l'écrit
 - 2 a Maintien du droit écrit
b Prépondérance du droit coutumier
 - 3 a Maintien du concept de *Proprietas/Possessio*
b Perte du concept
 - 4 a Maintien d'une autorité publique (Etat)
b Désagrégation de cette autorité
 - 5 a Maintien dans la discipline des grands
b Développement de l'arbitraire des grands
 - 6 a Espace social étendu
b Espace social réduit
 - 7 a Prépondérance du système domanial classique
b Prépondérance de la petite propriété alleutière
 - 8 a Economie en circuit fermé
b Economie d'échanges
 - 9 a Circulation monétaire régulière
b Raréfaction de la circulation monétaire
 - 10 a Maintien des gardes « publiques »
b Dominance du système des gardes privées (royales ou seigneuriales)
 - 11 a Paiement des *servitia* en argent
b Paiement des *servitia* en terres (*beneficium*)
 - 12 a Permanence à titre précaire des *beneficia* (*possessio*)
b Accaparement des *beneficia* → héréditaires (*proprietas*)
 - 13 a Rites d'hommage classique
b Simple serment de fidélité
 - 14 a Témoignages sur le maintien du latin comme langue de communication générale
b Témoignages sur la disparition du latin comme langue de communication générale

Source : Banniard, 67-68.

Cette légende éclaire deux autres aspects fondamentaux de la démarche de Banniard. Le premier concerne le choix des données. Le cadre chronologique dont traite cette figure correspond à l'une des périodes qui fait débat au sein de la communauté des médiévistes depuis l'entre-deux guerres, puisque ce tableau soulève inévitablement la question suivante : quel est le rythme du développement des institutions féodales entre la fin du VII^e siècle, et le début du XII^e siècle qui expliquerait l'explosion du nombre de documents disponibles constatée par les spécialistes pour la période d'après le tournant de l'an mil ?⁹⁰¹ Si la fenêtre de tir pour étudier ce phénomène a pu varier en fonction des

⁹⁰¹ Pour un exposé récent sur ces questions voir l'article suivant, déjà cité : Barthélémy, « La mutation féodale a-t-elle eu lieu ? (note critique) ».

médiévistes - du IV^e au XVIII^e siècle pour Guerreau dans son livre de 1979 précédemment cité⁹⁰², aux seuls X^e, XI^e et XII^e siècles pour la synthèse de 1981 de Jean-Pierre Poly (1941-) et Éric Bournazel (1944-)⁹⁰³ - le chapitre de Banniard, auquel le tableau sert de conclusion, s'insère dans cette problématique générale qui traverse l'histoire médiévale dans la seconde moitié du XX^e siècle.

Sur ce sujet, Banniard présente une synthèse originale, qui exploite les résultats d'un programme de recherche entamé et poursuivi par d'autres médiévistes dans ces mêmes années⁹⁰⁴. Ce programme consiste à évaluer la pertinence et à affiner le modèle mutationniste proposé par Duby sur la base de son travail sur les archives mâconnaises, en s'appuyant sur l'étude d'autres aires géographiques. Dans son chapitre consacré aux institutions, Banniard reprend à son compte cette démarche, qu'il résume en ces termes :

*« En conséquence, après une assez longue période qui tendait à privilégier les aspects généraux du passage à la société féodale (mais on se fondait alors surtout sur des études de la Gaule et de l' « Europe du Nord et de l'Est »), la tendance actuelle est plutôt d'insister sur ses caractères particuliers en fonction des régions considérées. À propos de la Catalogne, P. Bonnassie vient de démontrer la fausseté de grands clichés. »*⁹⁰⁵

Sans revenir sur la manière dont les études régionales ont permis à Duby de développer son modèle pour les X^e et XI^e siècle, nous constatons ici que Banniard s'appuie sur des monographies non seulement récentes, mais portant exclusivement sur les régions du Sud de l'occident médiéval : 1) la thèse de Bonnassie de 1976 sur la Catalogne ; 2) les travaux de Michel Ruche (1934-) de 1977 sur l'Aquitaine ; 3) ceux de Élisabeth Magnou-Nortier (1930-) sur la Gaule du Sud, publié en 1974, 4) de Pierre Toubert (1932-) sur l'Italie de 1973 et 5) de Poly sur la Provence de 1976. Seule 6) la monographie de Claudio Sanchez-

⁹⁰² Guerreau, *Le Féodalisme, un horizon théorique*.

⁹⁰³ Jean-Pierre Poly et Éric Bournazel, *La mutation féodale: X^e-XII^e siècles* (Paris, France: Presses universitaires de France, 1980).

⁹⁰⁴ Pour un aperçu des travaux de Bonnassie et de certains de ses collègues médiévistes engagés sur ce terrain, on pourra notamment se référer à l'article suivant, résumé d'un colloque international organisé par le CNRS et l'École Française de Rome en 1978 : Élisabeth Carpentier, « Structures féodales et féodalisme dans l'Occident méditerranéen (Xe-XIIIe s.) », *Cahiers de Civilisation Médiévale* 26, n° 102 (1983): 141-46.

⁹⁰⁵ Banniard, *Le haut Moyen Âge occidental*, 83.

Albornoz (1893-1984) sur les Asturies est légèrement plus ancienne, puisqu'elle date de 1965.

Poursuivant la description de l'évolution des institutions entre le VII^e et le XI^e siècle, et mobilisant une littérature secondaire récente, dédiée à des régions du Sud, le tableau conclusif publié par Banniard fait la synthèse des modalités régionales du développement de la société féodale en s'inscrivant dans une dynamique historiographique amorcée dans les années 1970. Pour autant, sa synthèse se démarque du modèle de Duby, ne serait-ce que parce qu'elle élargit le cadre chronologique de la question aux VII^e, VIII^e et IX^e siècles.

Si l'on s'intéresse maintenant aux critères retenus par Banniard pour mener son analyse, on observe que cette sélection rejoint, elle aussi, certaines questions clefs du programme de recherche qui s'est développé à la suite de la publication du modèle de Duby. On y retrouve tout d'abord des critères qui portent sur les dimensions classiques de l'analyse de la féodalité : le politique et le juridique (critères 2, 3, 4 et 10). L'exemple le plus évident est à ce titre la reprise des deux principaux éléments de l'analyse classique des liens féodo-vassaliques avec, d'une part, l'évolution de l'élément « personnel » de ces liens, le rite d'hommage (critère 13 : « *le passage des rites d'hommages classiques au simple serment de fidélité* ») et, d'autre part, l'évolution de sa dimension « réelle », le *beneficium* (bienfait), parfois présenté comme l'équivalent du fief (critère 12 de la grille de Banniard : l'hérédité des *beneficia*, mais également critère 11 : paiement des *servitia* en argent ou en terre, c'est-à-dire sous forme de *beneficia*).

Banniard complète cette grille d'analyse classique en y ajoutant d'autres critères inspirés du programme de recherche dessiné par Duby dans un article de 1972, intitulé « Les sociétés médiévales : une approche d'ensemble ». On y retrouve des critères qui portent sur les formes du travail agricole (critère 7 : « *prépondérance du système domanial classique* » ou « *de la petite propriété alleutière* »), sur les aspects économiques de la société féodale (critères 8 : « *économie en circuit fermé* » ou « *économie d'échanges* » et 9 : « *circulation monétaire régulière* » ou « *raréfaction* ») et, plus largement, sur ses dimensions culturelles (critères 1 : « *maintien* » ou « *effacement de*

l'écrit » ; 6 : « *espace social étendu* » ou « *réduit* » ; et 14 : « *maintien* » ou « *disparition du latin comme langue de communication générale* »)⁹⁰⁶.

La légende du tableau révèle encore un autre aspect de la démarche de Banniard, qu'il est nécessaire d'éclairer par contraste avec la méthode que proposera Guerreau dans son article du *Médiéviste et l'ordinateur*. Cet aspect concerne la fonction épistémologique de ce tableau pour chacun des auteurs. En effet, le tableau de Banniard n'a pas pour objet de proposer une formalisation statistique des observations des monographies sur lesquelles il s'appuie, mais seulement d'en offrir, pour le lecteur, une vue synoptique.

Ceci dit, la fonction que Banniard attribue à cette figure ne saurait être complètement remplie, puisque, comme nous allons le voir, la représentation choisie entrave une lecture véritablement féconde du tableau. S'appuyant sur cette critique, Guerreau développe, quant à lui, l'intuition d'une toute autre fonction pour ce tableau, et se propose de le transformer selon une procédure que nous allons maintenant détailler.

1.2. Du tableau synoptique à la matrice d'incidence.

Nous l'avons mentionné en introduction, Guerreau justifie, dans son article de 1981, son intérêt pour le tableau de Banniard par la structure et le codage des données qui le constituent. Il considère en effet ce même tableau - et nous avons là un exemple typique de *gestalt switch* - comme présentant de fortes similarités avec les matrices d'incidences, utilisées comme formalisation initiale des données dans la mise en œuvre des analyses factorielles de correspondances. Pourtant, Guerreau insiste : ce tableau n'est pas tout à fait équivalent aux matrices d'incidence sur lesquelles d'autres acteurs avaient pu mener des analyses factorielles. Il identifie notamment deux ensembles de problèmes, qui vont

⁹⁰⁶ C'est, en substance, certaines des questions que se pose déjà le médiéviste belge Ganshof, historien du droit, dans son célèbre livre publié en 1944 : *Qu'est-ce que la féodalité ?*. La question de l'apparition des *miles* passionnera, quant à elle, Duby, quelques années plus tard. L'approche de ces questions sera ensuite profondément renouvelée avec le développement d'approches anthropologiques de ces phénomènes au milieu des années 1970. Sur ce point on regardera par exemple Elizabeth A. R. Brown, « The Tyranny of a Construct: Feudalism and Historians of Medieval Europe », *The American Historical Review* 79, n° 4 (1974): 1063-88 ; Susan Reynolds, *Kingdoms and communities in Western Europe, 900-1300* (Oxford, Royaume-Uni de Grande-Bretagne et d'Irlande du Nord: Clarendon press, 1984). Sur ces questions, on pourra encore lire la note critique, plus tardive, de Dominique Barthélémy portant un autre livre de Susan Reynolds : Dominique Barthélémy, « La théorie féodale à l'épreuve de l'anthropologie (note critique) », *Annales* 52, n° 2 (1997): 321-41.

l'obliger à transformer la structure des données pour en faire une matrice effectivement exploitable par son programme d'analyse factorielle de correspondances.

1.2.1. Transformation des données dans le but de mettre en œuvre des analyses factorielles.

Figure 35. Tableau synoptique publié par Banniard.

D) TABLEAU GÉNÉRAL :

	1		2		3		4		5		6		7		8		9		10		11		12		13		14		
	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b	
I	+	-	+	-	-	+	+	-	+	-	+	-	+	?	?	-	+	-	-	+	-	+	-	-	+	+	-	-	A/B
II	+	-	-	+	?	?	+	-	-	+	-	-	+	-	+	-	-	+	+	-	+	-	+	-	?	?	-	+	B
III	+	-	-	+	?	?	-	+	-	+	-	-	+	+	+	-	-	+	-	+	?	?	+	-	-	+	-	+	A, B
III	+	-	+	-	+	-	+	-	+	-	-	+	-	+	-	+	-	+	-	+	-	-	-	-	+	-	~	C	
IV	+	-	+	-	+	-	+	-	+	-	-	+	-	+	-	-	+	-	+	-	+	-	+	-	+	-	+	-	A/B
V	+	-	+	-	+	-	+	-	+	-	-	+	-	+	-	?	?	-	+	-	+	-	+	-	+	-	+	-	A, B, C
VI	-	A	+		-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	A/B
	+	B	-																										
VII	+	-	+	-	?	?	-	+	-	+	-	+	-	+	(?)	-	+	-	-	+	-	+	-	+	-	+	-	+	A, B, C
VIII	+	-	+	-	+	-	+	-	+	-	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	C

Source : Banniard, *Le haut Moyen Âge occidental*, 68.

Le premier problème identifié par Guerreau concerne l'hétérogénéité des données. En effet, contrairement à ce qui est nécessaire pour assurer la solidité épistémologique d'une analyse statistique, les observations que Banniard retient sur les « *pays ou régions* » (en chiffres romains) ne sont pas rigoureusement comparables, puisqu'elles concernent des époques différentes (en lettres capitales). Deux types de transformation des données sont nécessaires pour pallier à ce problème.

Tout d'abord, Guerreau choisit de ne conserver que les informations des lignes relatives à la période carolingienne (c'est-à-dire l'époque codée B qui va du VIII^e au IX^e siècle). Seules les données relatives aux régions suivantes sont donc prises en compte : Asturies, Catalogne, Aquitaine, Narbonnaise, Gaule franque et Italie carolingienne. Cependant, pour ne pas perdre les informations relatives aux trois lignes ainsi exclues (Espagne wisigothique, Latium-Sabine et Catalogne – seulement pour la période C),

Guerreau propose de les utiliser comme variables illustratives, c'est-à-dire en tant que variables exclues de la construction des graphiques d'analyse factorielle⁹⁰⁷. Par voie de conséquence, la période de l'an mil, source de tant de débats chez les médiévistes, se trouve exclue de l'analyse. Nous verrons cependant plus loin que, en dépit de cette décision, l'une des principales critiques formulées par Bonnassie dans sa réponse portera justement sur cette période.

Dans la continuité de cette première opération, Guerreau va ensuite prendre une autre décision importante : pondérer les modalités des variables « *pays ou régions* » au regard de leur « *importance approximative* »⁹⁰⁸. La Gaule franque, considérée comme plus importante se voit attribuée un poids de 5, l'Aquitaine et l'Italie un poids de 2, la Narbonnaise, les Asturies et la Catalogne un poids de 1. Cette pondération, qui repose apparemment sur les connaissances historiques de Guerreau, est-elle contestable ? Repose-t-elle sur le poids démographique des régions considérées ? Ou sur leurs importances politiques respectives ?

Une réponse tranchée à cette question ne peut être formulée sur la base de cet article de Guerreau. Cette opération de pondération demeure ainsi, pour nous, l'un des points obscurs de la méthode. Elle nous montre cependant le caractère fondamental du moment de la structuration des données et ses relations avec la construction d'un résultat en analyse de données. Pour illustrer la différence entre ces deux jeux de données pondérées et non-pondérées, Guerreau propose dans son article la reproduction de deux graphiques d'analyse factorielle : l'un portant sur les données brutes (non pondérées), l'autre sur les données pondérées. Comme nous le verrons, la comparaison entre ces deux graphes viendra dans un second temps justifier de l'intérêt de cette opération de pondération.

Le deuxième problème qui se pose à Guerreau dans cette entreprise tient à la formalisation et la symbolisation des modalités qu'a choisies d'utiliser Banniard. Contrairement à un codage statistique classique, c'est-à-dire, donnant une valeur univoque à toutes les données, certaines informations du tableau sont codées « ? » ou « ~ ». Il en résulte quelques lacunes dans le jeu de données présenté par Banniard. Cet état de fait s'explique par les lacunes des monographies régionales sur lesquelles il s'appuie

⁹⁰⁷ Comme nous le verrons, il lui sera possible de les projeter sur les graphiques une fois ceux-ci construits (voir 1.2.2.)

⁹⁰⁸ Guerreau, « Le Haut Moyen Âge factorialisé », 8.

pour construire son tableau. En effet, non seulement ces dernières ne donnent pas toutes des réponses précises au questionnaire construit par Banniard, mais la documentation utilisée par les auteurs de ces monographies est elle-même incomplète. La combinaison de ces deux éléments ne permet donc pas toujours à Banniard d'affirmer sans équivoque si un phénomène est avéré ou non.

Le codage qu'il définit en lien avec ce problème demeure cependant parfaitement adapté au procédé de visualisation qu'il choisit. C'est tout à fait le contraire pour Guerreau, qui cherche à appliquer une procédure d'analyse de données qui ne s'accommode pas du caractère lacunaire des observations. Pour éviter les cas indéterminés, Guerreau prend le parti de remplacer ces valeurs équivoques par « *des chiffres qui correspondent seulement à la répartition globale des 0 et des 1 entre les deux lignes appariées concernées* »⁹⁰⁹. Comme pour la pondération, Guerreau ne précise pas les modalités du calcul de ces valeurs, ce qui constitue donc une nouvelle zone d'ombre dans l'exposé de sa méthode. Signalons qu'il n'est cependant pas nécessaire de s'y attarder, puisque ce point technique n'entrera pas en compte dans les échanges qui suivront entre Bonnassie et Guerreau, sujet vers lequel tend notre analyse présente.

De l'ensemble de ces opérations résulte la matrice qui suit, sur laquelle va se baser le programme d'analyse factorielle de correspondances de Guerreau. Faisons toutefois remarquer qu'elle comprend aussi, dans les septième et huitième colonnes, certains résultats issues de l'analyse factorielle, sur lesquelles nous ne reviendrons que plus bas.

⁹⁰⁹ Guerreau, 8.

Figure 36. Matrice d'incidence construite par Guerreau sur la base du tableau de Banniard.

Figure 2

	Asturies	Catalogne	Aquitaine	Narbonnaise	Gaule franque	Italie carol.	Position sur l'axe 1	Position sur l'axe 2	Espagne wisig.	Catalogne	Latium-Sabine
Maintien du droit écrit	0	1	1	1	0	1	-0,244	0,529	1	0	1
Prépondérance du droit coutumier	1	0	0	0	1	0	0,488	-1,057	0	1	0
Maintien du concept de Proprietas / Possessio	0,67	0,67	1	1	0	0,67	-0,402	0,136	0	1	1
Perte du concept	0,33	0,33	0	0	1	0,33	0,809	-0,274	1	0	0
Maintien d'une autorité publique (Etat)	1	0	1	1	0	0	-0,816	-0,162	1	1	1
Désagrégation de cette autorité	0	1	0	0	1	1	0,816	0,162	0	0	0
Maintien de la discipline des grands	0	1	1	1	1	0	-0,194	0,377	1	0	0
Développement de l'arbitraire des grands	1	0	0	0	0	1	0,387	-0,753	0	1	1
Espace social étendu	0	0	0	0	1	1	1,314	0,280	1	0	0
Espace social réduit	1	1	1	1	0	0	-0,657	-0,140	0	1	1
Prépondérance du système domanial classique	0	0	0	0	1	1	1,314	0,280	1	1	1
Prépondérance de la petite propriété alleutière	1	1	1	1	0	0	-0,657	-0,140	0	0	0
Economie en circuit fermé	1	1	0	0	1	1	0,502	-0,401	1	0	1
Economie d'échanges	0	0	1	1	0	0	-1,005	0,802	0	1	0
Circulation monétaire régulière	0	0	1	0,2	0	0	-1,256	0,794	1	1	0
Raréfaction de la circulation monétaire	1	1	0	0,8	1	1	0,314	-0,199	0	0	1
Maintien des gardes publiques	1	0	1	0	0	0	-0,310	-0,650	0	0	1
Dominance du système des gardes privées	0	1	0	1	1	1	0,455	0,325	1	1	0
Paiement des servitia en argent	1	0,2	0	0	0	0	-0,395	-1,754	0	1	0
Paiement des servitia en terres (beneficium)	0	0,8	1	1	1	1	0,099	0,435	1	0	1
Permanence à titre précaire des beneficia	1	1	1	0	0	0	-0,667	-0,457	1	0	1
Accaparement des beneficia (→ héréditaires)	0	0	0	1	1	1	0,667	0,457	0	1	0
Rites d'hommage classique	0,4	0	0	0	1	1	1,022	-0,115	0	0	0
Simple serment de fidélité	0,6	1	1	1	0	0	-0,681	0,077	1	1	1
Témoignages sur le maintien du latin (comm. génér.)	0	0	1	0	0	1	-0,085	0,687	1	0,33	1
Témoignages sur la disparition du latin (comm. génér.)	1	1	0	1	1	0	0,042	-0,343	0	0,07	0

Source : Guerreau, 8.

1.2.2. Qu'est-ce qu'une analyse factorielle des correspondances ?

Avant de poursuivre cette étude de cas, et pour en comprendre plus précisément la méthode, il est nécessaire de présenter rapidement la procédure d'analyse factorielle

des correspondances. Sans revenir longuement sur son histoire, nous présenterons cette procédure d'analyse des données, en nous appuyant sur l'application qu'en propose ici Guerreau. Nous détaillerons les différentes étapes de calcul qu'elle implique, de la construction de la matrice d'incidence à la réalisation des graphes.

L'analyse factorielle des correspondances appartient aux méthodes d'analyses dites multivariées, qui ont pour objet de condenser le maximum d'informations apportées par des données organisées sous forme de tableaux statistiques (individus, variables, modalités) dans un espace bidimensionnel. De façon schématique, il est possible d'affirmer que les analyses factorielles ne donnent pas une réponse à un problème donné, mais fournissent, d'une réalité trop complexe pour être directement accessible à notre entendement des « images » simplifiées, qui prennent la forme de la meilleure projection possible sur un plan. Ces « images » permettent à l'analyste de saisir les traits essentiels de la structure générale des données, et, ainsi de la résumer de la façon la plus économe possible. Dans ce type de méthodes, la multiplication des analyses, c'est-à-dire la production de plusieurs projections sous forme de graphiques, qui sont autant d'images permettant d'explorer la structure des données, enrichit donc toujours la recherche.

Les conséquences de l'utilisation des analyses multivariées sur la recherche ont été largement discutées par les utilisateurs qui s'en sont emparés, notamment en rapport avec, répétons-le ici, l'apparition des moyens électroniques de calcul programmables et l'explosion des capacités de mémoires et de calculs de ces instruments. Dans une communication de 1974, le zoologiste irlandais Robert Edmund Blackith (1923-2000) comparait même les apports de ces nouveaux types d'analyses à ceux du microscope :

*« Some techniques open up new approaches to biological problems; microscopy did not simply enable us to see better what we had already seen dimly, it allowed us to discover micro-organisms with all that discovery implies. Similarly, multivariate analysis does not simply enable us to discriminate ever more effectively as we incorporate more characters into the discriminant, it opens up ways of handling masses of data meaningfully, data which would have defeated us by sheer bulk »*⁹¹⁰

⁹¹⁰ Robert Edmund Blackith, « Multivariate Analysis: Not so Much a Technique as a Weltanschauung », *Biometrics* 2, n° 30 (1974): 387.

Le titre de cette communication, « Multivariate analysis: Not so Much a Technique as a Weltanschauung », éclaire sous un autre jour le caractère décisif qu'a eu l'irruption de ces nouvelles méthodes statistiques dans les sciences du vivant, et plus particulièrement en biologie.

Comme le souligne Blackith, bien davantage qu'une nouvelle technique d'analyse, ou qu'un simple nouvel instrument pour la recherche, ces méthodes ouvrent la voie vers une nouvelle *manière de voir le monde* (Weltanschauung) aux chercheurs qui les utilisent. Les analyses multivariées apparaissent ainsi au début des années 1970 comme un vecteur potentiel de transformation profonde de l'épistémologie des sciences du vivant, et ce, en offrant un nouveau mode d'analyse des jeux de données issus de l'observation, qui passe par la production de nouvelles formes de visualisation.

De cette réflexion proposée par un zoologiste, que peut-on tirer pour notre propre sujet ? La première chose essentielle repose sur la nature particulière de ces nouvelles procédures statistiques⁹¹¹. Fortement liées aux développements des outils de calculs électroniques, elles font apparaître une nouvelle facette de l'impact des ordinateurs sur la recherche : leurs utilisations n'ont pas seulement des conséquences sur la quantité des données traitées par les chercheurs et le rythme auquel il est possible de les traiter⁹¹², mais aussi un impact que nous qualifierons ici, à défaut d'autres mots, de « qualitatif ». Les historiens utilisent en effet ces nouvelles procédures statistiques pour affiner leurs analyses, et ce même si les ensembles de données qu'ils étudient sont de faible dimension, comme c'est le cas avec l'exemple étudié dans ce chapitre.

Cette remarque sur les analyses multivariées rejoint certaines réflexions plus générales de l'une des pionnières de l'utilisation des ordinateurs au Royaume-Uni. Dans un article de vulgarisation de 1962, la linguiste et philosophe des sciences Margaret Mastermann (1910-1986), fondatrice du premier centre de linguistique informatique à l'université de Cambridge, compare notamment les ordinateurs aux télescopes («

⁹¹¹ De très nombreuses études portant sur l'histoire de l'utilisation des statistiques dans l'étude des phénomènes sociaux ont été publiées depuis, *a minima*, le début des années 1980. Nous ne pouvons faire référence à toutes ici, et nous ne citerons donc qu'un article, publié par un spécialiste influent dans les milieux académiques français, qui revient plus largement sur cette historiographie : Alain Desrosières, « L'histoire de la statistique comme genre: style d'écriture et usages sociaux », *Genèses. Sciences sociales et histoire* 39, n° 1 (2000): 121-37.

⁹¹² Pour des développements précis sur ces questions, nous renvoyons ici à la première partie de cette thèse (chapitre 1 et 2).

a telescope for the mind »), mettant elle aussi, d'une manière analogue à celle de Blackith, l'accent sur les apports de ces outils sur les opérations d'observation⁹¹³.

Ces discussions sur les apports « qualitatifs » des ordinateurs, menées alors dans les marges des disciplines concernées, sont reprises aujourd'hui par d'autres acteurs sous des angles plus critiques, face à l'évolution des pratiques informatiques en sciences humaines⁹¹⁴. Dans d'autres contextes, les analyses multivariées se présentent, encore aujourd'hui, pour certains, comme une réponse aux conflits méthodologiques posés par l'articulation souvent artificielle entre méthodes quantitatives et méthodes qualitatives⁹¹⁵.

L'application de l'analyse factorielle des correspondances par Guerreau qui nous intéresse ici présente, à la lumière de ces considérations, un intérêt tout particulier, en raison de l'outil de calcul qu'il choisit. Pour des raisons que nous avons développées dans le chapitre précédent, ce médiéviste utilise une calculatrice programmable de poche Texas Instrument, plutôt que de travailler avec des ordinateurs. Ce choix technique a pour conséquence le fait qu'il ne peut travailler que sur des tableaux de faibles dimensions, composés d'au maximum dix colonnes⁹¹⁶. Le type d'application de l'analyse factorielle proposée ici par Guerreau s'inscrit donc d'emblée comme un exemple frappant de ces usages « qualitatifs » des outils de calculs électroniques. De plus, cette application des analyses factorielles de correspondances par Guerreau permet de nuancer l'affirmation de Blackith présentée plus haut, et largement répandue, selon laquelle les procédures d'analyse multivariées présenteraient un intérêt seulement lorsqu'elles sont en prise avec de grandes masses de données.

Ces seules considérations théoriques et historiques ne sauraient évidemment permettre au lecteur de ce chapitre de comprendre précisément en quoi consiste l'analyse factorielle des correspondances. Précisons d'ailleurs que ce type d'analyse factorielle ne constitue que l'une des nombreuses procédures d'une famille de méthodes qui compte, entre autres, les analyses en composantes principales, les analyses des correspondances

⁹¹³ Masterman Margaret, « Freeing the Mind: Articles and Letters from The Times Literary Supplement during March-June, 1962 », 1962.

⁹¹⁴ Voir par exemple Willard McCarty, « A Telescope for the Mind? », in *Debates in the Digital Humanities*, éd. par Matthew K. Gold, NED-New edition (University of Minnesota Press, 2012), 113-36.

⁹¹⁵ Une réflexion en ce sens, fondée sur la pratique d'un autre type d'analyse factorielle, se trouve dans : Yann Renisio et Rémi Sinthon, « L'analyse des correspondances multiples au service de l'enquête de terrain », *Geneses* n° 97, n° 4 (27 novembre 2014): 109-25.

⁹¹⁶ Alain Guerreau, « Analyse factorielle au moyen d'une calculatrice de poche programmable », *Le médiéviste et l'ordinateur* 2, n° 1 (1979): 10-12.

multiples, les analyses factorielles de données mixtes, les analyses factorielles multiples et les analyses factorielles multiples hiérarchiques, dont chacune possède ses particularités techniques et ses domaines d'application.

L'analyse factorielle des correspondances se base, comme les autres procédures de cette famille, sur la notion probabiliste d'indépendance, utilisée, particulièrement en statistique, pour l'analyse de tableaux de contingence⁹¹⁷. Le principe général de la procédure consiste à construire, à partir des données d'un tableau d'observation, un tableau théorique qui correspondrait à la situation d'indépendance entre les variables en lignes et les variables en colonnes. Pour comprendre, en pratique, ce qu'est l'indépendance⁹¹⁸, prenons pour exemple les données de la matrice construite par Guerreau. Nous nous intéresserons ici exclusivement aux variables colonnes, c'est-à-dire aux « *pays ou régions* », qui sont au cœur de l'analyse du médiéviste. Précisons toutefois ici que la particularité de l'approche de Guerreau est de partir d'un tableau d'incidence, ce qui pose toute une série de problèmes théoriques sur lesquels nous ne reviendrons pas.

La première étape est de construire les « profils » de chacune des colonnes du tableau. L'objectif est de transformer le tableau des effectifs en un tableau de fréquences, pour permettre la comparaison des colonnes entre elles, en dépit des différences d'effectifs. Pour cela, on divise chaque valeur du tableau par son effectif total correspondant en colonne. On obtient alors le tableau suivant :

⁹¹⁷ Pour une présentation claire et adaptée aux non-mathématiciens nous renvoyons ici à Philippe Cibois, *L'analyse factorielle: analyse en composantes principales et analyse des correspondances* (Paris, France: Presses universitaires de France, 2000).

⁹¹⁸ Cibois.

Figure 37. Tableau de fréquences construit sur la base du tableau de contingence de Banniard.

	Asturies	Catalogne	Aquitaine	Narbonnaise	Gaule	Italie	Profil moyen
Maintien du droit écrit	0	0,0769	0,0769	0,0769	0	0,0769	0,0513
Prépondérance du droit coutumier	0,0769	0	0	0	0,0769	0	0,0256
Maintien du concept de <i>proprietas/possessio</i>	0,0515	0,0515	0,0769	0,0769	0	0,0515	0,0514
Perte du concept de <i>proprietas/possessio</i>	0,0254	0,0254	0	0	0,0769	0,0254	0,0255
Maintien d'une autorité public	0,0769	0	0,0769	0,0769	0	0	0,0385
Désagrégation de cette autorité	0	0,0769	0	0	0,0769	0,0769	0,0385
...
Total	1	1	1	1	1	1	1

Une fois ces nouvelles valeurs établies pour chaque case, qu'on dénomme dans le cadre de la procédure des « profils-colonnes », il devient possible de les comparer avec le « profil-moyen », qui correspond à la situation théorique d'indépendance. L'indicateur qui intéresse alors l'analyste est l'écart des valeurs entre les « profils-colonnes » et le « profil moyen ».

Pour le cas de la première ligne, qui concerne le critère « *maintien du droit écrit* », la Catalogne, mais aussi l'Aquitaine, la Narbonnaise ou l'Italie carolingienne présentent un écart de seulement 0,0256 (0,0769 contre 0,0513 pour le « profil-moyen »), contre 0,0513 pour les Asturies et la Gaule franque (0 contre 0,0513). Ces écarts attestent dans notre cas d'un intérêt scientifique certain de ce critère pour comparer ces régions. Au contraire, on observe que l'écart entre la fréquence de la case « *perte du concept de proprietas/possessio* » pour les Asturies, la Catalogne et l'Italie, et de la fréquence du « profil-moyen » est très faible (0,0254 contre 0,0255). On peut en conclure que ce critère ne pèse pas beaucoup pour différencier ces régions.

Contrairement à ce que l'on peut remarquer dans la matrice publiée par Guerreau et reproduite plus haut, nous n'avons pour le moment dégagé ni les axes, ni les positions de ces profils-colonnes sur les axes. C'est pourtant l'objectif de cette procédure que de rendre compte de ces différences sous la forme d'une projection, qui prend alors la forme d'un nuage de points représentant les profils (dans notre cas, pour les « profils colonnes »⁹¹⁹). Pour ce faire, la construction suit un principe théorique simple : le profil moyen est placé dans ce nuage et sera considéré comme son origine. S'il y a indépendance entre les variables, tous les profils seront confondus avec le « profil moyen », et le nuage sera donc réduit à un unique point : l'origine des axes. Au contraire, plus les données s'écartent de l'indépendance, plus les « profils » s'écarteront de cette origine.

Cependant, comme nous l'avons déjà mentionné, l'enjeu de cette analyse est d'évaluer le poids sur l'ensemble des données de toutes les variables en même temps. Le nuage est donc nécessairement multidimensionnel, et il faut donc, pour le projeter sur un plan à deux axes, établir quels sont les axes les plus à même de représenter le maximum d'informations du nuage de points. C'est là qu'interviennent nécessairement des outils de calculs électroniques (ordinateurs, calculatrices programmables de poche⁹²⁰). Ils permettent de calculer une suite d'axes orthogonaux d'inertie maximum, c'est à dire, résumant le plus possible d'informations de la matrice constituée plus haut⁹²¹.

Un programme d'analyse factorielle donne alors à son utilisateur un tableau des valeurs propres, qui lui indique les axes offrant le meilleur taux d'inertie de représentation du nuage sur un plan à deux axes. On peut alors demander au programme les coordonnées de chaque point colonne dans le repère défini par les deux axes choisis. L'une des particularités de l'analyse factorielle, par contraste avec les autres procédures de la même famille, est de proposer comme méthode de calcul des distances, non la distance

⁹¹⁹ L'une des particularités de l'analyse factorielle des correspondances est l'exploitation de ce que les spécialistes appellent sa dualité : on considère en effet, au contraire des autres types d'analyses factorielles (analyse des correspondances multiples et analyse en composantes principales notamment), que les lignes et les colonnes jouent des rôles exactement symétriques. L'inertie du nuage des profils lignes par rapport au profil moyen des lignes est égal à l'inertie du nuage des profils colonnes au regard du profil moyen des colonnes. Il est donc possible de représenter, sur le même plan, les points-lignes et les points-colonnes, et donc d'évaluer leurs poids respectifs dans la structure des données.

⁹²⁰ Nous ne reviendrons pas ici sur les modalités de programmation des calculatrices programmables de poche. Nous projetons toutefois de mener une étude sur ce sujet après notre doctorat, en nous intéressant plus généralement aux modalités de l'auto-apprentissage de la programmation par les historiens. Nous reviendrons sur ce point dans la conclusion de ce chapitre, et nous citerons quelques-unes des sources que nous avons identifiées.

⁹²¹ Pour une présentation du concept d'inertie voir Cibois, *L'analyse factorielle*.

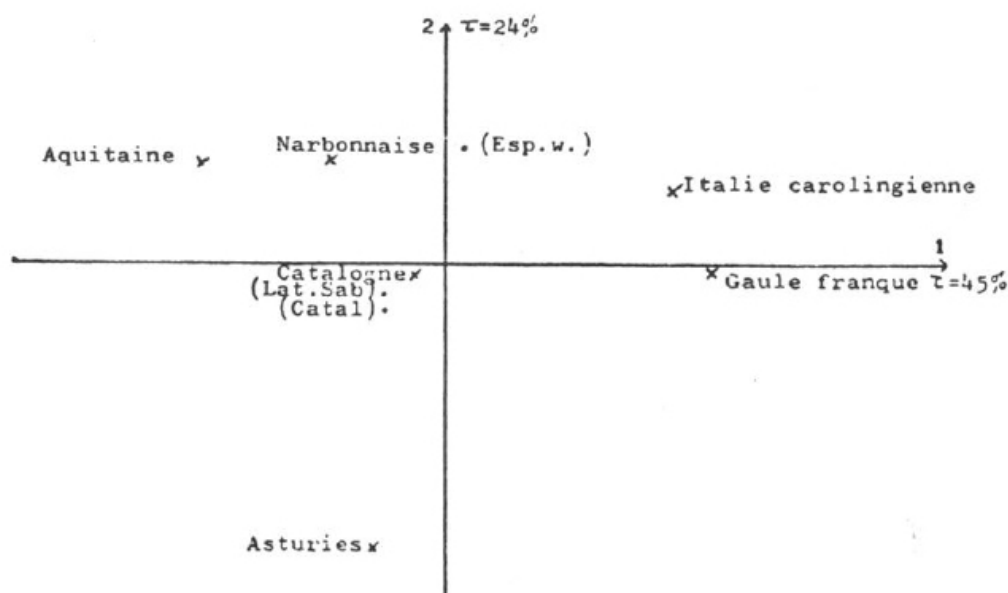
eulidienne, mais la distance du khi-deux⁹²². Il est dès lors possible de représenter le nuage des points-colonnes sur un plan. Pour vérifier si la représentation est bonne, l'analyste doit par ailleurs s'assurer qu'en choisissant différents axes, et en faisant ainsi varier la forme du nuage de point, la disposition des points-colonnes et des points-lignes ne change pas radicalement.

1.2.3. Les graphiques publiés par Guerreau.

Ci-dessous est reproduit le premier graphique d'analyse factorielle des correspondances construit par Guerreau, réalisé à partir de la matrice présentée au 1.2.1. qui concerne les données non-pondérées (figure 38). Il faut préciser que seuls les points-colonnes, correspondant donc aux « profils-colonnes », apparaissent sur le graphique. On voit encore que les valeurs d'inertie des axes obtenues à l'aide du programme sont reproduites sur le graphique (valeur τ), et que Guerreau fait également figurer sur le graphique les variables illustratives qu'ils distinguent entre parenthèses.

Figure 38. Graphique d'analyses factorielles des correspondances des données non-pondérées (plan factoriel 1-2).

Figure 3 Plan factoriel 1-2 de l'A.F.C. du tableau brut
(les variables illustratives sont indiquées entre parenthèses)

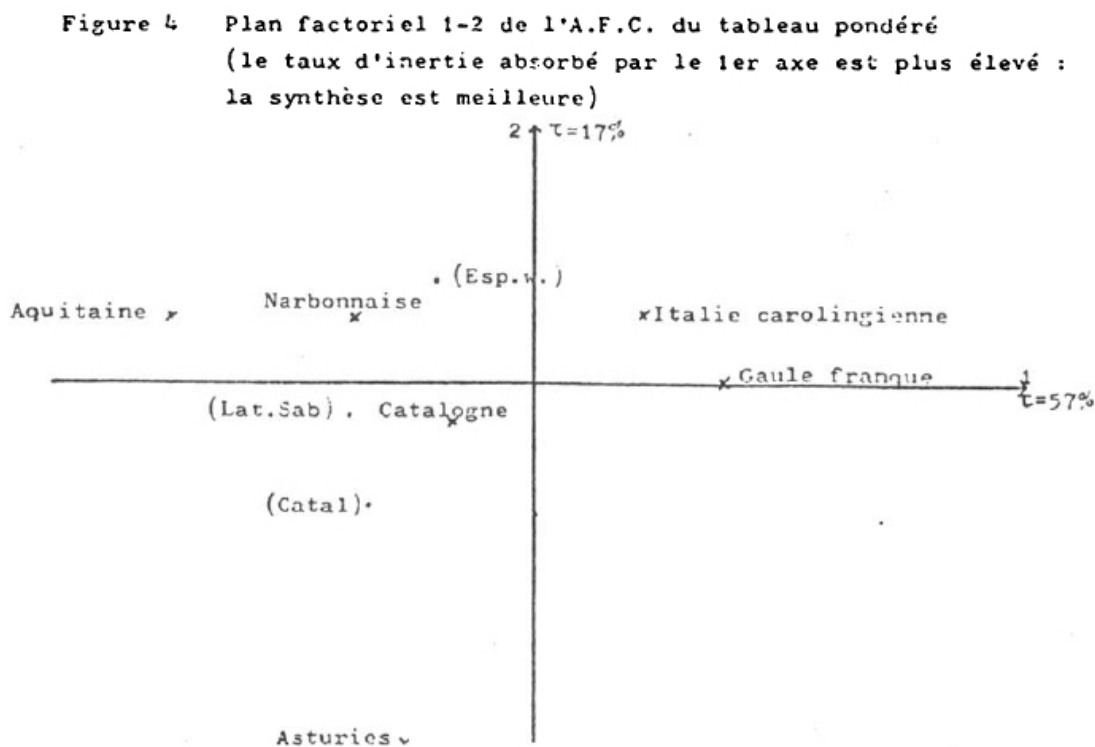


Source : Guerreau, « Le Haut Moyen Âge factorialisé », 9.

⁹²² Cibois.

Ci-après est reproduit le second graphique d'analyse factorielle publié par Guerreau dans son article, réalisé à partir, cette fois, des données pondérées selon la procédure dont il a été question plus haut (figure 39). Deux éléments de ces figures intéressent particulièrement Guerreau pour son analyse. Le premier, c'est la ressemblance entre les nuages de points des deux graphiques, qui montre que « *les deux analyses fournissent des résultats très voisins* ». Cette ressemblance atteste, d'après lui, de la robustesse de la représentation des données. L'analyse factorielle est donc concluante et il indique qu'il « *est dès lors légitime de penser qu'elle renvoie à une structure* »⁹²³.

Figure 39. Graphique d'analyses factorielles des correspondances des données pondérées (plan factoriel 1-2).



Source : Guerreau, 9.

Le deuxième élément que remarque Guerreau concerne la partie de l'inertie globale prise en compte par ces deux premiers axes, et indique ainsi à quel point ces graphiques captent ce qui différencie les profils-colonnes les uns par rapport aux autres.

⁹²³ Guerreau, 9.

Sur le second graphique par exemple, l'axe 1 (abscisse) synthétise particulièrement bien la majorité de l'information, avec un taux d'inertie de 57 %. À la lumière de ces éléments, Guerreau conclut qu'il existe un agencement ordonné des régions (une « *structure* »), comme suit (de droite à gauche sur le graphique) : Gaule franque, Italie Carolingienne, Catalogne, Asturies, Narbonnaise, Aquitaine.

Il conclut de cet agencement qu'il est

*« particulièrement significatif, puisqu'on voit sans difficulté qu'il permet de distinguer trois zones : 1) Le centre, le plus avancé (Gaule franque et Italie carolingienne), 2) une périphérie assez avancée (Catalogne et Asturies), 3) une zone intermédiaire attardée (Narbonnaise et Aquitaine). »*⁹²⁴

Pour décrire cet agencement, Guerreau emprunte ainsi un modèle « *centre-périphérie* », développé par l'historien français Fernand Braudel (1902-1985) dans son étude sur la dynamique du capitalisme pendant la période moderne, modèle repris et popularisé par l'historien américain, héritier de Braudel, Immanuel Wallerstein (1930-2019). Ce concept, dont se sont emparés, depuis les années 1980, les économistes, mais également les géographes et les spécialistes des sciences politiques, repose sur l'idée qu'il existe un déséquilibre dans le développement des phénomènes sociaux et économiques en fonction des régions⁹²⁵. Guerreau en reprend le principe, tout en s'appuyant sur le travail préalable de Banniard, qui distinguait dans son tableau, répétons-le, les manifestations avancées et attardées de phénomènes attestant du développement des institutions féodales.

Pour bien comprendre l'intérêt de cette conclusion de Guerreau pour son raisonnement, il ne faudrait cependant pas oublier que ces graphiques ne constituent pas le résultat final de l'analyse. En effet, et comme nous l'avons vu, les analyses factorielles ne sont que des moyens heuristiques qui permettent d'explorer un ensemble de données. Pour Guerreau, ces graphiques ne servent, au fond, qu'à mettre de l'ordre dans les données agrégées par Banniard, ordre qui n'apparaissait pas clairement dans le tableau de ce dernier.

⁹²⁴ Guerreau, 9.

⁹²⁵ Pour une réflexion récente sur le caractère désuet de ce modèle de description en histoire voir : Georges Saunier, « Quelques réflexions sur le concept de Centre et Périphérie », *Hypothèses* 3, n° 1 (2000): 175-80.

Dans ce cas, l'analyse factorielle de correspondances a permis à Guerreau de voir apparaître sur le graphique une structure centre-périphérie qui avait échappé à Banniard. Ce résultat, propre à la méthode d'observation choisie par Guerreau, permet la formulation d'une nouvelle hypothèse historique sur la distribution du phénomène féodal en Europe du Sud aux VIII^e et IX^e siècles. Une fois cette hypothèse formulée, Guerreau s'appuie sur celle-ci pour élaborer une meilleure visualisation synthétique que celle qui était proposée précédemment par Banniard.

1.3. Nouvelle visualisation synthétique des données.

1.3.1. Nouvelle représentation, nouvel outil heuristique ?

Les observations formulées par Guerreau sur la base des deux graphiques d'analyse factorielle des correspondances examinées ci-dessus lui permettent de construire le tableau que nous reproduisons ci-dessous (figure 40). Un seul coup d'œil permet de saisir de nombreuses différences qui le distinguent du tableau de Banniard, et de soutenir que le tableau présenté ci-dessous est plus lisible que le premier, présenté plus haut.

Le premier élément qui donne lieu à une meilleure lisibilité de ce second tableau tient aux opérations de transformation des données déjà décrites plus haut. Rappelons-le, Guerreau a fait le choix d'éliminer un certain nombre d'informations incomparables, mais aussi de supprimer le dédoublement de chacun des critères (« *avancés* » et « *attardés* »), pour ne conserver que le caractère le plus avancé. Ensuite, c'est la manière de représenter les informations dans ce nouveau tableau qui en facilite la lecture. Comme nous pouvons le constater, la présence ou l'absence des phénomènes est indiquée par un nouveau codage graphique : les cases sont hachurées lorsque le phénomène est attesté, et laissées vides lorsque ce n'est pas le cas. Un problème demeure toutefois : l'absence de légende ne permet pas de comprendre la différence entre les cases hachurées verticalement, et celles qui le sont dans le sens de la diagonale.

Un autre changement, plus fondamental, vient témoigner de l'intérêt de la méthode présentée par Guerreau dans cet article. Il s'agit de la réorganisation de l'ordre des colonnes à l'aune de la structure centre-périphérie des régions dégagée grâce à l'analyse factorielle des correspondances. En colonne, on retrouve ainsi, à gauche du

tableau, les régions de « *la périphérie intermédiaire attardée* » (Aquitaine, Narbonnaise), puis celle de « *la périphérie assez avancée* » (Catalogne, Asturies), et enfin à droite, celle du « *centre, le plus avancé* » (Gaule franque, Italie carolingienne). Figurent enfin, à l'extrême droite du tableau, les trois régions exclues de la construction du graphique et utilisées dans les analyses au titre de variables illustratives, l'Espagne wisigothique, la Catalogne des X^e et XI^e siècles, et le Latium-Sabine.

Figure 40. Tableau proposé par Guerreau en réponse à celui de Banniard.

Figure 5

	Aquitaine	Narbonnaise	Catalogne	Asturies	Italie carolingienne	Gaule franque	Espagne wisigothique	Catalogne	Latium-Sabine
Prépondérance du système domanial classique							1		
Espace social étendu							2		
Rites d'hommage classique				?			3		
Perte du concept de Proprietas / Possessio			?		?		4		
Prépondérance du droit coutumier							5		
Développement de l'arbitraire des grands							6		
Economie en circuit fermé							7		
Raréfaction de la circulation monétaire		?					8		
Désagrégation de l'autorité public.							9		
Témoignages sur la disparition du latin comme lan. de comm. génér.							10		
Dominance du système des gardes privées							11		
Accaparement des beneficia (→ héréditaires)							12		
Paiement des servitia en terres (beneficium)							13		

Source : Guerreau, « Le Haut Moyen Âge factorialisé », 10.

Un autre changement, plus fondamental, vient témoigner de l'intérêt de la méthode présentée par Guerreau dans cet article. Il s'agit de la réorganisation de l'ordre des colonnes à l'aune de la structure centre-périphérie des régions dégagée grâce à l'analyse factorielle des correspondances. En colonne, on retrouve ainsi, à gauche du tableau, les régions de « *la périphérie intermédiaire attardée* » (Aquitaine, Narbonnaise), puis celle de « *la périphérie assez avancée* » (Catalogne, Asturies), et enfin à droite, celle du « *centre, le plus avancé* » (Gaule franque, Italie carolingienne). Figurent enfin, à l'extrême droite du tableau, les trois régions exclues de la construction du graphique et utilisées dans les analyses au titre de variables illustratives, l'Espagne wisigothique, la Catalogne des X^e et XI^e siècles, et le Latium-Sabine.

Un troisième élément apparaît encore si l'on regarde de plus près cette figure : les lignes du tableau, qui correspondent aux critères, ont, elles aussi, été réordonnées. Ce changement dans la disposition des lignes permet à Guerreau de mettre en évidence des relations entre certains critères qui n'apparaissent ni dans le tableau de Banniard, ni dans les graphiques. Guerreau constate que trois ensembles de critères semblent fortement corrélés. Le premier concerne les lignes 1-2-3, qui représentent les phénomènes de « *prépondérance du système domanial classique* », d' « *espace social étendu* » et les « *rites d'hommages classiques* ». Une autre corrélation est mise en lumière pour les lignes 7 et 8, qui portent sur des phénomènes économiques : l'existence d'une « *économie en circuit fermée* » et la « *raréfaction de la circulation monétaire* ». Enfin, un troisième ensemble de variables corrélées est identifié par Guerreau, aux lignes 9, 11 et 12 (respectivement « *désagrégation de l'autorité publique* », « *dominance du système des gardes privées* », « *accaparement des bénéficiaires (→ héréditaires)* »), et regroupe des critères renvoyant à des observations sur des phénomènes politiques et militaires.

Ce second tableau fait ainsi apparaître beaucoup plus clairement que le précédent les grandes articulations des données que révélaient les graphiques d'analyses factorielles. Sur ce dernier tableau, il devient par exemple évident qu'il existe une opposition frappante entre l'Aquitaine, dont une seule case est hachurée, et la Gaule franque, pour laquelle au contraire une seule case est laissée vide.

La construction de ce tableau fait cependant apparaître deux autres problèmes pour Guerreau. Ils concernent deux lignes en particulier : 1) la ligne 4, qui, selon

Guerreau, « ne paraît guère pouvoir se distinguer de la ligne 5 »⁹²⁶ et 2) la ligne 13, qui, soulevant un paradoxe, est par suite détachée du reste du tableau. En effet, selon lui, il n'est pas possible d'affirmer qu'il existe une raréfaction de la circulation monétaire alors même que le paiement des *servitia*, c'est-à-dire les services dus au suzerain, se faisait en argent ? Cette contradiction apparente le conduit à conclure qu'il existe, sur ce point, une erreur dans le processus de collecte des données, sans qu'il ne précise toutefois l'erreur qu'il identifie. La réorganisation des lignes montre que le moment de la construction du tableau, comme le tableau lui-même, sont pour Guerreau des outils heuristiques, au contraire de Banniard qui conçoit la construction du tableau comme la conclusion d'une analyse et le tableau comme une image illustrant celle-ci.

1.3.2. Nouvelle synthèse ? Nouveaux résultats ?

Cette analyse met en évidence que le tableau réalisé par Guerreau constitue un outil de visualisation synthétique plus lisible que celui que proposait Banniard dans son *Que sais-je ?*. Il assume ainsi, lui aussi, une fonction synoptique. Par contraste, il s'inscrit cependant plus clairement dans un argumentaire relatif à un problème historiographique. Toutefois, il faut insister sur le fait qu'il a aussi une portée plus limitée.

En effet, ce second tableau ne concerne que la période carolingienne, et non plus l'ensemble du haut Moyen-Âge comme le premier (VII^e-XI^e siècle). De ce fait, il ne vise plus à proposer une réponse à la question historiographique centrale qui intéressait Banniard, à savoir, celle du rythme de développement des institutions féodales entre le début du VII^e siècle et la fin du XI^e siècle. Plus largement, c'est même toute l'ambition diachronique de la démarche de Banniard qui disparaît avec la procédure de Guerreau, puisque ce second tableau ne permet de comparer efficacement les régions que de manière synchronique. La procédure mise en place par Guerreau aboutit donc bien à une autre synthèse, qui ne relève pas exactement de la question à laquelle le tableau synthétique de Banniard visait à répondre.

Plus profondément, c'est en fait la nature de la synthèse proposée par Guerreau qui semble différente. Par contraste avec Banniard, Guerreau utilise le tableau qu'il a construit pour formuler des hypothèses sur l'existence de corrélations entre certaines

⁹²⁶ Guerreau, 9.

lignes. En montrant qu'il existe des relations particulières entre certains critères et certains ensembles de régions, Guerreau formule des axes de recherches pour les historiens de l'Europe du Sud au Moyen-Âge. Quelle est la particularité de l'Italie carolingienne et de la Gaule franque au regard de la prépondérance du système domanial classique ? Ou encore, qu'est-ce qui explique l'existence d'une économie en circuit fermé dans l'ensemble des régions, sauf en Aquitaine et en Narbonnaise ? Sans juger ici de la pertinence historique de ces questions, la procédure de Guerreau a le mérite de les faire apparaître.

La synthèse que propose ne se présente donc pas comme l'aboutissement de son raisonnement, puisque le tableau permet au lecteur de faire émerger des questions historiographiques originales (ce n'est « *qu'un moyen* » écrit-t-il⁹²⁷). Cette procédure, que Guerreau qualifie, dans son article, de « *procédure de visualisation synthétique* », a donc pour finalité de transformer le tableau initial en outil heuristique.

À ce titre, l'analyse factorielle de correspondance est présentée comme un « *outil d'un très bon rendement* »⁹²⁸, « *appelé à se développer dans des proportions considérables* »⁹²⁹, puisqu'elle permet « *l'exploitation dans des conditions optimales* »⁹³⁰ du tableau construit par Banniard. Pourtant, comme nous allons le voir dans ce contexte, la difficulté technique de cette procédure pour certains médiévistes, tout comme les difficultés qui sont attachées à sa compréhension, vont entraîner de fortes critiques quant à l'intérêt de l'analyse factorielle de correspondances en histoire.

2. Simplifier les opérations de synthèse historique.

En 1981, dans la rubrique « *Courrier des lecteurs* » du numéro 6 du *Médiéviste et l'ordinateur*, Bonnassie, l'auteur de la monographie sur la Catalogne sur laquelle Banniard construit l'une des lignes de son tableau, réagit à l'article de Guerreau. Bonnassie est alors un médiéviste important en France, reconnu notamment comme l'un

⁹²⁷ Guerreau, 10.

⁹²⁸ Guerreau, 10.

⁹²⁹ Guerreau, 10.

⁹³⁰ Guerreau, 10.

des principaux défenseurs de la thèse de la mutation féodale de l'an mil⁹³¹. Son doctorat d'État, soutenu en 1976 devant un jury présidé par Duby, propose l'un des modèles les plus radicaux du paradigme mutationniste⁹³².

Partant de l'étude de la Catalogne, Bonnassie développe l'idée d'une rupture soudaine du système économique et social entre les années 1020 et 1060, période caractérisée par la destruction de l'ordre public et marquée par des violences, des guerres privées, le développement des *castra* et la victoire finale des spécialistes de la guerre que sont les *milites*. À partir de 1060, ce système se cristallise et les structures héritées de la période précédente s'institutionnalisent à la faveur des rois, des princes ou des comtes, avec comme moyens de domination privilégiés le fief et les relations de vassalité. Au début des années 1980, Bonnassie soutient que ce modèle fulgurant de genèse et d'évolution du système féodal s'applique à toute l'ère hispano-occitane, c'est-à-dire à une région qui va du Rhône à la Galice⁹³³.

Les sujets abordés par Banniard et Guerreau sont ainsi au centre des préoccupations de Bonnassie. Aucune source ne permet cependant de saisir les raisons qui l'ont poussé à lire *Le Médiéviste et l'ordinateur*, rien ne laissant supposer dans ses articles publiés qu'il utilisait lui-même des moyens informatiques dans son travail de recherche. Dans sa réponse à Guerreau, sa curiosité pour la procédure d'analyse factorielle se manifeste en ces termes :

« *l'entreprise est hardie et même assez excitante* »⁹³⁴.

Il ajoute toutefois :

« *les résultats ne [sont] pas à la hauteur de la tentative* »⁹³⁵.

⁹³¹ Voir par exemple les déclarations en ce sens par l'un des spécialistes de cette question : Barthélemy, « La mutation féodale a-t-elle eu lieu ? »

⁹³² Bonnassie, « La Catalogne ».

⁹³³ Pour saisir l'importance de la thèse de Bonnassie et de son modèle catalan dans ce contexte, on pourra notamment se reporter au colloque déjà cité et présenté dans : Carpentier, « Structures féodales et féodalisme dans l'Occident méditerranéen (X^e-XIII^e s.) ».

⁹³⁴ Bonnassie et Guerreau, « Échos du numéro 5 », 17.

⁹³⁵ Bonnassie et Guerreau, 17.

La critique de que Bonnassie développe porte sur trois points : 1) le choix et le codage des informations, c'est-à-dire sur la formalisation de l'observation historique que suppose la procédure de Guerreau ; 2) le rôle de l'analyse factorielle des correspondances dans l'opération de synthèse de ces observations et ; 3) le recours à l'ordinateur comme « solution miracle » aux difficultés posées par la synthèse en histoire.

Dans un second article publié dans la rubrique « *Courrier des lecteurs* », publié dans ce même numéro de 1981, Guerreau répond précisément à son aîné sur chacun de ces trois points, dans le but de défendre sa méthode. En analysant les critiques formulées par Bonnassie à la lumière des réponses de Guerreau, en repérant les incompréhensions qui structurent cet échange, en devinant, aussi, les non-dits qui traversent ce débat, il nous sera possible de mieux comprendre quels sont les principaux désaccords épistémologiques entre ces deux historiens dans ce contexte et, partant, de saisir plus largement certains traits de la réception de l'analyse factorielle des correspondances en histoire médiévale.

2.1. Informations « vagues », codages « subjectifs ».

La première critique formulée par Bonnassie porte sur le choix des critères permettant de comparer les régions concernées par l'analyse. Si le texte de Bonnassie s'adresse à Guerreau, sa critique porte, de fait, sur les choix opérés initialement par Banniard. Selon Bonnassie, les critères retenus sont « *beaucoup trop vagues* » :

« *Que signifient par exemple des critères tels que « espace social étendu/espace social réduit », ou « maintien des gardes publiques/dominances des gardes privées ? »*⁹³⁶.

Mais tout autant que les critères, c'est le formalisme qu'a choisi Guerreau pour coder ces données qui pose problème à Bonnassie.

Ce deuxième volet de sa critique s'exprime en ces termes :

⁹³⁶ Bonnassie et Guerreau, 17.

« qui peut être capable d'affecter, sur de tels critères, un 1 ou un 0 (puisqu'il n'y a pas d'autres choix) ? »⁹³⁷.

En d'autres termes, comment peut-on rendre compte de la complexité des observations historiques dans le langage booléen ? Voilà, en substance, la question posée par Bonnassie. Dès lors, pour lui :

« le choix [d'attribuer un 1 ou un 0 pour chaque critère] est très aléatoire et il est en tout cas largement subjectif »⁹³⁸.

Cette première critique de Bonnassie s'adresse donc autant à Guerreau qu'à Banniard : les limites, imposées par la rigidité du formalisme mis en œuvre, conduisent ces historiens à abandonner la nuance que peut apporter une description plus traditionnelle.

Dans la suite de son texte, Bonnassie développe les raisons qui le pousse à affirmer que ce codage ne peut être que « très aléatoire » et « largement subjectif ». Il insiste en effet sur le fait qu'il n'existe pas, en histoire, d'informations brutes, mais seulement des observations résultant d'interprétations successives. Pour le cas qui nous intéresse, Bonnassie formule le problème en ces termes :

« Pour prendre l'exemple catalan, il s'agit même d'interprétations au quatrième degré : Guerreau a interprété Banniard, qui a interprété ma thèse, laquelle tentait d'interpréter des documents, lesquels interprétaient (ô combien !) la réalité. Au terme d'un tel cheminement, que peut-il bien subsister de la matière initiale, c'est-à-dire de l'histoire vécue ? Sûrement pas grand-chose. Ou plus exactement un mélange, à proportions égales, de vérités et d'erreurs »⁹³⁹.

Sur la base de ce modèle, le point clef concerne la déperdition d'informations à chaque répétition d'interprétation. En définitive, le choix d'attribuer la valeur 1 ou 0 à chacun des critères ne saurait ainsi procéder que d'une relation trop lointaine avec les sources.

⁹³⁷ Bonnassie et Guerreau, 17.

⁹³⁸ Bonnassie et Guerreau, 17.

⁹³⁹ Bonnassie et Guerreau, 18.

À la lumière de ces considérations, il nous faut toutefois formuler deux remarques, particulièrement importantes pour notre sujet. Tout d'abord, notons que cette critique ne vise pas spécifiquement les méthodes historiques assistées par ordinateurs, mais qu'elle s'applique plus largement à la majorité des tentatives de synthèses de monographies régionales⁹⁴⁰. Ensuite, cette citation nous permet de mieux saisir la conceptualisation de la finalité de l'observation historique défendue par Bonnassie. Pour lui, l'historien a pour rôle de poursuivre, à travers sa critique des documents (leurs « *interprétations* »), une « *réalité* », une « *histoire vécue* » par les acteurs, conception que ne partage pas tous ses contemporains⁹⁴¹.

Pour montrer l'accumulation des imprécisions qui résultent des interprétations successives, Bonnassie propose, dans son texte, de soumettre la Catalogne au même questionnaire, sur la base de sa propre expertise de spécialiste de cette région. Ses réponses pour chacun des 13 critères diffèrent fortement de celles de l'auteur du *Que sais-je ?* :

« *Pour la Catalogne carolingienne, il y a concordance pour 5 questions, divergence pour 3, opposition pour 5* »⁹⁴².

Par cette opération, Bonnassie illustre le problème que pose la démarche de Banniard et, partant, celle de Guerreau. Il ne remet cependant pas en cause la subjectivité de leur codage à l'aune d'une autre démarche, plus objective, mais leur oppose sa propre subjectivité, c'est-à-dire celle du monographe, plus proche des sources. De façon tout à fait paradoxale, Bonnassie range donc ici la rigidité du formalisme informatique du côté de la subjectivité, et y oppose une autre subjectivité, celle de l'historien spécialiste.

⁹⁴⁰ Sur ce point voir par exemple Colloque International sur la philosophie de L'Histoire. [1980. Ottawa], *La Philosophie de l'histoire et la pratique historique d'aujourd'hui*, éd. David Carr (Ottawa, Canada: Éditions de l'université d'Ottawa, 1982).

⁹⁴¹ La démonstration la plus flagrante de la perception de ces différences a été formulée, près de 15 ans plus tôt, par le philosophe Michel Foucault dans l'introduction de *L'archéologie du savoir*. On y reconnaîtra une allusion directe aux méthodes de ce qu'on appelle aujourd'hui l'histoire quantitative : « *Or, par une mutation qui ne date pas d'aujourd'hui, mais qui n'est sans doute pas encore achevée, l'histoire a changé sa position à l'égard du document : elle se donne pour tâche première, non point de l'interpréter, non point de déterminer s'il dit vrai et quelle est sa valeur expressive, mais de le travailler de l'intérieur et de l'élaborer : elle l'organise, le découpe, le distribue, l'ordonne, le répartit en niveaux, établit des séries, distingue ce qui est pertinent de ce qui ne l'est pas, repère des éléments, définit des unités, décrit des relations.* » in Michel Foucault, *L'archéologie du savoir* (Paris, France: Gallimard, 1969).

⁹⁴² Bonnassie et Guerreau, 18.

Pour répondre à cette critique, quoiqu'elle ne lui soit adressée que par ricochet, Guerreau propose, dans sa réponse à Bonnassie, de reprendre l'analyse factorielle avec ces nouvelles données corrigées. Il publie donc trois nouveaux graphiques d'analyse factorielle de correspondances, qui montrent que seul le point correspondant à la Catalogne bouge, mais que la forme du nuage de points ne change pas. La structure centre-périphérie dégagée par Guerreau est donc toujours visible.

Alors, « *pourquoi cette faible variation ?* »⁹⁴³ demande Guerreau. La réponse qu'il apporte nous permet de mieux saisir la fonction épistémologique qu'il attribue à l'analyse factorielle des correspondances dans sa méthode :

*« Tout simplement parce que, ni dans l'optique de Banniard ni dans celle de Bonnassie, la Catalogne carolingienne n'occupe une position extrême ou originale. L'analyse factorielle apparaît ici parfaitement dans un de ses rôles essentiels : ramener à de justes proportions des différences d'évaluation, dont l'importance réelle n'est pas toujours proportionnelle à l'intérêt que leur accordent les spécialistes »*⁹⁴⁴.

Sans nous appesantir pour le moment sur l'intérêt fondamental que présente cette citation pour la compréhension du rôle attribué à l'analyse factorielle dans l'épistémologie défendue par Guerreau, nous constaterons simplement le décalage entre sa réponse et la critique de Bonnassie analysée ci-dessus.

Guerreau confirme par ailleurs, plus loin dans l'article, qu'il refuse de s'engager sur ce terrain :

*« Mon propos n'était nullement de discuter le bien-fondé du tableau de M. Banniard, mais seulement d'en proposer une figuration graphique : il n'est que trop clair qu'une discussion sérieuse, même brève, de ce tableau nécessiterait au moins un gros article, sinon un livre entier. »*⁹⁴⁵.

⁹⁴³ Bonnassie et Guerreau, 18.

⁹⁴⁴ Bonnassie et Guerreau, 19.

⁹⁴⁵ Bonnassie et Guerreau, 19.

Le décalage qui apparaît entre ce premier argument de Bonnassie et la réponse de Guerreau va nous permettre de saisir en profondeur les désaccords épistémologiques entre les auteurs. Ce décalage concerne autant les opérations de synthèse historique que l'intérêt des analyses factorielles dans la réalisation de cette synthèse.

2.2. L'analyse factorielle de correspondances aplatit-elle la chronologie ?

Dans la suite de son courrier de lecteur de 1981, Bonnassie tente d'expliquer pourquoi ses réponses aux questions sur la Catalogne sont différentes de celles de Banniard.

« Quelle est la raison de telles divergences ? Elles résultent d'abord manifestement de différences dans les niveaux de synthèse (synthèse régionale [pour la Catalogne] en ce qui me concerne, continentale – européenne – en ce qui concerne Banniard-Guerreau) : plus la synthèse s'élargit, plus les données de bases doivent être simplifiées, donc maltraitées. Mais cette schématisation me semble surtout grave par l'écrasement qu'elle fait subir à la chronologie. Les X^e et XI^e siècles ne peuvent pas, à mon sens, être traités comme un « ensemble ». Ou alors, c'est nier carrément la révolution féodale des années – je simplifie à mon tour – 1020-1060. »⁹⁴⁶.

Deux critiques différentes, mais liées entre elles, apparaissent à la lecture de cette citation. La première regarde la question de la fonction épistémologique de la synthèse historique. Nous y reviendrons plus bas. La seconde concerne « l'écrasement de la chronologie » de l'histoire qu'implique la démarche de Guerreau.

Si l'on s'intéresse d'abord à cette seconde question, on se rend compte que s'opère ici un double malentendu. Il porte tout d'abord sur la période qu'a choisie d'analyser Guerreau, qui est, nous l'avons vu, la période carolingienne (VIII^e et IX^e siècles). L'analyse de Guerreau ne porte donc pas directement sur la « mutation féodale » de l'an mil (1020-1060). Le reproche de ne pas tenir compte de cette période, et par suite de

⁹⁴⁶ Bonnassie et Guerreau, 18.

pratiquer un « *écrasement de la chronologie* »⁹⁴⁷ n'est pas pertinent au vu des périodes traitées. Cette critique-ci paraît donc concerner ici plutôt le tableau de Banniard, et non la reprise de ce tableau par Guerreau.

Mais au-delà de ce malentendu, un autre transparait dans les critiques de Bonnassie. En effet, ce dernier laisse entendre qu'il comprend le résultat de l'analyse factorielle des correspondances comme celui d'une analyse diachronique, alors même que la méthode suppose, de par les données choisies, une comparaison synchronique entre les régions (voir 1.2). La réponse de Guerreau est en ce sens très directe, puisqu'il insiste sur ce point en ces termes :

*« il s'agit d'une analyse synchronique, qui ne vise donc ni à aplatir ni à gonfler la chronologie, mais à rechercher les positions relatives de ces 6 régions à un moment donné »*⁹⁴⁸.

Pour Guerreau, cette incompréhension vient d'une méconnaissance de la méthodologie de l'analyse factorielle des correspondances, dont il se propose de rappeler quelques principes.

À la fin de l'introduction de sa réponse, Guerreau affirme :

*« revenons aux problèmes techniques, sans la connaissance desquels la discussion générale risque de s'enliser »*⁹⁴⁹.

Suit alors une présentation, intitulée « *Rappels techniques* », qui définit le tableau des correspondances et propose une très rapide présentation des possibilités offertes par l'analyse factorielle pour exploiter ce type de tableau. Plus loin, Guerreau insiste encore, s'adressant certainement aux lecteurs autant qu'à Bonnassie lui-même :

« Il faut naturellement connaître la méthode pour pouvoir en apprécier et en interpréter correctement le produit graphique. On doit absolument éviter de se laisser abuser par le discours fallacieux de ceux qui vont répétant que

⁹⁴⁷ Bonnassie et Guerreau, 18.

⁹⁴⁸ Bonnassie et Guerreau, 20.

⁹⁴⁹ Bonnassie et Guerreau, 18.

*l'analyse factorielle serait comme une machine dont on pourrait se servir sans en connaître le fonctionnement interne : aucune technique intellectuelle, si rigoureuse soit-elle, n'est jamais réductible à une mécanique. »*⁹⁵⁰

Le message est clair : pour Guerreau, l'incompréhension entre les deux médiévistes naît avant tout d'une méconnaissance de cette « *technique intellectuelle* ». Les graphiques d'analyses factorielles des correspondances ne sont pas des photographies de la réalité, limpide pour tous les lecteurs, mais des produits techniques. Ils nécessitent, pour être intelligibles, une compréhension du « *fonctionnement interne* » de cette procédure, et ne peuvent donc être interprétés sans ouvrir la boîte noire que constitue l'analyse factorielle des correspondances.

Pourtant, si cette difficulté que rencontre Bonnassie est flagrante, et certainement liée à un déficit de connaissances en analyse des données, la réponse de Guerreau ne s'arrête pas là. La réponse qu'il va formuler au second point de la critique de Bonnassie va révéler que ce désaccord va bien au-delà de la compréhension technique de la méthode.

2.3. Niveaux de synthèse versus structure. L'histoire comme science sociale.

Au second élément de la critique de Bonnassie, qui incrimine la nécessité de simplifier les données de base pour permettre une synthèse à l'échelle continentale, Guerreau donne la réponse qui suit :

*« les développements de P. Bonnassie sur les "niveaux de synthèse" me paraissent singulièrement anachroniques. L'usage du vieux couple aristotélicien extension-compréhension ne correspond plus à l'état actuel des sciences sociales ; une synthèse, contrairement à ce qu'on lit dans l'ouvrage préhistorique de H. I. Marrou (qui est malheureusement encore conseillé aux étudiants) n'est pas une question de "niveaux" superposés, du canton à l'Europe, et de la monographie locale au manuel. »*⁹⁵¹

⁹⁵⁰ Bonnassie et Guerreau, 19.

⁹⁵¹ Bonnassie et Guerreau, 20.

Cette véhémence attaque de l'épistémologie défendue par Bonnassie passe ici par son association avec une autre conception de la synthèse en histoire, plus ancienne, celle de l'historien Henri-Irénée Marrou (1904-1977). Cette dernière peut notamment être résumée par son « analogie du muscle », proposée dans son fameux livre *De la connaissance historique* publié presque 30 ans plus tôt, en 1954⁹⁵², ici qualifié par Guerreau d' « *ouvrage préhistorique* »,.

Cette analogie se présente comme suit :

« On pourrait comparer la réalité historique à un muscle ; nous en étudions la structure sur une coupe faite à un niveau donné [...] : comme le muscle se révèle divisé en faisceaux subdivisés en fibres et fibrilles, le passé fait apparaître une structure plus ou moins parfaitement hiérarchisée de faits de civilisation, congères ou systèmes, et supersystèmes idéologiques (pour conserver la terminologie de Sorokin). Mais une étude plus complète exigera de l'anatomiste qu'il suive, de niveau en niveau, la continuité de chaque fibre ou faisceau, qu'il analyse leur nature et leurs rapports dans leurs modifications graduelles ; de même l'historien découvrira que chacun des éléments de la réalité historique, du fait de civilisation isolé ou élémentaire à la plus vaste synthèse, est inséré dans un développement continu au cours duquel il ne cesse de se transformer, comme ne cessent de se modifier les relations (de neutralité, d'antagonisme ou d'intégration) établies avec les éléments voisins. »⁹⁵³

C'est dans la droite ligne de cette description de la synthèse historique en niveaux successifs que s'inscrit, selon Guerreau, la critique adressée par Bonnassie.

Marrou propose dans son livre de 1954 une « *philosophie critique de l'histoire* »⁹⁵⁴, dans laquelle il se propose de mettre en avant la nature ambivalente du caractère scientifique de sa discipline. Il compare ainsi l'histoire aux sciences dites exactes, et notamment à la biologie, pour combattre

⁹⁵² Henri-Irénée Marrou, *De la connaissance historique* (Paris, France: Éditions du Seuil, 1954). Nous nous basons par la suite, pour les citations, sur la 6^{ème} réédition, qui date de 1975 : Henri-Irénée Marrou, *De la connaissance historique* (Paris, France: Éditions du Seuil, 1975).

⁹⁵³ Marrou, 176.

⁹⁵⁴ Nous empruntons ici le terme au titre de l'introduction du livre de Marrou.

« la conception positiviste de l'histoire qui proposait comme idéal à celle-ci de se hausser au niveau des sciences exactes et purement objectives »⁹⁵⁵.

Disséquant les logiques de la méthode historique, il s'intéresse notamment à l'usage des concepts dans le processus de synthèse historique. Il reprend sur ce point les conclusions du physicien Jacques Perrin (1870-1942), auteur en 1929 des *Éléments de la Physique*, mais aussi celles du philosophe Raymond Aron (1905-1983) sur les limites de l'objectivité historique, formulées dès 1938 dans *l'Introduction de la philosophie de l'histoire*⁹⁵⁶.

Marrou les formule en ces termes :

« Lorsque l'expérience scientifique s'étend à de nouveaux domaines, on découvre que les concepts utilisés jusque-là se trouvent dépendre très étroitement des conditions expérimentales dans lesquelles ils avaient été élaborés »⁹⁵⁷.

Appliquée à l'histoire, cette remarque rend notamment compte des limites de l'utilisation des concepts lorsque l'historien change d'échelle d'analyse. Pour Marrou, comme pour Bonnassie, un concept élaboré dans le cadre d'une analyse régionale ne peut donc être appliqué directement à une autre échelle d'analyse, sans que son sens en soit profondément transformé par cette opération intellectuelle.

Si l'on remet en perspective la conception de la synthèse développée par Marrou dans son analogie du muscle, avec cette limite inhérente à l'utilisation des concepts, on comprend mieux la suite de la réponse de Guerreau :

« Les concepts n'existent jamais à l'état individuel et isolé, mais seulement au sein d'une structure qui leur préexiste et d'où ils tirent tout leur sens. Qu'une

⁹⁵⁵ Antoine Guillaumont, « Henri Irénée Marrou. De la connaissance historique », *Revue de l'histoire des religions* 149, n° 1 (1956): 115-16. Les historiens positivistes visées ici sont, notamment, les historiens méthodistes Charles Seignobos (1854-1952) et Charles-Victor Langlois (1863-1929). Sur cette dénomination d'un pan des historiens français, dans le contexte qui nous intéresse, voir Charles-Olivier Carbonell, « L'histoire dite « positiviste » en France », *Romantisme* 8, n° 21 (1978): 173-85.

⁹⁵⁶ Raymond Aron, « Introduction à la philosophie de l'histoire : essai sur les limites de l'objectivité historique » (Paris, France: Gallimard, 1938).

⁹⁵⁷ Marrou, *De la connaissance historique*, 1975, 146. Pour la citation de Perrin se reporter à la page 190.

*structure conceptuelle puisse paraître différente au niveau du canton et au niveau du continent n'est qu'une illusion d'optique due à une perspective mal maîtrisée. La "synthèse", non seulement n'appauvrit pas, mais vise à enrichir la structure conceptuelle, qui doit intégrer la capacité de rendre compte simultanément des situations sociales synchroniques variées et en montrer les articulations »*⁹⁵⁸.

La critique de l'usage du « *vieux couple aristotélicien compréhension-extension* » formulée par Guerreau prend donc forme dans une affirmation d'un autre usage des concepts en histoire.

En effet, Bonnassie semblait fonder les divergences qui existent entre lui et le couple Banniard-Guerreau sur la différence d'échelle d'observation. Pour lui, le passage à une synthèse continentale nécessitait de « simplifier », et donc de « maltraiter » - pour reprendre ses termes - ou d'appauvrir les observations dégagées dans les monographies⁹⁵⁹. La difficulté de s'entendre sur les concepts pris en compte dans l'analyse (c'est-à-dire, dans notre cas, les critères), et donc sur les réponses au questionnaire initialement élaboré par Banniard, naissait donc de cette simplification, de cette transformation des observations historiques en données. En reprenant les termes choisis par Guerreau, la *compréhension* d'un concept, c'est-à-dire l'ensemble des attributs qu'on ne peut lui ôter sans le détruire, perdait de sa substance à mesure que cette compréhension était étendue à plus d'éléments, c'est-à-dire en l'espèce à davantage de régions (ce que Guerreau qualifie d'*extension*).

Par contraste avec cette épistémologie perçue chez Bonnassie par Guerreau, ce dernier recherche une utilisation structuraliste des concepts. Pour en préciser la teneur, nous nous appuyerons sur le chapitre V de son livre précédemment cité *Le féodalisme, un horizon théorique*, dans lequel il se propose justement de « *réfléchir de manière plus systématique sur les conditions de validité de l'emploi de tel ou tel concept ou système conceptuel* »⁹⁶⁰. Dans le début de ce chapitre, Guerreau s'appuie sur un livre d'épistémologie des sciences sociales, intitulé *Le métier de sociologue*, publié en 1968

⁹⁵⁸ Guerreau et Bonnassie, « Échos du numéro 5 », 20.

⁹⁵⁹ Rappelons les mots exacts de Bonnassie : « *plus la synthèse s'élargit, plus les données de bases doivent être simplifiées, donc maltraitées* ». Guerreau et Bonnassie, 18.

⁹⁶⁰ Guerreau, *Le Féodalisme*, 147.

par trois sociologues structuralistes français : Jean-Claude Chamboredon (1938-2020), Pierre Bourdieu (1930-2002) et Jean-Claude Passeron (1930-)⁹⁶¹.

Guerreau y présente notamment le problème que pose l'utilisation d'un concept isolé en ces termes, citant les trois sociologues :

« Le concept isolé (ou groupé par 2 ou 3) est impuissant : « les concepts les plus capables de déconcerter les notions communes ne détiennent pas à l'état isolé le pouvoir de résister systématiquement à la logique systématique de l'idéologie [sic] : à la rigueur analytique et formelle des concepts dits « opératoires » s'oppose la rigueur synthétique et réelle des concepts que l'on a appelés systémiques parce que leur utilisation suppose la référence permanente au système complet de leurs interrelations » ».

L'option épistémologique choisie par Guerreau, qui consiste à élaborer des concepts « systémiques » plutôt que des concepts « opératoires », s'inspire donc directement des travaux en sociologie. Elle repose sur un idéal épistémologique qui s'appuie, à la lumière de cette citation, sur deux principaux éléments. D'une part, l'idée que la finalité de la recherche en sciences sociales serait de « déconcerter les notions communes », objectif dont il serait intéressant, dans de futures recherches, d'étudier plus précisément le sens et la portée. D'autre part, on retrouve dans cette citation l'idée que certains types de concept, du fait de la nature de la rigueur qu'ils incarnent, ont plus de valeur, ou, en d'autres termes, sont plus à même de faire avancer la recherche vers la finalité précédemment évoquée.

Dans la suite de son chapitre, Guerreau développe ce raisonnement, justifiant pour le cas de l'histoire la valeur épistémologique qu'il défend ici avec les sociologues, à savoir, une rigueur « synthétique et réelle » plutôt qu'« analytique et formelle ». Selon lui, les couples, ou triades, de concepts utilisés par les historiens (« profane/sacré », « moderne/ancien », « jeunes-adultes-vieillards », etc..) n'opèrent que des divisions formelles et artificielles, qui ne saurait être considérées comme des conclusions pour un travail historique :

⁹⁶¹ Pierre Compilateur Bourdieu et al., éd., *Le métier de sociologue : préalables épistémologiques* (Paris, France, 1973).

« Dès lors qu'on a affaire à un tout documentaire et/ou social, le but ne consiste pas à opérer des divisions, mais à chercher une structure ou un système. La recherche empirique des divisions n'est jamais qu'un moyen : il n'y a que la logique aristotélicienne qui soit pure classification. La recherche des divisions (par opposition et/ou sur diverses échelles) est un procédé destiné à mettre en évidence les rapports et les liaisons pertinentes, qui structurent le tout que l'on étudie »⁹⁶².

Pour illustrer la limite des approches mettant en œuvre des concepts « opératoires », Guerreau prend justement, l'exemple de la « synthèse », en insistant en particulier sur le caractère relatif des oppositions entre les groupes sociaux :

« Considérer les groupes sociaux comme des objets réels est sans doute aujourd'hui la manière la plus commune de s'interdire l'accès à toute explication. Non que les études fondées sur une telle base soient inutiles : mais il est insensé d'attendre de leur accumulation une quelconque « synthèse » [...]. Par exemple, si l'on continue indéfiniment d'étudier les villes d'un côté, les campagnes de l'autre, il n'y a pas de raison pour qu'on cesse un jour de considérer cette opposition comme fondamentale »⁹⁶³.

Comme l'illustre cette dernière citation, les logiques de l'analyse de Guerreau sont donc exactement opposées à celles de Banniard, mais également à celles de Bonnassie. Pour Guerreau, comme pour les autres structuralistes, les éléments d'un ensemble n'ont pas de signification en eux-mêmes, puisque ces éléments ne permettent pas de dégager un système. Au contraire, ce sont les jeux de relations qui les lient et les opposent qui permettent de comprendre leurs significations, d'où la nécessité de construire des concepts dont l'« utilisation suppose la référence permanente au système complet », comme celui de centre-périphérie discuté plus haut.

⁹⁶² Guerreau, *Le Féodalisme*, 154.

⁹⁶³ Guerreau, 154.

Au sein de cette épistémologie, le rôle de l'analyse factorielle des correspondances prend tout son sens. Pour rappel, Guerreau le formulait en ces termes :

« ni dans l'optique de Banniard ni dans celle de Bonnassie, la Catalogne carolingienne n'occupe une position extrême ou originale. L'analyse factorielle apparaît ici parfaitement dans un de ses rôles essentiels : ramener à de justes proportions des différences d'évaluation, dont l'importance réelle n'est pas toujours proportionnelle à l'intérêt que leur accordent les spécialistes »⁹⁶⁴.

Dans le cadre de la synthèse des monographies régionales portant sur le développement de la féodalité en Europe du Sud, l'analyse factorielle des correspondances offre la possibilité au médiéviste de proposer une synthèse de type structuraliste, c'est-à-dire qui met en balance les observations historiques au sein d'un système qui prend en compte l'ensemble de la documentation. Au contraire du tableau de Banniard, qui n'était qu'une accumulation ordonnée d'observations, la synthèse que propose Guerreau repose sur une prise en compte de tous les éléments du système telle qu'il a été formalisé par Banniard.

Si l'on met en balance la critique formulée par Bonnassie avec l'épistémologie défendue par Guerreau, on comprend comment le recours à l'analyse factorielle permet d'éviter une synthèse analytique à la Marrou. En ramenant les observations à une évaluation empirique de leurs importances relatives, mais aussi en intégrant ces observations dans une construction théorique plus globale – ici l'existence d'une structure centre-périphérie - la procédure que propose Guerreau soulève de nouvelles questions théoriques, non plus sur la place de tel ou tel élément, mais sur les concepts qui se réfèrent à tout le système étudié.

Cette opposition épistémologique fondamentale entre les deux auteurs met en lumière les raisons de leurs différends, qui ne reposent donc ni directement sur un rejet des méthodes informatiques par Bonnassie, ni sur les limites de ce dernier dans la compréhension de l'analyse factorielle des correspondances, comme pouvait le laisser penser le début de notre analyse. Comme nous allons le voir, ces deux auteurs sont d'ailleurs en accord quant aux limites de l'utilisation des ordinateurs pour les historiens.

⁹⁶⁴ Bonnassie et Guerreau, « Échos du numéro 5 », 19.

2.4. « *Saint Ordinateur, priez pour nous* ».

Dans sa réponse à Guerreau, publiée, dans le même numéro du *Médiéviste et l'ordinateur*, Bonnassie affirme en conclusion :

*« l'ordinateur, quels que soient ses mérites – et ils sont grands – est bien incapable de mesurer l'ampleur de telles distorsions [il parle des distorsions d'interprétation dont nous avons déjà longuement parlé]. Nous ne pouvons lui demander de nous livrer « la clef d'une loi générale sur le développement de la féodalité ». Il peut certes nous aider, si nous le programmons bien (ou plutôt le moins mal possible, eu égard à la déficience de nos matériaux), à formuler une telle loi. Mais celle-ci ne sortira jamais que de notre tête. »*⁹⁶⁵

Cette critique, qui porte sur un certain type d'usage des ordinateurs, n'est pas directement adressée à Guerreau. Elle s'adresse plutôt, sous la forme d'un avertissement déjà exprimé ailleurs à de nombreuses reprises⁹⁶⁶, et notamment aux lecteurs du bulletin de liaison. Elle rejoint en fait l'affirmation de Guerreau, citée plus haut, qui affirme que l'analyse factorielle de correspondances est une « *technique intellectuelle* », et non « *une mécanique* »⁹⁶⁷.

Ces considérations sur le fait que l'informatique est bien plus qu'un ensemble de techniques au service de l'historien, ou, en d'autres termes, qu'une nouvelle science auxiliaire, font écho à d'autres réflexions d'historiens ayant utilisé des ordinateurs dans les années 1970 et 1980. Dans une table ronde organisée en 1980 par l'Institut d'Histoire Moderne et Contemporaine (IHMC) et l'Institut d'Histoire du Temps Présent (IHTP) et intitulée « Histoire et informatique », déjà cité dans le chapitre précédent, le contemporainiste Prost affirmait par exemple :

« L'informatique n'est pas une technique, c'est une méthodologie. Et je pense que les déboires rencontrés par les historiens qui ont utilisé l'informatique tiennent à

⁹⁶⁵ Guerreau et Bonnassie, 17.

⁹⁶⁶ À l'époque, et jusqu'à aujourd'hui, la formule consacrée pour mentionner cet avertissement est le suivant : *garbage in, garbage out* ».

⁹⁶⁷ Guerreau et Bonnassie, 20.

ce qu'ils l'ont employée comme un outil, alors qu'elle n'a pas de sens en dehors d'une démarche. Il y a des gens qui possèdent déjà une démarche très voisine de ce que l'on peut pratiquer avec l'informatique : et pour ces historiens, dont je dirais qu'ils montrent une démarche en attente d'informatique, à mon avis, leur formation ne pose pas de gros problèmes. Et puis il y a ceux dont la démarche est étrangère à l'enchaînement et à la méthodologie de l'informatique ; et ceux-là, ce n'est pas en leur apprenant un langage de programmation qu'on les amènera à se servir d'un ordinateur. Pour moi, c'est vraiment fondamental, l'informatique est une méthodologie, ce n'est pas seulement une technique. »⁹⁶⁸.

Il semble, dans notre cas, que Guerreau et Bonnassie partagent cette représentation de l'informatique. La problématique du recours aux instruments de calcul que sont les ordinateurs ne constitue donc pas, dans le cas qui nous intéresse ici, l'une des causes de cette controverse entre Guerreau et Bonnassie.

En dépit de cet accord sur le fond, Guerreau est toutefois beaucoup plus radical que son aîné sur ce point⁹⁶⁹. Selon lui, le danger ne porte pas sur une potentielle mauvaise utilisation des ordinateurs, mais plutôt sur les conditions d'utilisation dans lesquelles ses contemporains s'emparent de ces instruments. Dans sa réponse à Bonnassie, il insiste en particulier sur l'absence de formation solide en statistiques de ses collègues, sans laquelle il est, selon lui, impossible d'utiliser intelligemment un ordinateur ou un outil de calcul de ce type. Pour Guerreau, l'ordinateur n'est en effet qu'une calculatrice aux capacités de mémoire et de calculs faramineuses, mais il ne doit pas empiéter sur l'autonomie des historiens dans leurs recherches.

Plutôt que d'avoir recours à un centre de calcul et à un ou plusieurs spécialistes de l'informatique, comme les collectifs que nous avons étudiés dans les chapitres 1 et 2, Guerreau fait donc le choix d'utiliser une calculatrice programmable de poche pour mener ses analyses. Dans sa réponse à Bonnassie, comme dans la majorité des articles dans

⁹⁶⁸ Antoine Prost, « La collecte et l'analyse des données », in *Table Ronde Histoire et Informatique* (Paris, 1981), 9.

⁹⁶⁹ Cet accord désaccordé est exprimé en ces termes par Guerreau : « *la critique de P. Bonnassie sur ce point ne s'adresse pas à moi : la mienne est plus radicale que la sienne* », in Guerreau et Bonnassie, « Écho du numéro 5 », 19.

lesquels il présente des procédures de calculs, il indique le modèle de son instrument de calcul :

« une calculatrice de poche (la mienne pèse 245 g. et coûte 480 F) suffit à résoudre 9/10 des difficultés rencontrées à présent par les médiévistes sur le chemin du développement des analyses numériques. »⁹⁷⁰.

Au-delà des considérations sur l'équipement, c'est donc avant tout la méthode choisie par le médiéviste qui conditionne l'intérêt des outils de calculs électroniques pour ses recherches.

La spécificité du modèle de travail que défend Guerreau, qui se caractérise par un choix d'outil de calcul basique et une indépendance technique et scientifique du point de vue des calculs, se doit d'être commentée. Un élément en particulier, directement lié aux problématiques de fond de ce débat, alimente en fait cette décision. Il concerne l'organisation collective de la recherche en France dans les années 1970, qui est alors dominée, depuis Paris, par le moderniste Braudel, professeur au Collège de France, et son collègue Camille-Ernest Labrousse (1895-1988), professeur à la Sorbonne.

Ces deux chercheurs, convaincus des vertus des méthodes quantitatives et de l'exploitation des sources statistiques, dirigent, entre 1960 et 1980, de nombreuses thèses, qui prennent la forme de monographies locales ou régionales. Elles ont le plus souvent pour objet d'exploiter, voire d'« épuiser » les données statistiques dont regorgent les archives départementales pour la période moderne, et d'établir, par exemple, des séries d'évolution des prix à l'échelle locale⁹⁷¹. Dans la continuité de ces travaux, entre 1977 et 1980, Braudel et Labrousse dirigent la publication, en quatre tomes, d'une *Histoire économique et sociale de la France (1450-1950)*, qui se présente comme la synthèse de résultats de ce programme de recherche qu'ils ont contribué à édifier⁹⁷².

⁹⁷⁰ Guerreau et Bonnassie, 20.

⁹⁷¹ Voir Jean-Yves Grenier et Bernard Lepetit, « L'expérience historique. À propos de C.-E. Labrousse », *Annales* 44, n° 6 (1989): 1337-60 ; Michel Vovelle, « La mémoire d'Ernest Labrousse », *Annales historiques de la Révolution française* 276, n° 1 (1989): 99-107. Parmi les étudiants qui poursuivent des thèses sous la direction officielle ou non-officielle de Labrousse on retrouve notamment Jean Bouvier (1920-1987), Pierre Vilar (1906-2003), Emmanuel Le Roy Ladurie (1929-), Adeline Daumard (1924-2003), Maurice Agulhon (1926-2014) ou Jean-Claude Perrot (1928-).

⁹⁷² Fernand Braudel et Ernest Labrousse, éd., *Histoire économique et sociale de la France* (Paris, France: Quadriga : Presses universitaires de France, c1977,1993, 1977). Nous précisons par ailleurs ici que nous

En histoire médiévale, ce modèle est moins développé, pour des raisons qu'il faudrait étudier. Cependant, cette organisation de la recherche se reproduit en partie, pour certains sujets, comme le montre l'exemple des six monographies choisies par Banniard pour construire son tableau. Et c'est justement ce modèle, celui de la multiplication des monographies régionales, dont la synthèse serait opérée par quelques pontes de la discipline, que Guerreau et d'autres médiévistes de sa génération rejettent en affirmant leur indépendance⁹⁷³.

Cet aspect du modèle de la recherche défendu par Guerreau se corrèle directement avec l'épistémologie qu'il met en oeuvre, dont nous avons discuté au 2.3. En effet, là où le modèle de Braudel et Labrousse décrit plus haut, allait de pair avec une synthèse par niveaux, telles que le mentionnait Bonnassie, l'organisation sociale de la recherche défendue par Guerreau se corrèle aisément avec une approche théorique et structuraliste. L'option technique qu'il choisit s'articule donc à la fois avec, d'une part, l'épistémologie qu'il revendique et, d'autre part, avec l'organisation sociale de la recherche qu'il défend.

L'incompatibilité des « *démarches* » de ces deux historiens, que l'on a définie dans ce chapitre sur la base de l'épistémologie défendue par les acteurs (objet visé par les études historiques, conception de l'observation en histoire, rôle et méthode de la synthèse), mais également des activités liées à celle-ci (mise en oeuvre de procédures mathématiques, utilisation d'instrument de calcul, production de visualisation mais également positionnement par rapport à une organisation politique de la recherche), sous-tend les échanges et les oppositions entre ces deux médiévistes. Elle révèle, dans notre contexte, l'existence d'une profonde rupture entre deux manières de faire de l'histoire, dont on peut affirmer qu'elle apparaît en relation étroite avec l'utilisation d'instruments de calcul électroniques programmable par l'un des deux acteurs.

préparons un article sur ce sujet, qui vise à examiner la division sociale des opérations historiques en France dans les années 1970-1980.

⁹⁷³ Pour découvrir certains éléments de contexte de cette opposition voir Jacques Rougerie, « Faut-il départementaliser l'histoire de France? », *Annales* 21, n° 1 (1966): 178-93. La position d'historien comme Genet ou Guerreau, qui rejettent alors l'histoire dite départementaliste, et s'engagent dans l'étude de phénomène transversaux (respectivement la genèse de l'État moderne et le féodalisme) nous a été révélée dans des entretiens que nous avons menés auprès d'eux entre 2017 et 2019.

Conclusion.

La controverse étudiée dans ce chapitre nous permet, dans la continuité du chapitre précédent, d'avancer un certain nombre de conclusions qui portent sur les modalités des échanges au sein de la communauté discursive que nous avons définie dans les chapitres 3 et 4. À la différence du chapitre précédent, nous ne nous sommes pas concentrés sur le *genre* et la fonction des textes de la rubrique « *Article de fond* » du bulletin de liaison *Le Médiéviste et l'ordinateur*, mais sur les échanges qui prennent forme du fait de l'existence d'une autre rubrique, celle du « *Courrier des lecteurs* ».

Le débat entre deux médiévistes, Guerreau et Bonnassie, nous a permis d'examiner certaines oppositions qui pouvaient exister à l'encontre de l'utilisation d'une des procédures d'analyses multivariées les plus mobilisées par les médiévistes à partir des années 1970 : l'analyse factorielle des correspondances. L'intérêt de ce cas pour notre sujet reposait sur la nécessité d'utiliser des outils de calcul électroniques programmables pour mettre en œuvre cette procédure. Partant, et comme pour d'autres cas que nous avons présentés dans les chapitres 1, 2 et 5, cet exemple offrait la possibilité d'examiner le lien que la mise en œuvre des analyses factorielles des correspondances entretenait avec les pratiques de recherche et l'épistémologie défendue par les acteurs.

Sur la base de notre analyse, le principal élément que nous avons établi en ce sens est le fait qu'un médiéviste, Guerreau, formé aux méthodes de l'analyse des données et ayant développé depuis la fin des années 1970 une pratique de l'histoire fondée sur l'utilisation des moyens électroniques de calculs, repère dans les travaux d'un de ses collègues une forme de visualisation dont il considère qu'elle peut être exploitée tout autrement. Dans notre exemple, le tableau synoptique de Banniard est en effet perçu par Guerreau, du fait de la structure des informations que ce dernier y repère, comme un tableau d'incidence qui doit lui permettre de mettre en œuvre des analyses factorielles des correspondances. La manifestation de cette différence de perception d'une même visualisation peut être apparentée à ce que Thomas Samuel Kuhn (1922-1996) qualifiait

de *gestalt switch*⁹⁷⁴. Les questions qu'elle soulève par rapport à notre sujet sont dès lors d'au moins deux types.

Elles concernent tout d'abord les modalités selon lesquelles l'utilisation des moyens électroniques de calcul programmables explique cette différence de perception entre les deux médiévistes. Plusieurs hypothèses peuvent être avancées ici mais une seule nous apparaît véritablement pertinente à examiner : l'utilisation des méthodes d'analyse des données, et l'utilisation des moyens électroniques de calcul programmables que cela implique, procéderaient d'un changement profond des compétences des médiévistes et, plus largement, d'un changement fondamental pour la discipline, que l'on pourrait qualifier de changement de paradigme ou de changement de régime d'objectivité en fonction des affinités.

Pour préciser cette hypothèse, il faut toutefois insister sur le fait que contrairement à d'autres, nous l'avons vu, Guerreau met en œuvre des analyses factorielles des correspondances dans un environnement de travail singulier dans notre contexte. À la différence d'autres médiévistes, comme par exemple Hélène Millet (1944-), Guerreau n'engage pas de collaboration avec des statisticiens ou des spécialistes de l'informatique, à qui il déléguerait l'élaboration et la mise en œuvre sa méthodologie statistique. Il ne travaille pas non plus dans un centre de calcul, mais réalise ses calculs chez lui, seul, à l'aide d'une calculatrice programmable de poche. Il revendique ainsi une indépendance méthodologique qui le conduit à développer une série de compétences en statistiques qui sont, selon nous, au cœur de la modification de perception que nous observons ici.

Un autre élément abonde dans ce sens : contrairement aux collectifs que nous avons étudiés dans les deux premiers chapitres, Guerreau n'utilise pas son instrument de calcul pour analyser de très grand corpus et produire de nouveaux instruments de travail. Il utilise les moyens électroniques de calcul pour mettre en œuvre de nouvelles formes d'analyse, et montre, notamment avec l'exemple d'application des analyses factorielles que nous avons étudié dans ce chapitre, que l'intérêt de cette nouvelle instrumentation de la recherche ne repose pas uniquement sur ses apports « quantitatifs », mais plus essentiellement sur ses apports « qualitatifs ».

⁹⁷⁴ Voir Thomas Samuel Kuhn, *The structure of scientific revolutions* (Chicago, Etats-Unis d'Amérique: the University of Chicago Press, 1962). Pour une approche du lien plus large entre *gestalt switch* et structuralisme voir Jean Piaget, *Le structuralisme* (Paris, France: Presses universitaires de France, 1968).

Cette conception de la fonction des moyens électroniques de calcul programmables pour la recherche en histoire médiévale s'explique, nous l'avons dit, par l'existence d'un autre mouvement de fond au sein de la discipline dans les années 1970. Il s'agit du développement d'approche structuralistes, inspirées plus particulièrement des travaux de Bourdieu. Dans le chapitre 2, nous avons notamment analysé un autre avatar des effets de ce modèle méthodologique en histoire médiévale, en examinant l'usage que le médiéviste Jean-Philippe Genet (1944-) proposait du concept de *champ*. Pour rappel, Bourdieu comme Genet, fondaient leurs approches méthodologiques et leurs constructions théoriques sur ce type de procédures mathématiques⁹⁷⁵, ce que l'on retrouve dans l'utilisation des analyses factorielles des correspondances que propose ici Guerreau.

La différence de perception entre Guerreau et Banniard s'inscrit donc ici également en lien avec une évolution plus large des relations qu'entretiennent les historiens avec d'autres sciences sociales. L'utilisation des moyens électroniques ne se présente ainsi que comme la dimension instrumentale d'une évolution plus profonde, à la croisée des sciences sociales et des mathématiques, et dont ils participent en offrant aux chercheurs les moyens de mettre en œuvre de nouvelles procédures de calculs. Plus qu'une conséquence directe de l'utilisation des ordinateurs, comme le gain de temps ou la fiabilité des calculs, que nous avons observées dans les chapitres 1 et 2, ce changement de perception apparaît ainsi comme une conséquence indirecte de ce type de pratique.

Cependant, il ne faudrait pas ici généraliser. En effet, nous avons montré tout au long de ce chapitre que Guerreau défendait une position originale dans le champ des études médiévales, mais également dans le domaine plus spécifique des utilisateurs des moyens électroniques de calcul en histoire médiévale. On ne peut d'ailleurs affirmer si l'épistémologie qu'il défend dans ses échanges avec Bonnassie est partagée par d'autres acteurs, quoique de fortes ressemblances avec le travail de Genet, étudié dans le chapitre 2, sont à souligner. Nous reviendrons sur cette similitude dans la conclusion générale, en essayant d'en tirer toutes les conclusions quant aux conséquences historiographiques de l'utilisation des ordinateurs en histoire médiévale.

⁹⁷⁵ Voir par exemple Julien Duval, « L'analyse des correspondances et la construction des champs », *Actes de la recherche en sciences sociales* N° 200, n° 5 (2013): 110-23 ou encore Paul-André Rosental, « Introduction. Outil ou fétiche : la laïcisation de l'analyse factorielle dans les sciences sociales », *Histoire & Mesure* 12, n° 3 (1997): 185-96.

Les questions que soulèvent cette différence de perception entre Guerreau et Banniard regardent encore un autre ensemble de questions, qui portent cette fois sur les conséquences de l'utilisation des moyens électroniques de calcul programmable sur l'évolution de l'épistémologie de l'histoire dans les années 1970 et 1980. Nous l'avons dit, le fait que Guerreau perçoive le tableau synoptique de Banniard comme un tableau d'incidence le conduit à proposer « *une procédure de visualisation synthétique* » qui vise à transformer cette figure. De l'examen cette procédure nous avons dégagé plusieurs éléments.

Tout d'abord, un élément apparaît fondamental : il s'agit de l'évolution de ce que nous avons appelé la *fonction épistémologique* des visualisations produites par les acteurs sur la base de cette procédure. Guerreau propose en effet, de façon similaire à ce que nous avons vu dans les chapitres 1 et 2, une utilisation heuristique du tableau de Banniard, mais également des graphiques qu'il produit au fil de sa démarche. Le tableau initial, auquel, nous l'avons montré, Banniard assignait la fonction de résumer au lecteur les conclusions d'une analyse, devient pour Guerreau un élément permettant de mettre en œuvre une procédure de calcul. Cette transformation de la fonction de ce tableau semble essentiellement attachée à la manière de percevoir les images développées par Guerreau, et dont les raisons ont été examinées plus haut.

Ensuite, un autre type de conséquence peut encore être dégagée de ce premier élément que nous avons repéré. Il s'agit de la mutation qui l'accompagne quant à des aspects épistémologiques plus fondamentaux de l'histoire : la pratique de l'observation historique, la définition de l'objet de l'histoire, l'utilisation des concepts, etc. Sur ce point il semble cependant extrêmement difficile, voire impossible, d'affirmer si nous avons affaire à une autres des raisons que nous avons examinées ci-dessus, ou à une conséquence du *gestalt switch* que nous avons identifié. Est-ce le recyclage de méthodes élaborées sur la base de l'utilisation des analyses factorielles dans d'autres sciences sociales (sociologie, linguistique) qui a conduit Guerreau à développer l'épistémologie qu'il défend ici dans sa réponse à Bonnassie ? Ou est-ce que la construction de cette épistémologie est antérieure à cet intérêt pour ce type de méthodes ?

Nous retombons ici sur l'une des questions que soulevait Alexandre Koyré (1892-1964) en 1956⁹⁷⁶ dans sa critique de l'empirisme anglais, et plus particulièrement du travail de l'historien des sciences australien Alistair Cameron Crombie (1912-1996). Dans ce livre, il affirmait, en se fondant sur ses propres travaux sur Galilée, et à propos de la naissance de la science « *moderne* », que la théorie primait sur les aspects matériels de la recherche, ou, en d'autres termes, sur ce qu'il appelait la *praxis* :

« *Quant moi, je ne crois pas que la naissance et le développement de la science moderne puissent s'expliquer par le fait que l'esprit se soit détourné de la théorie du profit de la praxis [...]. En fait invention de la charrue, du harnais, de la bielle manivelle et du gouvernail arrière n'a rien voir avec le développement scientifique.* »⁹⁷⁷

Si le contexte qui nous intéresse ici est fondamentalement différent, tout particulièrement parce que l'invention des ordinateurs, tels que nous les avons définis, a bien à voir avec le développement scientifique, la question qui se pose dans notre cas nous semble comparable : existe-t-il, dans l'apparition de nouvelles formes de pratique de l'histoire médiévale dans les années 1970 en France, un primat de la théorie sur la pratique ? Et, donc, quelles est le poids de l'un des instrument clef de cette *praxis* - les outils électroniques de calcul - dans les évolutions que nous observons ?

Dans le contexte qui nous intéresse ici, cette question apparait d'autant plus prégnante que les méthodes que développent les acteurs, nous l'avons montré tout au long de cette thèse, s'inscrivent dans des contextes matériels très contraignants : des instruments de calculs coûteux, dont il faut définir et justifier l'intérêt en amont, l'acquisition de connaissances mathématiques et techniques extérieures à la discipline (programmation, analyse de données), la production de nouvelles formes de visualisation (imprimés par les ordinateurs ou tracés « à la main »). Dès lors, comment affirmer, pour le cas de Guerreau, dans quelle mesure l'utilisation des ordinateurs a contribué aux évolutions décisives que présente l'épistémologie qu'il défend contre celle de Marrou.

⁹⁷⁶ Alexandre Koyré, « Les origines de la science moderne, une interprétation nouvelle », *Diogène*, n° 16 (1956): 14. Ce texte a été republié dix ans plus tard dans Alexandre Koyré, *Études d'histoire de la pensée scientifique* (Paris, France: Presses universitaires de France, 1966).

⁹⁷⁷ Koyré, *Études d'histoire de la pensée scientifique*, 74.

Cette piste de recherche, que nous proposons ici d'ouvrir, devra selon nous être poursuivie par une étude serrée des modalités de la formation des médiévistes aux méthodes et aux techniques de l'utilisation des outils de calculs. Ce travail passe notamment par l'étude des manuels de programmation et d'utilisation des moyens électroniques de calculs destinés aux historiens ou aux chercheurs en sciences humaines et sociales, dont nous avons réunis un corpus au cours de l'année 2020 dans le but d'entamer cette recherche⁹⁷⁸. Pour le cas précis de notre sujet, il apparaît encore que poursuivre l'étude de l'utilisation des analyses factorielles, que nous avons pu présenter au fil de ce chapitre comme un vecteur de circulation aussi bien de théories que de pratiques, serait un autre moyen d'aller plus loin dans l'examen de cette question centrale dans notre contexte.

Pour le cas plus particulier des pratiques des analyses factorielles en sciences humaines et sciences sociales, il semble toutefois important de noter que les objectifs attachés à leurs utilisations évoluent au courant des années 1990, en se détournant parfois des usages structuralistes (chapitre 6) ou classificatoires (chapitre 2) que nous avons observés au fil de cette thèse⁹⁷⁹. Leurs utilisations est tout de fois remise en cause en histoire dans le contexte français à cette même période, essentiellement en raison des dimensions théoriques que ces procédures charrient avec elle. Les fortes critiques adressées contre le structuralisme, combinées au développement d'autres approches méthodologiques comme celle de la *microstoria*, conduisent les acteurs à s'emparer de nouveaux outils mathématiques et, notamment, de la théorie des graphes dès la fin des années 1990⁹⁸⁰. Là encore, il faudrait étudier non seulement comment et pourquoi ces différentes procédures mathématiques sont attachées à différents programmes historiographiques, mais également examiner comment ces représentations circulent entre les acteurs. Voilà encore une nouvelle piste pour la suite de nos recherches.

⁹⁷⁸ Voir notamment le livre : Edward Shorter, *Historian and the Computer: A Practical Guide* (New York: W W Norton & Co Inc, 1975), et certaines des références que Shorter cite dans la partie sur l'initiation à la programmation : Daniel D. McCracken, *A Simplified Guide to FORTRAN Programming* (New York: Wiley, 1974); Donald J. Veldman, *Fortran Programming for the Behavioral Sciences* (New York: Holt, Rinehart and Winston, 1967).

⁹⁷⁹ Voir le numéro 12 de la revue *Histoire & Mesure*, publié en 1997, et en particulier l'introduction de ce numéro : Rosental, « Introduction. Outil ou fétiche ».

⁹⁸⁰ Voir Claire Lemercier, « Analyse de réseaux et histoire », *Revue d'histoire moderne contemporaine* n°52-2, n° 2 (2005): 88-112.

Conclusion générale.

Quel fut l'impact de l'utilisation des ordinateurs sur les pratiques de recherche des médiévistes entre 1966 et 1990 ? Nous avons pris cette question comme problématique générale de notre thèse et nous sommes ici en mesure d'y apporter quelques ingrédients de réflexion. Au fil des études de cas réalisées dans ces six chapitres, nous avons réuni de nombreux éléments qui nous permettent de dégager, à différentes échelles, certaines des mutations clefs causées par l'utilisation de calculateurs électroniques programmables en histoire médiévale.

Revenons à présent, à titre de conclusion de ce travail, sur deux des transformations essentielles qui portent respectivement sur les *organisations collectives* de la recherche et les *pratiques des sources* développées par les médiévistes. Nous insisterons dans un troisième temps sur les conséquences historiographiques qui résultent de la combinaison de ces deux transformations essentielles.

Organisations collectives de la recherche.

Comme nous l'avons montré, l'une des évolutions centrales que nous observons dans ce contexte tient aux modifications profondes que les usages des ordinateurs induisent sur les formes d'organisation collective du travail.

À l'échelle de ce que nous avons qualifié de collectifs de travail, nous avons identifié deux conséquences cruciales directement liées à cette nouvelle instrumentation de la recherche.

En premier lieu, on constate que le recours aux ordinateurs va de pair, pour les médiévistes, avec la multiplication des relations extra-disciplinaires. Les deux cas du *catasto* et d'ALINE, que nous avons examinés et comparés dans les chapitres 1 et 2, attestent de manifestations différentes de ce même phénomène. Christiane Klapisch-Zuber (1936-) et David Herlihy (1930-1991) travaillent avec des experts démographes et cartographes. Les médiévistes de l'Équipe de Recherche Autonome 713 (ERA 713) s'inspirent quant à eux des méthodes de la linguistique ou de la sociologie.

La différence entre ces deux projets porte également sur les modalités de l'inscription des spécialistes d'autres disciplines dans ces collectifs. Nous avons vu que l'équipe constituée pour étudier le *catasto* intégrait des démographes (Jacques Dupâquier (1922-2010), Louis Henry (1911-1991)) en tant que consultants. Par contraste, les médiévistes de l'ERA 713 travaillent sans spécialiste de la linguistique, et c'est Jean-Philippe Genet (1944-) qui a pour fonction de garantir, au sein de l'équipe, la pertinence des analyses statistiques envisagées.

La comparaison entre ces deux études de cas révèle également plusieurs points communs. Ces deux collectifs font tous les deux, du fait des modèles extra-disciplinaires qu'ils se sont choisis, appel à des modes d'analyse extérieurs au champ de l'histoire (démographie, sociologie, linguistique). Par ailleurs, de façon tout à fait nouvelle, des experts de l'informatique et/ou des statistiques, le plus souvent formés en mathématiques (François Hucher, Jacques Mondelli ou Michel Demonet), intègrent ces collectifs pour contribuer à la programmation et la conduite des opérations sur les ordinateurs.

Ces influences extra-disciplinaires ne sont pas sans avoir de nombreuses implications sur les cultures locales que ces collectifs élaborent. Avant de revenir sur cette dimension, insistons toutefois sur le fait que si les acteurs intègrent des préoccupations nouvelles à leurs approches, ils ne renoncent pas pour autant aux spécificités méthodologiques propre aux métiers de l'historien. On peut prendre pour exemple le choix d'enregistrer les âges des contribuables florentins en accord avec les données du *catasto*, et ce en dépit de la suggestion des démographes de corriger les âges erronés pour obtenir des résultats en accord avec la réalité démographique (chapitre 1). Ce dernier montre comment les pratiques de ces collectifs s'inventent en recyclant des méthodes développées par d'autres collectifs de travail, tout en les adaptant aux spécificités de la discipline historique, de leurs problématiques scientifiques et de leurs sources.

En un second lieu, il faut insister sur le fait que vouloir utiliser des ordinateurs pour poursuivre des recherches suppose, *a minima* pour les années 1960 et 1970 - et, dans certains cas, également dans les années 1980, nous y reviendrons - de s'inscrire dans des structures de recherche qui permettent d'accéder à ces nouveaux outils de calcul. Nous l'avons vu, la majorité des collectifs qui utilisent des ordinateurs dans les années 1970 ont recours aux services de centres de calcul, qui font alors figures de lieux incontournables. En effet, ces infrastructures, et en particulier les plus grands centres des

universités et du CNRS, assurent un accès à des ordinateurs puissants, comme les IBM 360 du CIRCE, et fonctionnels, de par la présence d'opérateurs en charge de l'utilisation de ces nouvelles technologies.

Toutefois, ces environnements techniques imposent une grande rigidité dans l'organisation matérielle des opérations de calcul. Tout d'abord, l'accès à ces grands centres de calcul est subordonné à l'inscription dans certains types de cadres administratifs, comme la Recherche Coopérative sur Programme (chapitre 1) ou les Équipes de Recherche Autonomes (chapitre 2). Ensuite, une fois ce type d'organisation mis en place, les heures de calculs sont payantes et accordées au compte-goutte par le CNRS. Enfin, les ordinateurs ont des capacités de mémoire et de calcul limitées par rapport à ce que l'on connaît aujourd'hui, et ce, en dépit des rapides évolutions technologiques pendant la période qui nous intéresse. Les calculs sont de fait très lents, et il faut souvent attendre plusieurs heures pour voir apparaître un résultat sur les imprimantes.

Dans un contexte de ce type, l'organisation sociale des calculs repose sur un schéma complexe : il faut non seulement avoir les budgets pour embaucher des spécialistes de l'informatique, mais également recevoir l'aval administratif des institutions compétentes pour accéder aux centres de calculs. Les disparités entre les collectifs de travail sont ainsi flagrantes, et conditionnent le recyclage des méthodes d'un collectif à l'autre. L'exemple de ALINE étudié dans le chapitre 2 illustre, par exemple, comment le réemploi des méthodes lexicométriques par les médiévistes de l'ERA 713 se fait au prisme des moyens techniques, financiers et humains à la disposition de ces acteurs.

Un autre élément éclaire la rigidité de l'organisation des calculs à cette période. En effet, la mise en œuvre de traitements informatiques par les médiévistes repose, jusqu'au début des années 1980, sur une interface chercheur-machine privilégiée : les cartes perforées IBM à 80 colonnes. Si nous revenons plus loin sur la diversité des usages de ces technologies intellectuelles, et sur leurs implications sur les pratiques des sources, il convient d'insister dès à présent sur les règles de travail collectif que les acteurs définissent pour organiser l'enregistrement des données sur ce type de support.

Pour le cas de l'enquête du *catasto*, nous avons par exemple montré que les acteurs mettent en place des règles très strictes pour le dépouillement et l'enregistrement des données, basées sur un type de *documents de travail* original : les bordereaux de perforation. L'établissement de ce document, ainsi que les méthodes qui président à l'inscription des données sur celui-ci, constituent les clefs de voûte de cette culture locale. À ces méthodes sont par ailleurs attachées de nombreuses procédures de correction des données qui structurent elles-aussi les pratiques collectives, ce qui étaye l'hypothèse selon laquelle organisation matérielle et organisation sociale sont intimement liées.

Au fil de nos analyses, nous avons montré que de nombreux historiens s'opposent à ces nouvelles pratiques en raison de l'obligation de saisir les données sous une certaine forme. Dans le chapitre 1, nous avons vu que pour le médiéviste italien Elio Conti (1925-1986), les méthodes mettant en œuvre des ordinateurs empêchent d'émettre des doutes ou des commentaires, ce qui le conduit à les rejeter. L'opposition de Conti est fondée, dans ce cas, sur une vision radicalement différente des opérations de critique des sources, qui nécessiteraient, selon lui, une certaine souplesse au moment de l'exploitation des observations.

Éclairée sous un autre angle, on peut relire sa critique comme une opposition à un changement de pratique de la recherche qu'induirait l'emploi des ordinateurs. Celle-ci reposerait matériellement sur le remplacement d'une technologie intellectuelle traditionnelle de l'historien, la fiche papier ou cartonnée, qui s'articule avec certaines formes de travail collectif, par une autre, le bordereau de perforation, qui suppose la définition de règles collectives du type de celles que nous avons observées.

L'organisation sociale du travail des médiévistes subit une autre modification substantielle du fait de l'utilisation des cartes perforées IBM à 80 colonnes. Comme les données enregistrées ne sont plus modifiables, de nombreuses procédures de vérification multilatérale, visant à prévenir et à détecter les erreurs, sont définies au moment du dépouillement des sources. Comme nous l'avons vu dans l'étude de l'enquête du *catasto* (chapitre 1), de ALINE (chapitre 2) et dans l'analyse des discussions du colloque de Rome (chapitre 3), les données ainsi produites acquièrent alors pour les acteurs un statut scientifique similaire à celui d'une édition critique ou d'une publication. Le modèle de travail collectif qui se constitue dans ces contextes est, à notre connaissance, tout à fait

original en histoire médiévale, et constitue l'une des conséquences marquantes de l'utilisation des ordinateurs par ces chercheurs.

Ce modèle d'organisation est cependant remis en cause dès la fin des années 1970. Au fil de notre étude sur les modalités d'échanges entre les auteurs du *Médiéviste et l'ordinateur* (chapitres 5 et 6), nous avons par exemple vu qu'un autre médiéviste, Alain Guerreau (1948-), pourtant fervent défenseur des outils de calculs programmables, s'opposait non pas au principe du recours aux ordinateurs ou aux cartes perforées IBM, mais aux limites de l'organisation sociale des calculs qu'ils supposaient : dépendance des médiévistes à des statisticiens ignorants de l'épistémologie de l'histoire, nécessité de faire valider les projets par les mandarins de la discipline, manque de formation des historiens en statistiques, etc.⁹⁸¹

Pour permettre aux historiens de ne plus dépendre de cette organisation rigide, Guerreau proposait de s'emparer d'un autre outil de calcul programmable, une calculatrice de poche disposant d'une mémoire interne, qui permettait de travailler de façon autonome et individuelle. Par-là, Guerreau anticipait déjà les impacts de la révolution micro-informatique, qui allait être largement mise à profit par les historiens pour développer une autonomie vis-à-vis des centres de calculs.

On observe en effet tout au long des années 1980 de fulgurantes évolutions techniques qui relèguent les modalités de la saisie des données des années 1970, et en particulier l'utilisation des cartes perforées, dans les limbes de l'histoire de l'informatique. Par exemple, la saisie directe sur support magnétique (cartes magnétiques, disques dur, disquettes, etc.) via micro-ordinateur, attesté dans notre corpus à partir de 1983, assouplit considérablement les procédés d'enregistrement des données, puisqu'elle permet les corrections et les modifications une fois les données enregistrées.

Le développement massif de la micro-informatique bouleverse ainsi en profondeur les modalités de l'organisation collective du travail par rapport à la période précédente. Les utilisateurs peuvent désormais travailler à domicile, le plus souvent seuls, en se connectant en réseau à des ordinateurs plus puissants de grands centres de calculs⁹⁸².

⁹⁸¹ Cette position est clairement affirmée dans l'un des articles analysés dans le chapitre 5 : Alain Guerreau, « Analyse factorielle et analyses statistiques classiques : le cas des ordres mendiants dans la France médiévale », *Annales* 36, n° 5 (1981): 869-912.

⁹⁸² Nous aurions aimé étudier ces évolutions majeures de la fin de la période qui nous intéresse à l'échelle d'un projet d'un collectif. Nous avons d'ailleurs identifié un chantier possible ainsi que les archives de celui-ci l'Institut de Recherche et d'Histoire des Textes (IRHT) à Orléans. Malheureusement, la crise

Lors d'un colloque de 1996 réunissant plusieurs acteurs centraux de notre thèse, parmi lesquels Genet, Marie-Anne Polo de Beaulieu (1950-) ou Monique Zerner (1935-), certains incriminent les ordinateurs personnels d'être à l'origine d'une individualisation des recherches, qui aurait ruiné l'organisation méta-collective et la collaboration entre les chercheurs qui s'était mise en place depuis le milieu des années 1970⁹⁸³.

Pour résumer, l'organisation des collectifs de travail utilisant des ordinateurs en histoire médiévale entre 1966 et 1990 s'articule autour de trois facteurs cruciaux : 1) l'environnement de travail dans lequel ils développent leur projets (outil de calcul choisi, lieux de l'utilisation de ces outils de calcul, interface chercheur-machine) ; 2) le degré de dépendance vis-à-vis des informaticiens⁹⁸⁴ et 3) la dimension des équipes de travail.

À l'échelle de ces cultures de travail locales, nous pouvons également relever une autre conséquence majeure des usages des calculateurs électroniques par les médiévistes, qui se répercute à l'échelle de la discipline toute entière. Comme nous l'avons montré, nombreux sont en effet les acteurs qui produisent, au fil de leurs activités, de nouveaux *instruments de travail* destinés à la communauté des médiévistes. Parmi eux, on peut citer les bases de données - dont nous avons étudié un exemple avec l'enquête du *catasto* (chapitre 1) -, les éditions de textes sur cartes perforées - comme celles de la base MEDITEXT de l'ERA 713 (chapitre 2) -, les programmes informatiques - dont nous avons examiné un exemple avec ALINE (chapitre 2) - et plus largement toutes les formes d'édition de données ou de sources que nous avons mentionnées dans les chapitres 3 et 4 de cette thèse (index, moteur de recherche par requête, systèmes experts, etc.), qu'elles soient publiées sur support de stockage « numérique » ou sur papier⁹⁸⁵.

sanitaire ne nous a plus permis d'avoir accès aux archives depuis mars 2020, laissant ainsi ce projet en friche.

⁹⁸³ Voir par exemple « L'utilisation de l'informatique par les historiens d'aujourd'hui : bilan et perspectives », *Cahiers de la Méditerranée* 53, n° 1 (1996) : 211-17, ou encore l'article déjà plusieurs fois cité Lucie Fossier et al., « Vingt ans d'informatique en histoire médiévale », *Actes des congrès de la Société des historiens médiévistes de l'enseignement supérieur public* 20, n° 1 (1989) : 501-25.

⁹⁸⁴ Sébastien Plutniak a développé dans sa thèse un modèle tricéphale qui nous paraît adapté à la situation que l'on observe en histoire médiévale. Voir en particulier la conclusion de Sébastien Plutniak, « L'opération archéologique. Sociologie historique d'une discipline aux prises avec l'automatique et les mathématiques. France, Espagne, Italie, 2^e moitié du XX^e siècle » (Thèse de doctorat, Paris, EHESS, 2017).

⁹⁸⁵ Nous regretterons seulement ici de ne pas avoir eu la possibilité, en raison de la crise sanitaire, de consacrer un chapitre aux instruments de travail produits par les chercheurs de l'IRHT et du CRAL (CLEO, MEDIUM). Voir la note 1 ci-dessus.

Comme nous l'avons vu dans le chapitre consacré aux premières initiatives d'organisation méta-collective (chapitre 3), la majorité des acteurs envisagent ces nouveaux types d'instruments de travail comme susceptibles de provoquer une révolution de la circulation des données historiques et de leurs usages au sein de la discipline. Premièrement, du fait qu'ils portent sur des corpus de grandes dimensions, et qu'ils doivent donc concourir au développement des méthodes quantitatives en histoire médiévale. Deuxièmement, parce qu'ils rendent possible de nouvelles utilisations de ces corpus, qui passeraient non plus par leur lecture mais par leurs *consultations*⁹⁸⁶, sous la forme d'une interrogation par « requêtes ».

Plusieurs initiatives d'organisation articulées à la production et la circulation de ces instruments de travail se mettent ainsi en place au début des années 1970. Nous l'avons vu dans les développements consacrés au colloque de Rome (chapitre 3), l'une des idées défendues par certains médiévistes était de faire de la production ce type d'instruments de travail l'une des clefs des « progrès » de cette discipline. Dans un article de 1973, que nous avons déjà largement cité, le médiéviste Léopold Genicot (1914-1995) affirmait en conclusion :

*« Dans toutes les sciences, l'organisation et la collaboration sont aujourd'hui une des conditions majeures du progrès. L'histoire n'échappe pas à cette loi. »*⁹⁸⁷

Il mettait ainsi, au cœur de cette marche vers le « progrès », la nécessité d'organiser la collaboration entre les médiévistes, et définissait, dans le même mouvement, l'histoire comme une science comme les autres. Ce type de déclaration nous renseigne sur le fait que le recours aux ordinateurs coïncide avec des transformations plus profondes dans la conception que certains acteurs se font de la discipline au début des années 1970.

Ces acteurs envisagent alors, dans leur domaine de recherche, ces instruments de travail comme permettant de produire des savoirs *cumulables*, de par la possibilité de réutiliser les matériaux des historiens enregistrés sur supports informatiques. Si le terme

⁹⁸⁶ Ce terme a été proposé par Plutniak dans un article récent pour différencier entre deux modes d'utilisation distincts : celui des instruments de travail papier, et celui des instruments de travail que nous qualifierons de numérique.

⁹⁸⁷ Léopold Genicot, « Pour une organisation de la recherche en histoire médiévale », *Francia* 1 (1973): 696.

n'est pas directement employé par les auteurs des années 1970 et 1980, on trouve toutefois cette idée dans la correspondance entre Klapisch-Zuber et Herlihy. Pour rappel, dès les premières années de l'enquête du *catasto*, Herlihy formulait comme l'un de ses objectifs la constitution d'une bibliothèque de bases de données pour l'histoire médiévale, sur le modèle de l'Inter-University Consortium for Political and Social Research (ICPSR)⁹⁸⁸.

De même, Genicot affirmait dans la conclusion du colloque de Rome, dont les actes furent publiés en 1977, que ces nouveaux instruments de travail ne devraient pas seulement répondre aux « *besoins immédiats* » de la recherche, mais aussi aux préoccupations « *futures* » des historiens

*« y compris celles auxquelles nous ne pensons pas aujourd'hui et que le progrès de la problématique suscitera un jour »*⁹⁸⁹.

Cette ambition sur le caractère pérenne des éditions de sources et autres instruments de travail, qui anime les utilisateurs pionniers des ordinateurs dès le début des années 1970, nous la retrouvons aussi ultérieurement.

Au cours d'un colloque qui se tint en 1996 sur le thème « Histoire et informatique », un débat entre les participants les conduit à revenir à cette même problématique. Alain Ruggiero, maître de conférences en histoire contemporaine à l'université de Nice Sophia Antipolis, présente la question en ces termes :

« il est urgent de soulever le problème du savoir cumulable ainsi que celui de la réexploitation du matériel scientifique des historiens qui, c'est vrai, dans la plupart des cas n'est pas réutilisable. L'informatique nous permet d'espérer pouvoir réutiliser les matériaux mis au point par des collègues mais il demeure trop de réticences encore dans les mentalités. Après beaucoup d'espairs mis dans

⁹⁸⁸ Pour une présentation de ce consortium à la fin des années 1960, par l'un de ses initiateurs voir Jerome M. Clubb, « The Inter-University Consortium for Political Research: Progress and Prospects », *Historical Methods Newsletter* 2, n° 3 (1 juin 1969): 1-5.

⁹⁸⁹ « Conclusion générale », *Publications de l'École Française de Rome* 31, n° 1 (1977): 426.

l'outil informatique par la communauté des historiens, nous sommes arrivés au temps de la déception »⁹⁹⁰.

Ce diagnostic nous donne une idée claire du devenir des ambitions portées par des médiévistes comme Herlihy ou Genicot au début des années 1970. En effet, non seulement le problème est toujours urgent en 1996, mais Ruggiero insiste sur le fait qu'à l'engouement a succédé, vingt ans plus tard, la déception.

Toutefois, d'autres formes d'organisations méta-collectives se mettent en place à la fin des années 1970. Autour de Lucie Fossier (1924-) et de la section informatique de l'IRHT s'élabore par exemple la promotion d'un nouveau type de communication entre les utilisateurs des calculateurs électronique, dont le bulletin de liaison *Le Médiéviste et l'ordinateur* constituait la pierre angulaire, comme nous l'avons vu dans les chapitres 3, 4 et 5. Ce nouvel espace d'échanges, au sein duquel il était possible de discuter directement des aspects matériels et techniques de ces pratiques de recherches informatisées, permettait également de présenter les difficultés attachées à ce type de projet. Les éditeurs y encourageaient également la publication des *documents de travail* élaborés par les collectifs utilisant des ordinateurs. Un nouveau *genre* de texte s'était ainsi développé, sous l'impulsion des éditeurs du bulletin, en accord avec les difficultés qu'ils avaient identifiées.

Nuançons cependant la description de la spécificité du phénomène qui nous intéresse ici, puisqu'on peut affirmer que l'introduction d'autres instruments techniques a pu conduire les praticiens d'autres disciplines, et d'autres périodes, à aménager également de nouvelles formes écrites de partage de savoirs et de savoir-faire. Pour le cas particulier des historiens, il n'existe toutefois, à notre connaissance, aucune étude qui montre que les implications de l'introduction de nouvelles formes d'échange savant permettant à un collectif, dans de telles circonstances, de se constituer de nouvelles pratiques scientifiques. Ce type de source (bulletin de liaison, *newsletter*), nous l'avons montré dans le chapitre 4, constitue ainsi une mine d'or pour saisir, d'une part, les logiques de constitution d'un réseau de chercheurs confrontés à des problématiques communes, et, d'autre part, les pratiques de recherche et la science en train de se faire.

⁹⁹⁰ « L'utilisation de l'informatique par les historiens d'aujourd'hui », 212.

Dans notre contexte, nous pouvons encore relever un élément clef tout à fait important : la culture écrite des médiévistes mute au cours de ce processus, et ce en relation avec les nouvelles organisations collectives qu'ils mettent en place. Les stratégies élaborées par les éditeurs du bulletin de liaison jouent en ce sens un rôle crucial. Le premier élément en ce sens tient à l'apparition d'un *genre* de texte, proche du récit d'expérience, dans les pages du *Médiéviste et l'ordinateur*.

Le second élément de cette mutation repose sur le fait que les éditeurs favorisent la publication de nombreux types de technologies intellectuelles attachés à l'utilisation des ordinateurs : masque de saisie, modèle de cartes perforées, manuel de codage, tableau, matrices, graphiques, cartes, etc. Dans ce contexte de publication, non seulement de nouveaux éléments textuels apparaissent, mais les fonctions de technologies intellectuelles préexistantes comme les tableaux, les graphiques ou les cartes mutent également. L'importance accordée par les éditeurs du *Médiéviste et l'ordinateur* à ces deux aspects suggère qu'aux yeux des acteurs, les mutations des formes de l'écrit sont cruciales pour l'évolution de la discipline. On peut même avancer la thèse qu'ils en faisaient l'un des vecteurs de la *culture épistémique commune* qu'ils cherchaient à constituer.

À ce titre, la controverse entre Alain Guerreau (1948-) et Pierre Bonnassie (1932-2005) étudiée dans le chapitre 6 montre les effets de cette évolution des fonctions de certains types d'éléments textuels pour les médiévistes ayant adopté ces nouvelles pratiques. Pour rappel, nous y avons étudié les différentes lectures du tableau élaboré par le médiéviste Michel Banniard (1945-), que l'auteur percevait d'une façon différente de Guerreau. Pour le premier, ce tableau avait pour fonction de résumer de façon synoptique ses lectures sur le développement des institutions féodales en Europe du Sud entre le VII^e et le XI^e siècle. Guerreau y voyait quant à lui une matrice d'incidence permettant la mise en œuvre d'analyses factorielles des correspondances.

De la même manière, les échanges qui suivent entre Guerreau et Bonnassie s'articulent autour de la lecture et de la manipulation de ces différents types de technologies intellectuelles. Les acteurs mettent ainsi au centre des stratégies de constitution d'un réseau et d'une culture épistémique commune le partage de pratiques de production, de construction, de lecture et d'utilisation de textes.

Plus généralement, cette mutation des formes d'échanges dans un contexte de révolution technique constitue, selon nous, un cas particulièrement intéressant pour les historiens des sciences, au-delà du domaine de l'histoire de l'histoire ou de l'histoire des sciences humaines. Il est exemplaire des façons dont certains acteurs donnent corps à une culture épistémique commune en élaborant des stratégies qui visent à faire circuler des pratiques de production et de lecture de textes. De plus, notre approche de cette question illustre clairement l'apport des méthodes de la linguistique pour l'histoire des sciences. En effet, les méthodes de la linguistique nous ont donné les moyens d'examiner les modalités de la circulation des savoirs entre les médiévistes. Qui plus est, en mobilisant certains concepts empruntés aux linguistes nous avons pu dégager certains éléments d'analyse clefs pour notre sujet.

Des linguistes ont notamment mis au point une méthodologie claire, à laquelle nous avons pu recourir, pour saisir ce qu'ils appellent des *communautés discursives*, offrant ainsi des outils pour aborder certaines questions aujourd'hui centrales en histoire des sciences, mais aussi, plus largement, dans la société civile. Qu'est-ce qu'une communauté ? Qu'est-ce qu'une culture ? Comment employer ces concepts sans tomber dans un essentialisme dangereux ? Et donc, comment qualifier les membres de tels collectifs ? Que partagent-ils ? Enfin, comment de telles cultures évoluent-elles ?⁹⁹¹

La mise en œuvre du concept de communauté discursive, ainsi que les outils linguistiques mis au point pour détecter une entité de ce type, ont permis d'aborder ces questions en nous fondant sur des éléments centraux pour les collectifs et les cultures scientifiques : les pratiques de production, de manipulation et de lecture de textes ou d'éléments textuels.

Ces considérations méthodologiques générales nous apparaissent particulièrement importantes pour l'histoire des *digital humanities*, à laquelle notre travail entend contribuer. En effet, les acteurs de ce champ de recherche peinent, depuis plusieurs années, à en définir la nature, les limites, et à en saisir à la fois l'unité et la diversité. Le dépassement d'une approche par les relations interpersonnelles, les institutions, la participation à des rencontres ou les médias sur lesquels se basent les acteurs dans leur travail (textes, images, objets matériels, etc.) pourraient en effet s'effectuer par le biais

⁹⁹¹ À ce propos, voir par exemple Karine Chemla et Evelyn Fox Keller, éd., *Cultures without culturalism: the making of scientific knowledge* (Durham: Duke University Press, 2017).

d'une étude croisée des documents de travail produits par les chercheurs et des nombreuses *pratiques des textes* qui leurs sont attachés. Cette proposition offrirait en effet un angle d'analyse à même d'étudier la diversité et l'unité de ce très vaste champ de recherche⁹⁹².

Pratiques des sources.

Un second ensemble de problématiques nous a intéressé tout au long de ce travail. Il porte sur différentes activités fondamentales du métier d'historien : constitution des corpus, critique des sources et analyses mises en œuvre en lien avec les deux opérations précédentes, que nous avons désignées sous le vocable de *pratiques des sources*. Pour en permettre l'étude, nous avons proposé au fil de nos chapitres de découper schématiquement ces opérations en deux phases distinctes, avec, pour pivot, l'intervention de la machine. Ainsi, nous récapitulerons nos conclusions en deux temps : d'une part, les conséquences des usages des ordinateurs sur la constitution des corpus de données, en amont de leurs traitements par ordinateur, et, d'autre part, les implications de l'utilisation des calculateurs électroniques programmables sur l'analyse et le traitement de ces données.

Le premier ensemble de questions qui porte sur la constitution des corpus concerne ce que nous avons appelé la *modélisation* des sources. En pratique, pour le début de notre période (1966- 1983), le dialogue avec les machines est conduit par l'intermédiaire de cartes perforées IBM à 80 colonnes. Les acteurs doivent donc concevoir la façon dont ils représenteront, sur de telles cartes, les données qu'ils prélèvent dans leurs sources. C'est ce à quoi nous renvoyons ici par le terme *modèle*. L'utilisation des cartes perforées nécessitent donc la construction explicite d'un modèle des sources qui constituent le corpus.

⁹⁹² Nous développons cette idée dans l'introduction et la conclusion du chapitre 4. Voir également Manfred Thaller, « Controversies around the Digital Humanities: An Agenda », *Historical Social Research / Historische Sozialforschung* 37, n° 3 (141) (2012): 7-23; Dr Edward Vanhoutte, *Defining Digital Humanities: A Reader* (Ashgate Publishing, Ltd., 2013); Chris Alen Sula et Heather V Hill, « The early history of digital humanities: An analysis of Computers and the Humanities (1966–2004) and Literary and Linguistic Computing (1986–2004) », *Digital Scholarship in the Humanities* 34, n° Supplement_1 (1 décembre 2019): 190-206.

Le cas d'étude de l'enquête du *catasto* (chapitre 1) nous a donné l'occasion de montrer comment les acteurs abstraient de la source certaines informations selon une formalisation empruntée aux statistiques (individus, variables, modalités). L'étude de ALINE (chapitre 2) a illustré un autre type de modélisation des textes, qui consistait en l'enregistrement de chaînes de caractères sans pré-codage. Dans ce même chapitre, comme dans les chapitres 3, 4 et 5, nous avons dégagé l'existence de nombreux autres types de modélisation des textes. Nous avons vu comment Polo de Beaulieu proposait de considérer les textes comme des chaînes de caractères assortis de descripteurs grammaticaux ou sémantiques (noms propres, verbes, personnages historiques, etc...). Nous avons également étudié comment Fossier modélisait les textes sous la forme d'une arborescence de descripteurs, organisés sous la forme d'un thésaurus, à travers la méthode d'indexation documentaire.

En somme, dans chacune de ces cultures de travail, l'ordinateur amène les acteurs à formuler explicitement comment ils comprennent l'intérêt d'un texte. Les discussions qui prennent forme au colloque de Rome de 1975 (chapitre 3) s'articulent essentiellement autour de la modélisation, étant entendu qu'en fonction des problématiques que poursuivaient les historiens et des moyens humains et financiers dont ils disposaient pour mener leurs recherches, cette opération pouvait prendre des formes tout à fait différentes.

Dans les études de cas que nous avons construites, ces différents types de modélisation devaient par ailleurs être appliqués de manière uniforme à de nombreux documents, ce qui impliquait de penser la cohérence de ces corpus et de justifier leurs limites. À ce titre, nous avons vu les acteurs mettre en œuvre des ensembles documentaire variés, constitués sur la base d'une seule source de grande dimension, comme le *catasto* (chapitre 1). La logique de production de cet ensemble de documents justifiait dans ce cas à elle seule non seulement la cohérence du corpus et ses limites. Pour le cas d'étude du travail de Genet, nous avons au contraire montré que les corpus de textes politiques qu'il constituait étaient en constante évolution, au gré des questions de recherche sur lesquelles Genet avait résolu de se pencher (chapitre 2). Dans ce second cas, ce sont les problématiques historiques qui permettaient de définir la cohérence des corpus et de justifier de leurs limites.

À l'aune de ces deux exemples et d'autres rencontrés au fil de notre travail, on constate qu'une majorité de médiévistes impliqués dans l'utilisation des ordinateurs entendent mettre en œuvre des procédures d'analyses statistiques. Ce mode d'analyse suppose que la fonction du corpus soit radicalement modifiée par rapport à une autre démarche. En effet, le corpus devient, dans cette configuration, un espace de comparaison entre les données extraites des sources, et, ainsi, un élément central pour donner du sens aux résultats des analyses statistiques. Le corpus change ainsi de *fonction épistémologique*, en devenant un espace de comparaison, et donc, de mesure.

Avant de revenir sur les effets épistémologiques de ces deux modes de constitution des corpus, insistons toutefois sur le fait que ces évolutions ne sauraient être attribuées uniquement à l'introduction des calculateurs électroniques dans l'arsenal de l'historien. En effet, dès le début du XX^e siècle, le développement des méthodes quantitatives en histoire avaient également suscité de nouvelles pratiques des sources, qui supposaient déjà de nouvelles modalités de construction de corpus de données quantitatives. À la fin des années 1960, Michel Foucault (1926-1984) avait déjà pointé la mutation de l'épistémologie des historiens causée par l'histoire quantitative. Il voyait dans ce tournant la tendance méthodologique générale au sein de la discipline historique en France, par contraste avec les méthodes de l'épistémologie historique à la Gaston Bachelard (1884-1962), qui servait alors de modèle dominant en histoire des sciences⁹⁹³.

En 1971, dans un article que nous avons abondamment cité dans cette thèse, le moderniste François Furet (1927-1997) insistait lui aussi sur les mutations dues au développement des méthodes quantitatives en histoire, en rappelant que la situation était similaire pour toutes les sciences sociales mettant en œuvre des procédures de quantification et des méthodes statistiques. S'il se concentrait avant tout sur le cas de l'histoire dite *sérielle*⁹⁹⁴, qu'il définissait comme résultant d'une opération logique substituant « à l'« insaisissable » événement de l'histoire positiviste la répétition régulière de données sélectionnées et construites en fonctions de leur caractère comparable »⁹⁹⁵, Furet dégagait plusieurs des points que nous venons de mettre en avant

⁹⁹³ Voir le premier chapitre de Michel Foucault, *L'archéologie du savoir* (Paris, France: Gallimard, 1969).

⁹⁹⁴ Le terme, devenu commun dès les années 1960, est emprunté à Pierre Chaunu, « Histoire quantitative ou histoire sérielle », *Cahiers Vilfredo Pareto* 2, n° 3 (1964): 165-76.

⁹⁹⁵ François Furet, « Histoire quantitative et construction du fait historique », *Annales* 26, n° 1 (1971): 65.

: étude de nouveaux types de sources monumentales, mutation des pratiques de constitution des corpus, transformation des documents en données.

Dans cette analyse, les inflexions dues à l'utilisation des ordinateurs demeuraient toutefois étonnamment absentes. Furet présentait en effet ces nouveaux instruments comme des compagnons de route de ces évolutions, sans que, jamais, il ne se hasarde à affirmer autre chose que l'existence d'une corrélation entre l'instrument technique et les mutations dans les pratiques des sources. Pourtant, l'utilisation des ordinateurs constitue bien une nouvelle étape dans les mutations identifiées par Foucault et Furet pour les années 1960. Face à cette problématique, nous avons montré au fil de nos six chapitres qu'il demeurerait toutefois très difficile de dégager ce qui tiendrait d'un phénomène plutôt que de l'autre.

Deux conclusions différentes nous semblent devoir être dégagées sur la base des cas d'étude des chapitres **1** et **2**. Pour l'enquête du *catasto* (chapitre **1**), nous avons montré que les opérations de constitution des corpus s'inscrivaient dans la continuité des méthodes de l'histoire sociale quantitative et de la démographie. Le recours aux ordinateurs avait toutefois deux principaux effets sur ces opérations. Tout d'abord, sur la modélisation des données, comme nous l'avons déjà vu. Ensuite, sur les dimensions du corpus constitué par Klapisch-Zuber et Herlihy, qui étaient sans commune mesure avec les travaux réalisés sans ordinateurs (60 000 déclarations et 260 000 individus).

Par contraste, nous avons établi dans l'étude de ALINE (chapitre **2**) que les pratiques de constitution des corpus de Genet s'inspirait de méthodes de recherche informatisée, développées en linguistique (lexicométrie) et en sociologie (étude des *champs* à l'aide d'analyse factorielle). Dans ce cas, l'emploi des ordinateurs semble pouvoir être considéré comme le vecteur de *pratiques des corpus*, via l'importation depuis d'autres sciences sociales de méthodes pour lesquelles les opérations de constitution des corpus sont absolument centrales. Genet développa ainsi une utilisation des corpus tout à fait originale dans ce contexte, puisque ceux-ci devenaient tantôt la « matrice du sens » des mots, tantôt la représentation du *champ* du politique.

Les sources que nous avons mobilisées nous ont par ailleurs permis d'étudier un autre temps du travail des médiévistes, l'analyse et le traitement des données entrées dans l'ordinateur. Comme pour ce qui est de la modélisation des sources et des corpus, nous

avons vu que les impacts de l'utilisation des ordinateurs s'inscrivent dans la continuité de l'introduction des méthodes quantitatives en histoire⁹⁹⁶. Dès lors, quel rôle jouent les ordinateurs dans l'évolution de ces méthodes ? Et peut-on vraiment identifier des mutations épistémologiques qui tiennent spécifiquement à l'emploi de calculateurs électroniques ?

Le premier élément sur lequel insistent les acteurs pour justifier de l'intérêt des ordinateurs pour leurs recherches tient au gain de temps que permettent ces instruments dans la mise en œuvre de nombreuses opérations : repérer des mots dans un corpus, opérer des calculs sur des quantités, manipuler et classer des données, etc. La puissance de calcul des ordinateurs donne aux historiens les moyens d'envisager un changement d'échelle dans les analyses, qui peuvent porter sur beaucoup plus de données et être plus facilement répétées et reproduites. L'étude de cas du chapitre 1 illustre par exemple que le recours aux ordinateurs conduit les acteurs à multiplier le nombre de calculs mis en œuvre pour répondre à une même question, puisque les opérations statistiques de base (tris, classement, comptages) sont incroyablement plus rapides à réaliser, une fois les données enregistrées.

Ce bénéfice que les historiens entendent tirer des ordinateurs va de pair avec un autre argument sur lequel ils reviennent régulièrement, comme nous l'avons montré dans les chapitres 1, 2 et 3 : la confiance que les médiévistes accordent aux résultats des opérations effectuées à l'aide des ordinateurs est bien plus grande que celle qu'ils éprouvent envers des analyses faites « à la main », puisque le risque d'erreur de calcul, de classement, de tri ou de dénombrement est désormais nul. Cet élément central étaye la croyance, que nourrissent certains acteurs, selon laquelle l'utilisation des ordinateurs renforcerait « *pouvoir d'analyse* » des historiens quantitativistes⁹⁹⁷.

Ces atouts (gain de temps, fiabilité) qu'attachent les acteurs aux ordinateurs semblent marquer des travaux qui s'inscrivent dans la droite ligne des études d'histoire sociale quantitative des années 1950, comme par exemple, l'enquête du *catasto* (chapitre 1). Ils ne sont toutefois pas spécifiques aux calculateurs électroniques programmables.

⁹⁹⁶ Nous renvoyons ici encore à l'article fondamental de Furet, « Histoire quantitative et construction du fait historique ».

⁹⁹⁷ Nous reprenons ici le terme employé, dans un autre contexte, par Chaunu, pour qualifier les apports des méthodes quantitative en histoire moderne. Voir par exemple Pierre Chaunu, « Pour une histoire sérielle du Brésil au XVIII^e siècle : Réflexions préalables », *Revue d'histoire économique et sociale* 49, n° 4 (1971): 466.

En effet, on retrouve déjà ces mêmes raisons d'utiliser des moyens automatiques dans les études des années 1940 et 1950 présentés en introduction, mais les historiens s'intéressent alors aux machines mécanographiques non programmables⁹⁹⁸. Les arguments du gain de temps et de la fiabilité paraissent ainsi s'inscrire dans la continuité de ces démarches plus anciennes, et ne se présentent donc pas réellement comme une nouveauté liée à l'irruption des ordinateurs.

D'autres conclusions de notre travail mettent en évidence que, pour plusieurs acteurs, le recours aux calculateurs électroniques au sens large (ordinateur, calculatrice programmable de poche, etc.), n'a pas seulement vocation à s'attaquer à des corpus plus grands ou à s'assurer de la fiabilité des calculs. L'ordinateur est également mobilisé pour ses apports conceptuels et épistémologiques, du fait des nouvelles représentations des corpus qu'il permet d'élaborer. Comme nous l'avons vu, des médiévistes comme Genet (chapitre 2) ou Guerreau (chapitre 6) font ainsi état d'un intérêt méthodologique plus fin pour ces nouveaux instruments, sur lequel nous revenons plus en détail dans la section « Conséquences historiographiques » de cette conclusion.

Dans cette direction, deux conséquences tout à fait fondamentales de l'utilisation des ordinateurs sur l'évolution de la pratique des sources doivent toutefois être soulignées dès à présent. La première concerne la possibilité de mettre en œuvre de nouveaux types d'opérations statistiques complexes à l'aide de ces nouvelles machines. Cette révolution pour les statistiques, comme la qualifiait le statisticien britannique Frank Yates (1902-1994) dans un article de 1966⁹⁹⁹, permettait en effet non seulement la multiplication des opérations portant sur une ou deux variables, mais surtout le développement des analyses multivariées. À ce titre, l'emploi répandu des analyses factorielles des correspondances par les médiévistes français de notre corpus reflète de façon frappante cette évolution.

L'analyse factorielle des correspondances constitue, à proprement parler, une véritable nouveauté dans les méthodes des médiévistes (chapitre 1, 2, 4 et 6). Si elle se « démocratise » dans les années 1980, jusqu'à devenir un outil banal dans la boîte de

⁹⁹⁸ William O. Aydelotte, « A Statistical Analysis of the Parliament of 1841: Some Problems of Method », *Historical Research* 27, n° 76 (1954): 141-55 ; Merle Eugene Curti et al., *The making of an american community: a case study of democracy in a frontier country* (Stanford, États-Unis d'Amérique: Stanford University press, 1959); Bernard Bailyn et Lotte Bailyn, *Massachusetts shipping, 1697-1714* (États-Unis d'Amérique: Belknap Press of Harvard University Press, 1959).

⁹⁹⁹ Frank Yates, « Computers, the second revolution in statistics », *Biometrics* 22, n° 2 (1966): 233-51.

l'historien quantitativiste dans les années 1990¹⁰⁰⁰, cette nouvelle approche analytique modifie en profondeur l'épistémologie des acteurs qui s'en emparent dans les années 1970. Dans le chapitre 6, nous avons notamment montré à quel point cette méthode avait pu marquer l'épistémologie défendue par Guerreau, et, de fait, à quel point cela l'avait conduit à se détourner d'une épistémologie plus ancienne, attachée, entre autres, à la figure d'Henri-Irénée Marrou (1904-1977). Au contraire de la démarche défendue par Marrou, le recours aux analyses factorielles se corrèle à l'adoption d'une philosophie qui affirmait l'existence de structures, invisibles pour les acteurs étudiés, et indétectables par le chercheur sauf à mettre en œuvre ce type d'analyses. Dès lors, les analyses factorielles de correspondances représentent pour les historiens un outil étroitement solidaire du développement de nouvelles observations, de l'émergence de nouvelles questions historiographiques ainsi que de nouvelles constructions théoriques.

Le cas du développement du programme ALINE (chapitre 2) montre que ce type de méthode est un vecteur d'interdisciplinarité, puisqu'il charrie avec lui des concepts développés dans d'autres disciplines, et, dans ce cas, le concept de *champ* de la sociologie bourdeusienne. La reprise dans de nouveaux contextes de programmes d'analyses factorielles, qui, comme nous l'avons montré dans le chapitre 2, circulent grâce à des institutions comme l'ADDAD, favorise les transferts méthodologiques entre les disciplines. Tout comme certaines pratiques de modélisation ou de constitution de corpus circulaient sous la forme d'échanges d'instruments de travail (bases de données, d'éditions sur cartes perforées, etc.) entre les collectifs, les méthodes d'analyse circulent elles aussi sous des formes matérielles d'un contexte à l'autre, à travers les programmes informatiques, qui véhiculent avec eux des conceptions épistémologiques plus fondamentales.

Sur un tout autre plan, le principe même de la programmation des opérations, rendu possible par les calculateurs électroniques, induit une autre transformation des méthodes d'analyse des médiévistes. En effet, comme le faisait remarquer le démographe Louis Henry (1911-1991), dont nous avons rapporté les propos dans le chapitre 1, la formalisation des raisonnements que suppose la fabrication des programmes implique de

¹⁰⁰⁰ Un numéro complet de *Histoire & Mesure* revient d'ailleurs sur le développement de ces utilisations en 1997. Voir notamment son introduction Paul-André Rosental, « Introduction. Outil ou fétiche : la laïcisation de l'analyse factorielle dans les sciences sociales », *Histoire & Mesure* 12, n° 3 (1997) : 185-96.

définir explicitement toutes les étapes de ces raisonnements. Elle impose également une forme de systématisation qui consiste, pour chaque problème du même type, à appliquer la même procédure. L'enquête de Klapisch-Zuber et Herlihy (chapitre 1) nous a montré comment l'élaboration de programmes avait rendu possible la réplique du même type d'analyse pour différents ensembles de données extraits du *catasto*, de même que la réplique de méthodes empruntées à d'autres disciplines des sciences sociales (démographie, économie, géographie).

Le principe même de la programmation s'avère avoir d'autres implications d'ordre épistémologique. L'usage des ordinateurs rend nécessaire de définir en amont les principes des raisonnements qui vont être menés. Klapisch-Zuber et Herlihy avaient par exemple, avant même de commencer l'enregistrement des données sur les bordereaux, imaginé un certain nombre des traitements statistiques qu'ils désiraient mettre en œuvre. Dans ce cas particulier, les dimensions de la source ne permettaient pas leur dépouillement exhaustif avant l'enregistrement des données sur cartes perforées. Cette situation contraignait les médiévistes à définir, en amont de l'étude, des hypothèses précises, dans le but de les tester à l'aide des seules données retenues. Cette démarche impliquait par ailleurs de définir, dans le même mouvement, certaines des opérations que les programmes devaient permettre d'exécuter. Par contraste avec d'autres méthodologies historiques pour lesquelles la familiarité avec les sources dicte les objectifs de recherches, cette pratique apparaît comme tout à fait originale.

Cependant, pour certains historiens, les corrélatifs de cette nouvelle approche, en matière de traitement des sources, étaient irrecevables : le médiéviste Conti avait par exemple refusé de participer au dépouillement et à l'enregistrement des données du *catasto* en 1968. Le recours aux ordinateurs impose ainsi à Klapisch-Zuber et Herlihy d'opérer une inversion radicale des opérations de définition du sujet et des méthodes. On comprend, dès lors, que certains observateurs aient qualifié ce type d'approche, dans les années 1990, d'histoire expérimentale¹⁰⁰¹.

Toutefois, à ce modèle très rigide de la fin des années 1960, on peut opposer le cas du programme d'analyse automatique des textes ALINE, étudié dans le chapitre 2. Sur la base de cet exemple, nous avons montré que, par contraste avec l'enquête du

¹⁰⁰¹ Bernard Lepetit, éd., *Les formes de l'expérience : une autre histoire sociale*, L'évolution de l'humanité (Paris: Michel, 1995).

catasto, certains médiévistes avaient pu produire des outils collectifs, qui n'étaient pas nécessairement attachés à un type de corpus ou à une problématique. Cette autre conception du programme comme outil adaptable à différentes données et différentes problématiques constitue, au début des années 1970, une véritable évolution dans le contexte de l'histoire médiévale, qui deviendra la norme dans les années 1980. Les logiciels impliquent alors toujours systématisation et formalisation des raisonnements. Cependant, du fait qu'ils deviennent plus modulaires et se composent de « packages » exécutant chacun une séquence d'opérations, ils permettent de mettre en œuvre des méthodologies qui ne sont plus nécessairement attachées à des hypothèses de travail définies en amont.

Évoquons enfin une dernière conséquence majeure de l'introduction des outils de calcul électroniques dans les pratiques de recherche des médiévistes. Elle porte sur la multiplication de nouveaux types de documents de travail, produits des opérations mises en œuvre à l'aide des programmes informatiques. Ces « documents de sorties », dans le vocabulaire des acteurs, soulèvent de nombreuses questions tout à fait passionnantes pour notre sujet. Une première série d'entre elles concerne le fait que ces documents de travail soient produits de façon automatique, et non plus « à la main ». Ils se présentent dès lors, pour les chercheurs, comme de nouvelles « images » de leurs sources ou de leurs corpus. Quelles différences y a-t-il entre un tableau produit par un programme, puis imprimé sur un listing ou figurant sur un écran, et un tableau élaboré à la main par un médiéviste ? Entre une carte produite automatiquement et une carte dessinée par le médiéviste lui-même ?

Ces problématiques, autant cognitives qu'épistémologiques, nous ramènent à la question générale de l'impact de ces nouveaux environnements matériels sur les pratiques de l'écrit. Ce que nous avons montré au fil de nos analyses, c'est que ce changement du mode de production de certains documents de travail s'accompagnait d'une modification de la fonction épistémologique de certaines de ces technologies intellectuelles. Là où certains historiens utilisaient les cartes, les tableaux ou les graphiques pour illustrer les résultats de leurs analyses - ce que le géographe Jacques Bertin (1918-2010) qualifiait

dans son livre de 1967 d' « *image morte* »¹⁰⁰² - la majorité des médiévistes que nous avons étudiés ici utilisent les documents de sortie des ordinateurs comme des outils heuristiques.

Ce changement de la fonction de ces visualisations est-il lié au fait que ces documents de travail soient produits automatiquement ? La réponse n'est pas aisée à donner, notamment parce que certains des acteurs qui nous ont intéressés, comme par exemple le médiéviste Guerreau ou le moderniste Antoine Prost (1933-), traçaient eux-mêmes leurs graphiques d'analyses factorielles à la main, à l'aide de papier logarithmique et de tracettes. Pour autant, comme nous l'avons montré dans le chapitre 6, Guerreau utilise bel et bien le tableau de Banniard et les graphiques d'analyses factorielles qu'il produit sur sa base comme des outils heuristiques, qui lui permettent non pas de résumer un ensemble de lectures, mais de faire émerger des questions. Les acteurs produisent donc non seulement des nouveaux types de documents de travail aux fonctions épistémologique inédites, mais ce sont également les fonctions de technologies intellectuelles préexistantes (tableaux, graphiques, cartes) qui mutent dans ce contexte.

Sur ce point, nous avons établi que ces mutations reposaient essentiellement les modes de lecture de ces éléments textuels. En effet, le chapitre 6 nous a permis de montrer que pour deux médiévistes aux méthodes de travail différentes, un même tableau pouvait être perçu comme présentant des fonctions tout à fait différentes : illustrative pour l'un et heuristique pour l'autre. Le cas d'étude du chapitre 4 suggère que c'est par le truchement de la culture épistémique commune qui se met en place grâce à des initiatives comme celles des éditeurs du *Médiéviste et l'ordinateur* que les documents de travail sont mis en partage, en même temps que les nouvelles fonctions que les acteurs peuvent désormais leur attribuer¹⁰⁰³.

L'étude des modes de lecture de ces nouveaux documents, et de ce qu'en font les historiens, apparaît comme l'un des axes de recherches déterminant pour comprendre l'évolution plus générale des *pratiques des textes* que nous avons identifiée. Cette remarque semble d'ailleurs pouvoir s'appliquer à d'autres types d'acteurs, et mériterait,

¹⁰⁰² Jacques Bertin et al., *Sémiologie graphique : les diagrammes, les réseaux, les cartes* (Paris, France: Éditions de l'EHESS, 2013), 8.

¹⁰⁰³ Voir notamment l'article suivant, que nous avons analysé en détail dans la troisième partie du chapitre 4 : Jean-Philippe Genet, « Une application de l'analyse factorielle à l'étude du vocabulaire », *Le médiéviste et l'ordinateur* 5, n° 1 (1981): 11-15.

selon nous, la plus grande attention des observateurs intéressés par ce sujet. L'évolution de ces pratiques, dans des contextes où les ordinateurs personnels disposant d'un terminal vidéo sont plus répandus, et où le standard « What You See Is What You Get » s'impose, constituerait ainsi la suite logique de l'examen de cette problématique capitale pour comprendre les conséquences de l'utilisation des ordinateurs sur les cultures scientifiques de l'écrit.

Conséquences historiographiques.

Nous avons examiné la nouvelle culture qui se met en place tant du point de vue des types de collectifs qui se forment pour déployer des projets d'un nouveau type que de celui des pratiques de l'écrit qui la caractérisent. Ces différentes facettes présentent un intérêt d'autant plus grand qu'elles s'articulent pour produire des changements historiographiques au cours de la période qui nous intéresse. Revenons à titre de conclusion à ces observations.

Les évolutions des règles du travail collectif et les mutations des pratiques des sources nous permettent de montrer que s'opèrent une évolution majeure de l'épistémologie des médiévistes entre 1966 et 1989. Pour résumer, nous présenterons schématiquement que les usages des ordinateurs par les acteurs s'articulent autour de deux *pôles*, qui correspondent à deux rôles que les acteurs attribuent à ces machines.

D'une part, les ordinateurs sont considérés par certains médiévistes, essentiellement au début de la période qui nous intéresse, comme un outil de calcul plus rapide et plus performant pour l'analyse de corpus aux dimensions monumentales. Pour ces collectifs, qui poursuivent des objectifs scientifiques variés, l'ordinateur se présente comme un instrument facilitant la mise en œuvre des méthodes préexistantes de critique des sources, du fait qu'il permet d'automatiser la réalisation de certaines opérations (identification de mot, dénombrements, tris, classement, etc.). Dans ce type de travaux, nous l'avons montré dans les chapitres 1 et 3, l'utilisation des ordinateurs est attachée à des valeurs épistémiques centrales pour des programmes historiographiques qui envisagent l'histoire comme une science sociale : exactitude, cumulativité, fiabilité, exhaustivité.

Le cas de l'enquête du *catasto* constitue un exemple paradigmatique de ce rôle attribué aux ordinateurs en histoire médiévale. De même, nous avons rencontré au fil de nos six chapitres de nombreux projets qui s'inscrivaient dans la droite ligne de cette forme d'enquête, comme ceux qu'ont élaborés les équipes réunies autour de Fossier à l'Institut de Recherche et d'Histoire des Textes (IRHT), de Karl Ferdinand Werner (1924-2008) à l'Institut Historique Allemand ou de Cinzio Violante (1921-2001) à l'université de Pise (chapitre 3).

D'autre part, d'autres médiévistes s'étaient emparés des moyens électroniques de calcul avec des ambitions scientifiques toute différentes. Pour saisir cet écart, la critique qu'adressait Genet à l'étude du *catasto* dans sa recension du livre *Les Toscans et leurs familles* fait office de révélateur (chapitre 1). L'auteur y affirmait regretter le fait que l'ordinateur ait surtout été utilisé par Klapisch-Zuber et Herlihy pour « *peaufiner la description* », plutôt que pour chercher à « *faire apparaître les structures complexes à l'aide d'analyses factorielles* »¹⁰⁰⁴. Cette critique résume, dans les termes d'un acteur, l'une des différences épistémologiques fondamentales qui existe entre les deux types d'utilisation des ordinateurs que nous essayons ici de définir. Elle est d'autant plus pertinente à nos yeux que, rappelons-le, Herlihy avait rejeté l'utilisation des analyses factorielles des correspondances au moment de l'écriture du livre concluant l'enquête, en dépit de la publication par Klapisch-Zuber de plusieurs articles mobilisant ce type de méthodes (chapitre 1). Cette anecdote illustre de manière frappante l'écart qui existait entre les deux pôles que nous avons identifiés.

Les représentations des fonctions des ordinateurs par les acteurs du premier pôle, telles que nous les avons définies ci-dessus, éclairent la spécificité des usages que Genet ou Guerreau font des calculateurs électroniques programmables à partir de la deuxième moitié des années 1970. Comme nous l'avons analysé dans les chapitres 2 et 6, les moyens électroniques de calcul sont, pour ces deux chercheurs, des instruments à vocation heuristique, qui doivent permettre de renouveler les méthodes permettant l'étude de problématiques historiographiques centrales de la discipline (genèse de l'État moderne ou évolution des institutions féodales) et d'élaborer de nouvelles constructions conceptuelles (champ du politique, structure centre-périphérie). Dans cette optique, et

¹⁰⁰⁴ Jean-Philippe Genet, « David Herlihy et Christiane Klapisch-Zuber, *Les Toscans et leurs familles. Une étude du catasto florentin de 1427, 1978* », *Francia* 7 (1979): 711.

pour ces deux médiévistes, les analyses factorielles des correspondances représentaient l'une des procédures fondamentales, et justifiaient, presque à elles seules, l'intérêt de ces nouveaux outils de calculs.

L'étude d'une controverse au chapitre 6 a éclairé comment Guerreau employait ce type de procédure. Elle a également dégagé les implications épistémologiques de leur mise en œuvre, par contraste avec l'épistémologie défendue par un médiéviste aux méthodes plus traditionnelles, Bonnassie. L'analyse des différences de vues entre ces deux chercheurs prouve que la mise en œuvre des analyses factorielles des correspondances se conjugait avec de profondes modifications des pratiques de l'observation historique et des opérations de synthèse fondées sur ces observations. À ces disparités correspondait encore, dans ce cas, une différence d'objet d'étude : là où Bonnassie attribuait à l'histoire l'objectif de saisir la réalité historique des acteurs, Guerreau cherchait, quant à lui, à observer les structures qui se cachaient derrière les données.

À la lumière de notre travail, plusieurs éléments apparaissent pertinents pour expliquer la distinction entre ces deux pôles. Le premier tient directement aux modalités de la mise en œuvre des calculateurs électroniques élaborées par ces historiens. Par contraste avec la forme paradigmatique de division sociale des calculs qu'illustre l'exemple du *catasto*, et qui reposait sur une dépendance très forte aux informaticiens et aux centres de calculs, Genet et Guerreau s'étaient aménagés, dans les années 1970, des situations d'indépendance, plus ou moins marquées, vis-à-vis des informaticiens. Mes recherches paraissent étayer l'hypothèse que ce dernier type d'organisation a joué un rôle décisif sur les innovations méthodologiques et conceptuelles qui sont apparues dans ce contexte. Cette modalité d'utilisation des calculateurs électroniques s'inscrivait toutefois dans des environnements de travail dans lesquels les médiévistes avaient développé une grande maîtrise technique, non seulement de certaines méthodes statistiques, mais également dans le maniement des calculateurs électroniques programmables.

Pour éclairer les différences entre les usages développés par ces deux médiévistes et ceux que nous avons présentés plus haut, il semble pertinent d'insister sur un second élément. Celui-ci repose sur la périodisation des évolutions technologiques, dont le point de pivot central est le remplacement des machines mécanographiques par les outils de

calculs électroniques programmables dans l'arsenal technique des historiens à la fin des années 1960. En effet, chez les historiens du premier groupe que nous avons présenté ci-dessus, on retrouve une continuité forte entre les fonctions attribuées aux machines mécanographiques et celles dévolues aux ordinateurs. Il ne semble dès lors pas absurde d'affirmer que ces derniers instruments sont considérés, par ces acteurs, comme un équivalent, quoique plus puissant, d'une machine mécanographique, ou, en d'autres termes, comme une puissante machine de tri et de classement. L'engouement qu'avait pu provoquer l'utilisation des machines mécanographiques pour la recherche en histoire moderne, dont l'article largement cité de Adeline Daumard (1924-2003) et Furet de 1959 était l'un des points d'orgue¹⁰⁰⁵, aurait ainsi été perpétué par les médiévistes en direction des ordinateurs disponibles dans les centres de calcul à la fin des années 1960, sans pour autant que les nouvelles fonctionnalités (programmation, mémoire interne) de ces machines soient complètement exploitées.

Dans les années 1970, certains médiévistes s'étaient alors emparés des ordinateurs en les utilisant en accord avec leurs nouvelles fonctionnalités. L'un des principaux effets de ces nouveaux usages venait du fait que ces historiens recouraient alors massivement à de nouvelles procédures d'analyse des données, telle que les analyses factorielles des correspondances, la classification automatique ou le développement de systèmes experts, auxquels nous n'avons, dans ce travail, pas assez rendu justice. Ce regret motive cependant la possibilité de lancer, autour de ces deux derniers objets, des recherches futures, par exemple sur la modélisation et la réplique de certains raisonnements historiques.

Pour le cas des analyses factorielles des correspondances, nous avons vu que les possibilités qu'elles offraient concouraient au développement d'approches structuralistes mathématisées, en permettant la production d'une image d'un espace conceptuel¹⁰⁰⁶. L'analyse factorielle des correspondances apparaissait ainsi comme un nouvel instrument d'observation, à même de repérer les relations des données entre elles, et donc, les structures qui se cachaient dans ces ensembles d'observations formalisées.

¹⁰⁰⁵ Adeline Daumard et François Furet, « Méthodes de l'Histoire sociale : les Archives notariales et la Mécanographie », *Annales* 14, n° 4 (1959): 676-93.

¹⁰⁰⁶ Voir par exemple Julien Duval, « L'analyse des correspondances et la construction des champs », *Actes de la recherche en sciences sociales* N° 200, n° 5 (2013): 110-23.

L'examen de la postérité de deux *pôles* d'usages des ordinateurs que nous avons présentés ici suggère que ce sont avant tout les approches du second type qui se sont transmises aux générations suivantes, du moins dans le contexte académique français. En 1986, la création de la revue *Histoire & Mesure* a marqué un tournant, que l'on pourrait faire coïncider avec une mutation définitive de l'histoire quantitative. Rappelons que dans l'éditorial du premier numéro, Genet, l'un des fondateurs de la revue, affirmait :

« Pourquoi donc créer, aujourd'hui en 1985, une revue intitulée Histoire & Mesure ? Le besoin, l'urgence sont là. Ils sont révélés en histoire comme dans les autres sciences humaines et sociales, par l'ordinateur, qui a profondément transformé la pratique de nombreux historiens. Tout d'abord, il a obligé à travailler sur un ensemble de données, extraites de la source, qui sont construites et pourvues nécessairement d'une structure finie. Il a contraint l'historien à chercher les outils optimaux d'analyse et d'interprétation et à effectuer des choix décisifs dans l'arsenal sans cesse renouvelé que lui offre l'analyse des données. Dès lors il n'est plus d'histoire quantitative ou d'histoire qualitative. Tout se mesure : les flux, les évolutions et les ruptures, les équilibres et les déséquilibres peuvent être dégagés, circonscrits, dessinés. »

Avec cette présentation, par l'un des acteurs, des conséquences de l'utilisation des ordinateurs sur les pratiques de recherche - qui recoupe plusieurs des conclusions que nous avons formulées ici - il est possible de conclure que l'une des conséquences historiographiques majeures de l'utilisation des ordinateurs consiste dans le passage d'une *épistémologie du quantitatif* à une *épistémologie de la mesure*.

Une autre conséquence fondamentale de l'utilisation des ordinateurs par les médiévistes doit cependant être mise en lumière ici. Dans un article de 2014, Nicolas Perreaux présentait l'intérêt de plusieurs bases de données, produites par les médiévistes du premier pôle, pour le développement des recherches sur l'histoire de l'Occident médiéval¹⁰⁰⁷. Il y mentionnait par exemple les *Chartes Originales* de l'Atelier de

¹⁰⁰⁷ Nicolas Perreaux, « De l'accumulation à l'exploitation ? Expériences et propositions pour l'indexation et l'utilisation des bases de données diplomatiques », in *Digital diplomatics. The computer as a tool for diplomatist?*, eds. Antonella Ambrosio, Sébastien Barret, et Georg Vogeler (Köln, Weimar, Wien: Böhlau Verlag, 2014), 187-211.

Recherche sur les Textes Médiévaux (ARTeM)¹⁰⁰⁸ ou le *Deeds Project*, deux instruments de travail initiés par des médiévistes utilisant des ordinateurs dans les années 1970¹⁰⁰⁹. À propos de leurs utilisations par les médiévistes, Perreaux faisait le diagnostic suivant :

« *Le constat est donc simple : il existe un décalage important entre la création de bases de données – activité en quoi les médiévistes excellent, et c'est une bonne chose – et l'exploitation de celles-ci* »¹⁰¹⁰.

Le décalage identifié ici par Perreaux illustre une tout autre postérité pour les projets qui s'inscrivent dans le premier pôle que nous avons défini. Il nous incite par ailleurs à nous tourner vers de nouvelles questions, que nous considérons capitales pour comprendre l'impact de l'utilisation des ordinateurs sur le métier de médiévistes et sur leurs pratiques de l'écrit.

Les instruments de travail produits depuis le début des années 1970 ont-ils vraiment été sous-exploités par les médiévistes ? Si oui, pourquoi ? Si non, comment ont-ils été utilisés ? Et quelles différences peut-on identifier entre l'emploi de ces outils tel qu'envisagée par leurs producteurs et l'usage qu'en ont proposé les autres utilisateurs ? L'étude de ces problématiques nous apparaît comme la continuité logique de nos recherches. Elle permettra, nous l'espérons, de comprendre comment les chercheurs utilisent les instruments de travail numériques (moteur de recherche, bases de données, éditions électroniques, etc.) et, ainsi, de percevoir encore un peu plus finement les impacts de l'emploi des ordinateurs sur les cultures scientifiques au sein desquels le travail sur les écrits et par l'écrit joue un rôle clef.

¹⁰⁰⁸ Ce projet du début des années 1970 s'inscrit dans la continuité du travail menée par Lucie Fossier au Centre de Recherche et d'Applications Linguistique (CRAL) de l'université de Nancy au début des années 1970 et du projet d' « index verborum » des documents diplomatiques lorrains, dont nous avons analysé certains usages à travers l'article publié par Michel Parisse (1936-2020) dans les actes du colloque de Rome en 1977 : Michel Parisse, « Traitement des documents diplomatiques: exploitation de l'« Index Verborum » », *Publications de l'École Française de Rome* 31, n° 1 (1977): 335-45. Pour une présentation plus récente de ce corpus, voir Benoît-Michel Tock et al., éd., *La diplomatie française du Haut Moyen Age: Inventaire des chartes originales antérieures à 1121 conservées en France*, vol. 4, Atelier de recherche sur les textes médiévaux (Turnhout: Brepols Publishers, 2001). Disponible en ligne : <http://www.cn-telma.fr/originaux/colophon/>

¹⁰⁰⁹ Ce projet anglais, qui trouve son origine en 1975, est notamment présenté dans Michael Gervers, « The DEEDS project and the development of a computerised methodology for dating undated English private charters of the twelfth and thirteenth centuries », *Dating Undated Medieval Charters*, 2000, 13-35. Le corpus est disponible ici : <http://deeds.library.utoronto.ca/>

¹⁰¹⁰ Perreaux, « De l'accumulation à l'exploitation ? », 190-191.

Bibliographie générale.

- Actes des congrès de la Société des historiens médiévistes de l'enseignement supérieur public*, 10^e congrès, Lille, 1979. Le paysage rural : réalités et représentations., 1979. https://www.persee.fr/issue/shmes_1261-9078_1980_act_10_1.
- Adams, Margaret. « Punch Card Records: Precursors of Electronic Records ». *The American Archivist* 58, n° 2 (1 avril 1995): 182-201. <https://doi.org/10.17723/aarc.58.2.d61078182725616j>.
- d'Angeville, Adolphe, et Emmanuel Le Roy Ladurie (Préfacier). *Essai sur la statistique de la population française considérée sous quelques-uns de ses rapports physiques et moraux*. Paris, France: Mouton, 1969.
- Anheim, Étienne. « L'historiographie est-elle une forme d'histoire intellectuelle? La controverse de 1934 entre Lucien Febvre et Henri Jassemin ». *Revue d'histoire moderne contemporaine* n° 59-4bis, n° 5 (2012): 105-30.
- « Annales. Économies, Sociétés, Civilisations. 25^e année, N. 5, 1970. » 25, n° 5 (1970). https://www.persee.fr/issue/ahess_0395-2649_1970_num_25_5.
- d'Arbois de Jubainville, Henri, et Georges Dottin. *Recherches sur l'origine de la propriété foncière et des noms de lieux habités en France : période celtique et période romaine*. Paris, France: Ernest Thorin, 1890.
- Arcamone, Maria Giovanna. « Per lo spoglio elettronico dell'onomastica italiana altomedievale ». *Publications de l'École Française de Rome* 31, n° 1 (1977): 193-97.
- Aron, Raymond. « Introduction à la philosophie de l'histoire : essai sur les limites de l'objectivité historique ». Paris, France: Gallimard, 1938.
- Auffray, Yves. « I. Traitement informatique ». *Mélanges de la Casa de Velázquez* 12, n° 1 (1976): 141-45. <https://doi.org/10.3406/casa.1976.2223>.
- Aurell, Martin. « Les Cahiers de Civilisation Médiévale au service de l'interdisciplinarité. » In *Actes de la Table Ronde de l'Institut Historique Allemand*, 10-25. Paris, France, 2009. <https://hal.archives-ouvertes.fr/hal-00444036>.

- Austin, Erik W. « The Historical Data Resources of the Inter-University Consortium for Political and Social Research ». *Historical Social Research / Historische Sozialforschung*, n° 12 (1979): 43-45.
- Autrand, Françoise. « Le personnel du Parlement de Paris ». *Publications de l'École Française de Rome* 31, n° 1 (1977): 239-43.
- . « Naissance d'un grand corps de l'État : les gens du parlement de Paris, 1345-1454 ». Thèse d'État, Université Paris 1 Panthéon-Sorbonne, 1978.
- Aydelotte, William O. « A Statistical Analysis of the Parliament of 1841: Some Problems of Method ». *Historical Research* 27, n° 76 (1954): 141-55. <https://doi.org/10.1111/j.1468-2281.1954.tb01131.x>.
- . « Voting Patterns in the British House of Commons in the 1840s ». *Comparative Studies in Society and History* 5, n° 2 (1963): 134-63.
- Bailyn, Bernard, et Lotte Bailyn. *Massachusetts shipping, 1697-1714*. États-Unis d'Amérique: Belknap Press of Harvard University Press, 1959.
- Balard, Michel. « Les actes notariés de Gênes ». *Le médiéviste et l'ordinateur* 3, n° 1 (1980): 3-5. <https://doi.org/10.3406/medio.1980.1639>.
- Baldwin, Sandy. « The Idiocy of the Digital Literary (and what does it have to do with digital humanities)? » *Digital Humanities Quarterly* 007, n° 1 (1 juillet 2013).
- Banniard, Michel. *Le haut Moyen Âge occidental*. Paris, France: PUF, 1980.
- Barthélemy, Dominique. « La mutation féodale a-t-elle eu lieu? (note critique) ». *Annales* 47, n° 3 (1992): 767-77. <https://doi.org/10.3406/ahess.1992.279072>.
- . « La théorie féodale à l'épreuve de l'anthropologie (note critique) ». *Annales* 52, n° 2 (1997): 321-41. <https://doi.org/10.3406/ahess.1997.279569>.
- Bashe, Charles J, Lyle R. Johnson, John H. Palmer, Emerson W. Pugh. *IBM's early computers*. Cambridge, États-Unis d'Amérique: The MIT Press 1986.
- Battelli, Giulio. « Una proposta per un indice dei registri pontifici ». *Publications de l'École Française de Rome* 31, n°1 (1977): 19-29.
- Bautier, Robert-Henri. « Caractères spécifiques des chartes médiévales ». *Publications de l'École Française de Rome* 31, n° 1 (1977): 81-96.
- . « Les demandes des historiens à l'informatique ». *Publications de l'École Française de Rome* 31, n° 1 (1977): 179-86.

- Béaur, Gérard, et Béatrice Marin. « La Statistique Générale de la France – Présentation ». *L'Atelier du Centre de recherches historiques. Revue électronique du CRH*, 17 mars 2011. <http://journals.openedition.org/acrh/2891>.
- Bell, Daniel. *The coming of post-industrial society: a venture in social forecasting*. New York, États-Unis d'Amérique: Basic books, 1973.
- . *Vers la société post-industrielle*. Paris, France: Éditions Robert Laffont, 1973.
- Benedict, Philip. « Between Whig Traditions and New Histories: American Historical Writing about Reformation and Early Modern Europe ». In *Imagine Histories*. Princeton University Press, 1999. <https://doi.org/10.1515/9780691187341-016>.
- Benzécri, Jean-Paul. *Histoire et préhistoire de l'analyse des données*. Paris, France: Dunod, Bordas, 1982.
- . « Philosophie Thomiste Et Connaissance Mathématique de la Nature ». *Aquinas* 11, n° 1 (1968): 37.
- Berges, Wilhelm. *Die Fürstenspiegel des hohen und späten Mittelalters*. Leipzig, Allemagne, 1938.
- Bergounioux, Gabriel. « How Statistics Entered Linguistics: Pierre Guiraud at Work. The Scientific Career of an Outsider ». *Glottometrics* 33 (2016): 45-55.
- Berra, Aurélien. « Pour une histoire des humanités numériques ». *Critique, Des chiffres et des lettres : les humanités numériques*, 819-820 (août 2015): 613-26.
- Berthe, Maurice, et Pierre Bonnassie. « In memoriam Philippe Wolff (1913-2001) ». *Annales du Midi* 114, n° 237 (2002): 5-24.
- Bertin, Jacques, Marc Barbut, Serge Bonin, et Guy Arbellot. *Sémiologie graphique : les diagrammes, les réseaux, les cartes*. Paris, France: Éditions de l'EHESS, 2013.
- Besnard, Marie-Pierre. « Notre-Dame de Saint-Lô nella Manica. Una restituzione virtuale in 3D dopo la distruzione durante la Seconda guerra mondiale ». *Memoria e ricerca*, n° 37 (2011): 193-212.
- Biber, Douglas. *Variation across Speech and Writing*. Cambridge: Cambridge University Press, 1988. <https://doi.org/10.1017/CBO9780511621024>.
- Biget, Jean-Louis, Jean-Claude Hervé, et Yvon Thébert, eds., *Les cadastres anciens des villes et leur traitement par l'informatique : actes de la table ronde, 31 janvier-2 février 1985, Saint-Cloud*. Rome, Italie: École française de Rome, 1989, 1989.

- Billot, Claudine, et Guy Lobrichon. « À propos des thesaurus ». *Le médiéviste et l'ordinateur* 7, n° 1 (1982): 22-22.
- Blackith, Robert Edmund. « Multivariate Analysis : Not so Much a Technique as a Weltanschauung ». *Biometrics* 2, n°30 (1974): 386-87.
- Blakeslee, Merritt. « Lo dous jocx sotils : la partie d'échecs amoureuse dans la poésie des troubadours ». *Cahiers de Civilisation Médiévale* 28, n° 110 (1985): 213-22. <https://doi.org/10.3406/ccmed.1985.2297>.
- Bloch, Marc. « Mr Karl Michaelsson, Études sur les noms de personne français d'après les rôles de taille parisiens [rôles de 1292, 1296-1300, 1313] ». *Annales* 4, n° 13 (1932): 67-69. <https://doi.org/10.3406/ahess.1932.1586>.
- Bloch, Marc. *La société féodale*. 2 vol. Paris, France: Albin Michel, 1939.
- Bodler, Markus. « Les historiens français et les États-Unis dans les années 1950 et 1960 ». *Nouveaux mondes nouveaux - Novo Mundo Mundos Novos - New world New worlds*, 29 janvier 2010. <https://doi.org/10.4000/nuevomundo.58811>.
- Bodson, Arthur, et Suzanne Govaerts. « Codification d'un texte latin sur cartes mécanographiques 80 colonnes ». *Revue de l'Organisation Internationale pour l'Étude des Langues Anciennes par Ordinateur*, n° 1 (1966): 1-49.
- Bonnain, Rolande. « Michel Demonet, Tableau de l'agriculture française au milieu du XIX^e siècle, l'enquête de 1852. » *Études rurales* 121, n° 1 (1991): 275-275.
- Bonnassie, Pierre, et Alain Guerreau. « Échos du numéro 5 : Peut-on factorialisier le Moyen Âge ? » *Le médiéviste et l'ordinateur* 6, n° 1 (1981): 17-20. <https://doi.org/10.3406/medio.1981.979>.
- Bonnassie, Pierre, et Philippe Wolff. « La Catalogne : du milieu du X^e à la fin du XI^e siècle ». Toulouse, France: Association des publications de l'Université de Toulouse-Le Mirail, 1976 1975.
- Bonnet, Philippe. *Chapitre 2. Pour une histoire sociale de l'analyse des données. La méthodologie de Pierre Bourdieu en action*. Dunod, 2015. <https://www.cairn.info/la-methodologie-de-pierre-bourdieu-en-action--9782100703845-page-21.html>.
- Borghetti, Maria Novella. *L'œuvre d'Ernest Labrousse: genèse d'un modèle d'histoire économique*. Paris: Éditions de l'EHESS, 2005.

- Bourdieu, Pierre, Jean-Claude Chamboredon, Jean-Claude Passeron *Le métier de sociologue : préalables épistémologiques*. Paris, France : Mouton, 1973.
- Bourlet, Caroline, Charles Doutrelepon, et Serge Lusignan. *Ordinateur et études médiévales : Bibliographie*. Montréal, Canada: Institut d'études médiévales, 1982.
- Bourlet, Caroline, et Michel Parisse. « ... En direct de Nancy : l'automate et les chartes ». *Le médiéviste et l'ordinateur* 2, n° 1 (1979): 8-10.
<https://doi.org/10.3406/medio.1979.902>.
- Bourlet, Caroline, Lucie Fossier, Jean-Philippe Genet, Christiane Klapisch-Zuber, Jacques Lefort, Josette Metman, et Gian Piero Zarri. « Éditorial ». *Le médiéviste et l'ordinateur* 1, n° 1 (1979): 1-2.
- Bowles, Edmund A. « Towards a Computer Curriculum for the Humanities ». *Computers and the Humanities* 6, n° 1 (1971): 35-38.
- Bratley, Paul, Serge Lusignan, et Francine Ouelette. *JEUEMO: a text-handling system*. Edinburg, Scotland: Edinburgh University Press, 1974.
- . « Quelques réflexions et beaucoup d'information sur les problèmes de « sorties » ». *Le médiéviste et l'ordinateur* 12, n° 1 (1984): 15-20.
<https://doi.org/10.3406/medio.1984.1696>.
- Braudel, Fernand, et Ernest Labrousse, éd., *Histoire économique et sociale de la France*. Paris, France: Quadrige Presses Universitaires de France, 1977-1993.
- Braunstein, Philippe. « La Toscane dans le premier tiers du XV^e siècle : informatique et paysage social (note critique) ». *Annales* 35, n° 1 (1980): 42-51.
<https://doi.org/10.3406/ahess.1980.282607>.
- Bresson, Alain. « Recherches sur la société Rhodienne (480 avant J.-C.-100 après J.-C.) ». Lille, France: A.N.R.T., 1994.
- Bretelle-Establet, Florence. « The Issue of Textual Genres in the Medical Literature Produced in Late Imperial China ». In *Texts, Textual Acts and the History of Science*, édité par Jacques Virbel et Karine Chemla, 87-142. New York, États-Unis d'Amérique: Springer Verlag, 2015.
- Brian, Éric. « L'horizon nouveau de l'historiographie expérimentale ». *Revue d'histoire moderne contemporaine* n° 58-4bis, n° 5 (2011): 41-56.

- Brisset, Nicolas. « Qu'est-ce qu'un contexte? Quentin Skinner et l'histoire des idées ». *Æconomia. History, Methodology, Philosophy*, n° 10-3 (1 septembre 2020): 581-92.
- Brown, Elizabeth A. R. « The Tyranny of a Construct: Feudalism and Historians of Medieval Europe ». *The American Historical Review* 79, n° 4 (1974): 1063-88. <https://doi.org/10.2307/1869563>.
- Brown, Peter. *The world of late Antiquity: from Marcus Aurelius to Muhammad*. London, Royaume-Uni: Thames and Hudson, 1971.
- Bruhns, Hinnerk, Joachim Nettelbeck, et Maurice Aymard, éd., *Clemens Heller: imprésario des sciences humaines*. Paris, France: Maison des sciences de l'homme, 2017.
- Brunet, Etienne. « Hyperbase : logiciel documentaire et statistique pour l'exploitation des grands corpus ». *Tools for humanists*, 1989, 33-36.
- . « Qui lemmatise dilemme attise ». In *11^e Rencontres linguistiques en pays rhénan*, édité par Publications de l'Université Marc Bloch, 7-32. Strasbourg, France: L Kosé, A. Theissen, 1999. <https://hal.archives-ouvertes.fr/hal-01575442>.
- Burguière, André. « Archéologie du Centre de Recherches Historiques. Les enquêtes collectives de Marc Bloch et de Lucien Febvre et leur postérité ». *Les Cahiers du Centre de Recherches Historiques. Archives*, n° 36 (30 juin 2005). <https://doi.org/10.4000/ccrh.3039>.
- . *L'école des annales : une histoire intellectuelle*. Paris: O. Jacob, 2006.
- Burke, Peter. « The “Annales” in global context ». *International Review of Social History* 35, n° 3 (1990): 421-32.
- . *The French historical revolution: the Annales school, 1929 - 89*. Cambridge: Polity Press, 1990.
- Burnard, Lou. « Du literary and linguistic computing aux digital humanities : retour sur 40 ans de relations entre sciences humaines et informatique ». In *Read/Write Book 2: Une introduction aux humanités numériques*, édité par Pierre Mounier, 45-58. Marseille, France: OpenEdition Press, 2012. <http://books.openedition.org/oep/242>.

- . « Histoire de la TEI : un cas d'étude dans l'évolution des méthodes et pratiques scientifiques dans les SHS ? » *Les Dossiers du Grihl*, 20 avril 2012. <http://journals.openedition.org/dossiersgrihl/5306>.
- . « La TEI et le XML ». In *Qu'est-ce que la Text Encoding Initiative ?* Marseille, France: OpenEdition Press, 2015. <http://books.openedition.org/oepr/1298>.
- Burt, Cyril. « Experimental tests of general intelligence ». *British Journal of Psychology* 3, n° 1 (1909): 94.
- Cabantous, Alain. « André Zysberg, Les galériens. Vies et destins de 60 000 forçats sur les galères de France (1680-1748) ». *Annales* 44, n° 6 (1989): 1544-47. *Cahiers du C.R.A.L.* Nancy, France: Université de Nancy 2, Centre de recherches et d'applications linguistiques, 1967.
- Califano, Elio. « Registrazione diretta e integrale dei documenti ». *Publications de l'École Française de Rome* 31, n° 1 (1977): 253-56.
- Callu, Jean-Pierre. « Ordre du jour et Allocution à la mémoire de M. Jean Schneider, membre de l'Académie ». *Comptes rendus des séances de l'Académie des Inscriptions et Belles-Lettres* 148, n° 2 (2004): 797-98.
- Campbell-Kelly, Martin. « Punched card machinery ». In *Encyclopedia of Computer Science*, édité par Edwin D. Reily, Anthony Raston et David Hemmendinger, 1489-91. Chichester, Royaume-Uni: Wiley, 2003.
- Carbonell, Charles-Olivier. « L'histoire dite « positiviste » en France ». *Romantisme* 8, n° 21 (1978): 173-85. <https://doi.org/10.3406/roman.1978.5216>.
- Carpentier, Élisabeth. « Le paysage rural du Contado d'Orvieto à la fin du XIII^{ème} siècle d'après le cadastre de 1292 ». *Actes des congrès de la Société des historiens médiévistes de l'enseignement supérieur public* 10, n° 1 (1979): 177-89. <https://doi.org/10.3406/shmes.1979.1335>.
- . « Cadastres d'Orvieto ». *Le médiéviste et l'ordinateur* 3, n° 1 (1980): 5-7. <https://doi.org/10.3406/medio.1980.1640>.
- . « Le paysage rural du Contado d'Orvieto à la fin du XIII^{ème} siècle d'après le cadastre de 1292 ». *Revue du Nord* 62, n° 244 (1980): 177-89. <https://doi.org/10.3406/rnord.1980.3669>.

- . « Histoire et informatique. Recherches sur le vocabulaire des biographies royales françaises ». *Cahiers de Civilisation Médiévale* 25, n° 97 (1982): 3-30. <https://doi.org/10.3406/ccmed.1982.2387>.
- . « Structures féodales et féodalisme dans l'Occident méditerranéen (X^e-XIII^e s.) ». *Cahiers de Civilisation Médiévale* 26, n° 102 (1983): 141-46. <https://doi.org/10.3406/ccmed.1983.2222>.
- . « L'homme, les hommes et la femme. Étude sur le vocabulaire des biographies royales françaises (XI^e-XIII^e siècle) ». *Annales* 41, n° 2 (1986): 325-46. <https://doi.org/10.3406/ahess.1986.283277>.
- . *Orvieto à la fin du XIII^e siècle : ville et campagne dans le cadastre de 1292*. Paris, France: Éditions du CNRS, 1986.
- Carr, David, éd., *La Philosophie de l'histoire et la pratique historique d'aujourd'hui*. Ottawa, Canada: Éditions de l'Université d'Ottawa, 1982.
- Carter-Thomas, Shirley, et Elizabeth Rowley-Jolivet. « Open science notebooks: New insights, new affordances ». *Journal of Pragmatics* 116 (2017): 64-76.
- Casini, Bruno. *Il catasto di Pisa del 1428-29*. Giardini, 1964.
- Castellani, Christian. « Le rôle économique de la communauté juive de Carpentras au début du XV^e siècle ». *Annales* 27, n° 3 (1972): 583-611. <https://doi.org/10.3406/ahess.1972.422525>.
- Ceruzzi, Paul E. *A history of modern computing*. Cambridge, États-Unis d'Amérique: MIT Press, 1998.
- Cetina, Karin Knorr. « Epistemic Cultures: Forms of Reason in Science ». *History of Political Economy* 23, n° 1 (1991): 105-22.
- Charbonneau, Hubert. « Louis Henry et la démographie historique ». *Population* 50, n° 6 (1995): 1663-71. <https://doi.org/10.2307/1534477>.
- Charle, Christophe, éd., *Histoire sociale, histoire globale ? : actes du colloque des 27-28 janvier 1989*. Paris, France: Maison des Sciences de l'Homme, 1993.
- Chartier, Yves. « Édition critique et micro-ordinateur ». *Le médiéviste et l'ordinateur* 16, n° 1 (1986): 5-9. <https://doi.org/10.3406/medio.1986.1138>.
- Chaunu, Pierre. « Histoire quantitative ou histoire sérielle ». *Cahiers Vilfredo Pareto* 2, n° 3 (1964): 165-76.

- . « Pour une histoire sérielle du Brésil au XVIII^e siècle : Réflexions préalables ». *Revue d'histoire économique et sociale* 49, n° 4 (1971): 465-84.
- Chemla, Karine. « Commentaires, éditions et autres textes seconds : quel enjeu pour l'histoire des mathématiques ? Réflexions inspirées par la note de Reviel Netz », *Revue d'histoire des mathématiques* 5, n° 2 (1999): 127--148.
- . « Literacy and the History of Science: Reflections Based on Chinese and Other Sources » In *The Cambridge Handbook of Literacy*, édité par David R. Olson et Nancy Torrance. Cambridge, États-Unis d'Amérique: Cambridge University Press, 2009.
https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_2305132_1.
- . « On the Diagrammaticity of Ancient Texts and Its Importance for the History of Science: Based on the Example of the Early Chinese Mathematical Text The Gnomon of the Zhou ». *East Asian Science, Technology and Society: An International Journal* 14, n° 2 (1 juin 2020): 279-308.
<https://doi.org/10.1215/18752160-8538529>.
- Chemla, Karine, et Evelyn Fox Keller, édés. *Cultures without culturalism: the making of scientific knowledge*. Durham, Royaume-Uni: Duke University Press, 2017.
- Chevallier, Raymond. « Photographie aérienne et histoire économique et sociale de l'Antiquité : un livre, une exposition ». *Annales* 13, n° 2 (1958): 322-26.
<https://doi.org/10.3406/ahess.1958.2742>.
- Cheyne, Jean-Claude. « À la maison : c'est Byzance ! ». *Le médiéviste et l'ordinateur* 9, n° 1 (1983): 7-8. <https://doi.org/10.3406/medio.1983.1657>.
- Cibois, Philippe, « Analyse des données et sociologie ». *L'Année sociologique* 31, n° 1 (1981): 333-48.
- . *L'analyse factorielle : analyse en composantes principales et analyse des correspondances*. Paris, France: Presses universitaires de France, 2000.
- . « L'utilisation de l'analyse des correspondances ». *Le médiéviste et l'ordinateur* 5, n° 1 (1981): 16-18. <https://doi.org/10.3406/medio.1981.954>.
- Cloulas, Ivan. « L'inventaire automatisé des actes notariés : principes d'analyse et résultats d'expérimentation aux Archives Nationales de Paris ». *Publications de l'École Française de Rome* 31, n° 1 (1977): 127-31.

- Clubb, Jerome M. « The Inter-University Consortium for Political Research: Progress and Prospects ». *Historical Methods Newsletter* 2, n° 3 (1 juin 1969): 1-5. <https://doi.org/10.1080/00182494.1969.10593887>.
- Clubb, Jerome M., et Howard Allen. « Computers and Historical Studies ». *The Journal of American History* 54, n° 3 (1967): 599-607. <https://doi.org/10.2307/2937409>.
- Combe, Sonia. « L'historien face à l'ordre informatique. Classification et Histoire ». *Matériaux pour l'histoire de notre temps* N° 82, n° 2 (2006): 1-3.
- « Conclusion générale ». *Publications de l'École Française de Rome* 31, n° 1 (1977): 423-31.
- Conti, Elio. *La formazione della struttura agraria moderna nel contado fiorentino*. Rome, Italie: Istituto storico italiano per il Medio Evo, 1965.
- Coq, Dominique, et Ezio Ornato. « Les séquences de composition du texte dans la typographie du XVe siècle. Une méthode quantitative d'identification ». *Histoire & Mesure* 2, n° 1 (1987): 87-136. <https://doi.org/10.3406/hism.1987.1305>.
- Couturier, Marcel. « Démographie historique et mécanographie électronique ». *Annales de Démographie Historique* 1966, n° 1 (1967): 57-78. <https://doi.org/10.3406/adh.1967.922>.
- . « Vers une nouvelle méthodologie mécanographique. La préparation des données ». *Annales* 21, n° 4 (1966): 769-78. <https://doi.org/10.3406/ahess.1966.421421>.
- Créange, Marion. « Par et pour la recherche d'images : EXPRIM ». *Le médiéviste et l'ordinateur* 19, n° 1 (1988): 15-21. <https://doi.org/10.3406/medio.1988.1758>.
- Créange, Marion, et Lucie Fossier. « Essai d'exploitation sur ordinateur des sources diplomatiques médiévales ». *Annales* 25, n° 1 (1970): 249-84. <https://doi.org/10.3406/ahess.1970.422211>.
- Créange, Marion, et Marie-Christine Haton. « L'informatique universitaire à Nancy : un demi-siècle de développement ». *Technique et science informatiques* 33, n° 1-2 (2014): 127-41.
- Crozet, René, et Edmond-René Labande. « Liminaire ». *Cahiers de Civilisation Médiévale* 1, n° 1 (1958): 5-7.
- Curry, Niall. « Academic Writing and Reader Engagement: Contrasting Questions in English, French and Spanish Corpora ». PhD dissertation, University of Limerick,

2019. <https://www.routledge.com/Academic-Writing-and-Reader-Engagement-Contrasting-Questions-in-English/Curry/p/book/9781032011134>.
- Curti, Merle Eugene, Robert R. Daniel, Shaw Livermore, et Joseph Van Hise. *The making of an american community: a case study of democracy in a frontier country*. Stanford, États-Unis d'Amérique: Stanford University press, 1959.
- Dagnelie, Pierre. « L'ordinateur et l'enseignement de la statistique ». *Revue de Statistique Appliquée* 19, n° 1 (1971): 5-15.
- Daubeze, Yvette, et Jean-Claude Perrot. « Un programme d'étude démographique sur ordinateur ». *Annales* 27, n° 4 (1972): 1047-70. <https://doi.org/10.3406/ahess.1972.422580>.
- Daumard, Adeline, et François Furet. « Méthodes de l'Histoire sociale : les Archives notariales et la Mécanographie ». *Annales* 14, n° 4 (1959): 676-93. <https://doi.org/10.3406/ahess.1959.2865>.
- David, M. « Aux sources de l'histoire sociale ». *Revue historique de droit français et étranger (1922-)* 46 (1968): 32-62.
- Delalande, Nicolas, et Julien Vincent. « Portrait de l'historien-ne en cyborg ». *Revue d'histoire moderne contemporaine* n° 58-4bis, n° 5 (2011): 5-29.
- De Mautort, Laurent. « L'industrie des micro-ordinateurs (les principales lignes de force) ». *Revue d'économie industrielle* 38, n° 1 (1986): 64-73. <https://doi.org/10.3406/rei.1986.2198>.
- De Roover, Raymond. *The rise and decline of the Medici bank: 1397-1494*. Cambridge, États-Unis d'Amérique: Harvard University Press, 1963.
- Delacroix, Christian, François Dosse, Patrick Garcia, et Nicolas Offenstadt. *Historiographies : concepts et débats*. Paris, France: Gallimard, 2010.
- Delatte, Louis, et Étienne Évrard. « Un laboratoire d'analyse statistique des langues anciennes à l'Université de Liège ». *L'Antiquité Classique* 30, n° 2 (1961): 429-44. <https://doi.org/10.3406/antiq.1961.3417>.
- Delumeau, Jean. « Méthode mécanographique et trafic maritime : les terre-neuviens malouins à la fin du XVII^e siècle ». *Annales* 16, n° 4 (1961): 665-85. <https://doi.org/10.3406/ahess.1961.421718>.

- Demonet, Michel, Annie Geffroy, et Jean Gouazé. *Des tracts en mai 68 : mesures de vocabulaire et de contenu*. Paris, France: Fondation nationale des sciences politiques, Armand Colin, 1975.
- Demonet, Michel. « Tableau de l'agriculture française au milieu du XIX^e siècle : l'enquête de 1852 ». Thèse de doctorat d'État, Université Paris 1 Panthéon-Sorbonne, 1985.
- Desrosières, Alain, et Laurent Thévenot. « Les mots et les chiffres : les nomenclatures socio-professionnelles ». *Économie et Statistique* 110, n° 1 (1979): 49-65. <https://doi.org/10.3406/estat.1979.4260>.
- Desrosières, Alain. « L'histoire de la statistique comme genre : style d'écriture et usages sociaux ». *Genèses. Sciences sociales et histoire* 39, n° 1 (2000): 121-37. <https://doi.org/10.3406/genes.2000.1626>.
- « Directory of Scholars Active ». *Computers and the Humanities* 2, n° 2 (1967): 71-93.
- Dosse, François. « À l'école des Annales, une règle : l'ouverture disciplinaire ». *Hermès, La Revue* n° 67, n° 3 (2013): 106-12.
- Dougherty, Jack, et Kristen Nawrotzki, eds., *Writing history in the digital age*. Ann Arbor, États-Unis d'Amérique: University of Michigan Press, 2013.
- Dubois, Henri. « D. Herlihy et Ch. Klapisch-Zuber, Les Toscans et leurs familles, une étude du castato florentin de 1427, 1978 ». *Annales de Démographie Historique* 1979, n° 1 (1979): 457-66.
- Duby, Georges. « Dans la France du Nord-Ouest au XII^e siècle : les « jeunes » dans la société aristocratique ». *Annales* 19, n° 5 (1964): 835-46. <https://doi.org/10.3406/ahess.1964.421226>.
- . « Histoire des mentalités ». In *L'histoire et ses méthodes*, édité par Charles Samaran, 937-65. Paris, France : Gallimard, 1961.
- . « La diffusion du titre chevaleresque sur le versant méditerranéen de la chrétienté latine ». In *La Noblesse au moyen âge XI^e-XV^e siècles. Essai à la mémoire de Robert Boutruche*, édité par Philippe Contamine, 42-52. Paris, France : Presses Universitaires de France, 1976.
- . *La société aux XI^e et XII^e siècles dans la région mâconnaise*. Paris, France: Armand Colin, 1953.

- . « Présentation de l'enquête sur « Famille et sexualité au Moyen Âge » ». *Publications de l'École Française de Rome* 30, n° 1 (1977): 9-11.
- Dumont, Léo, Octave Julien, et Stéphane Lamassé. « Articuler histoire et informatique, enseignement et recherche : le PIREH de l'université Panthéon-Sorbonne ». *Humanités numériques*, n° 1 (1 janvier 2020). <https://doi.org/10.4000/revuehn.284>.
- Dumoulin, Olivier. « La tribu des médiévistes ». *Genèses. Sciences sociales et histoire* 21, n° 1 (1995): 120-33. <https://doi.org/10.3406/genes.1995.1328>.
- Dupâquier, Jacques. « De l'animal à l'homme : le mécanisme autorégulateur des populations traditionnelles ». *Revue de l'Institut de Sociologie*, n° 2 (1972): 177-211.
- . « Herlihy D., Klapisch-Zuber C. — Les Toscans et leurs familles. Une étude du Catasto Florentin de 1427 ». *Population* 34, n° 1 (1979): 213-16.
- Duval, Julien. « L'analyse des correspondances et la construction des champs ». *Actes de la recherche en sciences sociales* N° 200, n° 5 (2013): 110-23.
- « Éditorial ». *Histoire & Mesure* 1, n° 1 (1986): 5-6.
- Epstein, Steven A. *Genoa and the Genoese, 958-1528*. Chapel Hill, États-Unis d'Amérique: University of North Carolina Press, 1996.
- Escofier, Brigitte. « L'analyse factorielle des correspondances ». Thèse de doctorat de 3^e cycle, Université de Rennes, France, 1965.
- Evrard, Étienne. « Deux programmes d'ordinateur pour l'étude quantitative du vocabulaire ». *Revue de l'Organisation Internationale pour l'Étude des Langues Anciennes par Ordinateur*, n° 2 (1967): 81-96.
- « Famille et parenté dans l'Occident médiéval. Actes du colloque de Paris (6-8 juin 1974) » 30, n° 1 (1977). https://www.persee.fr/issue/efr_0000-0000_1977_act_30_1.
- Febvre, Lucien, Ernest Tonnelat, Alfredo Niceforo, et Louis Weber. *Civilisation : le mot et l'idée*. Paris, France: Félix Alcan, 1930.
- Febvre, Lucien. *Le problème de l'incroyance au XVI^e siècle : la religion de Rabelais*. Paris, France: Éditions Albin Michel, 1942.

- Federico, Giovanni, et Paolo Malanima. « Progress, Decline, Growth: Product and Productivity in Italian Agriculture, 1000-2000 ». *The Economic History Review* 57, n° 3 (2004): 437-64.
- Feller, Jean. « « Que sais-je ? » : une collection qui a réponse à tout ». *Communication & Langages* 28, n° 1 (1975): 71-75. <https://doi.org/10.3406/colan.1975.4251>.
- Fénelon, Jean-Pierre. « Justification mathématique des correspondances ». *Le médiéviste et l'ordinateur* 5, n° 1 (1981): 15-16. <https://doi.org/10.3406/medio.1981.953>.
- . *Qu'est-ce que l'analyse des données ? : exposé accessible aux non-mathématiciens*. Paris, France: Lefonen, 1981.
- Fiumi, Enrico. « Stato di popolazione e distribuzione della ricchezza in Prato secondo il catasto del 1428-29 ». *Archivio Storico Italiano* 123, n° 3 (447) (1965): 277-303.
- Flori, Jean. « La notion de chevalerie dans les chansons de geste du XII^e siècle : étude historique du vocabulaire », *Le Moyen-Âge* 81, n° 1, (1975) : 211-244.
- Fløttum, Kjersti, Trine Dahl, et Torodd Kinn. *Academic Voices*. Amsterdam, Pays-Bas: John Benjamins Publishing Company, 2006. <https://benjamins.com/catalog/pbns.148>.
- Fogel, Robert William. « The Limits of Quantitative Methods in History ». *The American Historical Review* 80, n° 2 (1975): 329-50. <https://doi.org/10.2307/1850498>.
- Fossier, Lucie. « L'utilisation de l'informatique pour l'exploitation des documents textuels médiévaux ». *Mélanges de l'école française de Rome* 87, n° 2 (1975): 617-18. <https://doi.org/10.3406/mefr.1975.2341>.
- . « À propos d'un formulaire d'enquête relatif à "L'information sur les recours aux ordinateurs" ». *Francia* 4 (1976): 719-21.
- . « Informatique et documents médiévaux ». *Annales* 31, n° 6 (1976): 1131-42. <https://doi.org/10.3406/ahess.1976.293778>.
- . « Les langages documentaires ». *Publications de l'École Française de Rome* 31, n° 1 (1977): 351-59.
- . « De la carte perforée au texte numérisé ». *Le médiéviste et l'ordinateur* 12, n° 1 (1984): 1-2.
- . « Éditorial ». *Le médiéviste et l'ordinateur* 20, n° 1 (1988): 1-3.
- , éd., *Le médiéviste et l'ordinateur : actes de la table ronde, Paris, CNRS, 17 novembre 1989*. Paris, France: Institut de Recherche et d'histoire des textes, 1990.

- Fossier, Lucie, André Vauchez, et Cinzio Violante. « Avant-propos ». *Publications de l'École Française de Rome* 31, n°1 (1977): 5-9.
- Fossier, Lucie, Élisabeth Carpentier, Jean-Philippe Genet, Hélène Millet, et Michel Parisse. « Vingt ans d'informatique en histoire médiévale ». *Actes des congrès de la Société des historiens médiévistes de l'enseignement supérieur public* 20, n° 1 (1989): 501-25. <https://doi.org/10.3406/shmes.1989.1523>.
- Fossier, Lucie, et Gian Piero Zarrì. *L'indexation automatique des sources documentaires anciennes*. Paris, France: Éditions du CNRS, 1975.
- Fossier, Lucie, et Jean-Philippe Genet. « Un type d'exploitation : le traitement de texte ». *Le médiéviste et l'ordinateur*1, n° 1 (1979): 2-9. <https://doi.org/10.3406/medio.1979.881>.
- Fossier, Robert. « Censiers et terriers ». *Publications de l'École Française de Rome* 31, n° 1 (1977): 169-74.
- . « Peuplement de la France du Nord entre le X^e et le XVI^e siècles ». *Annales de démographie historique*, 1979, 59-99.
- Foucault, Michel. *L'archéologie du savoir*. Paris, France: Gallimard, 1969.
- Fourquin, Guy. « Yves Renouard (1908-1965) ». *Revue du Nord* 47, n° 184 (1965): 167-69.
- Furet, François. « Histoire quantitative et construction du fait historique ». *Annales* 26, n° 1 (1971): 63-75. <https://doi.org/10.3406/ahess.1971.422459>.
- Galland, Bruno. « La publication des registres de lettres pontificales par l'École française de Rome. » *Bibliothèque de l'École des chartes* 154, n° 2 (1996): 625-34. <https://doi.org/10.3406/bec.1996.450837>.
- Ganshof, François-Louis. *Qu'est-ce que la féodalité ?* Bruxelles, Belgique: J. Lebègue, 1944.
- Gardin, Jean-Claude. « Présentation ». *Annales* 18, n° 2 (1963): 259-63.
- Gardin, Jean-Claude, et Paul Garelli. « Étude par ordinateurs des établissements assyriens en Cappadoce ». *Annales* 16, n° 5 (1961): 837-76. <https://doi.org/10.3406/ahess.1961.420758>.
- Gauthier, Etienne, et Louis Henry. « Crulai. Démographie d'une paroisse normande au XVII^e et au XVIII^e siècles. Présentation d'un cahier de l'I.N.E.D ». *Population* 13, n° 2 (1958): 283-86. <https://doi.org/10.2307/1524822>.

- Gauvard, Claude. « Christine de Pisan a-t-elle eu une pensée politique? À propos d'ouvrages récents ». *Revue Historique* 250, n° 2 (508) (1973): 417-30.
- Geffroy, Annie. « André Salem, « Méthodes de la statistique textuelle », thèse de doctorat d'État, Université Paris 3, 1993 ». *Mots. Les langages du politique* 38, n° 1 (1994): 104-8.
- Genet, Jean-Philippe. *Les idées sociales de Sir John Fortescue*. Paris, France: Publications de la Sorbonne, 1973.
- , éd., *Four English political tracts of the later Middle Ages*. London, Royaume-Uni: Offices of the Royal Historical Society, 1977.
- . « Histoire sociale et ordinateur ». *Publications de l'École Française de Rome* 31, n° 1 (1977): 231-37.
- . « Ordinateur, lexique, contexte ». *Publications de l'École Française de Rome* 31, n° 1 (1977): 297-317.
- . « David Herlihy et Christiane Klapisch-Zuber, Les Toscans et leurs familles. Une étude du catasto florentin de 1427, 1978 ». *Francia* 7 (1979): 705-11. <https://doi.org/10.11588/fr.1979.0.49376>.
- . « Éditorial ». *Le médiéviste et l'ordinateur* 5, n° 1 (1981): 1-2.
- . « Une application de l'analyse factorielle à l'étude du vocabulaire ». *Le médiéviste et l'ordinateur* 5, n° 1 (1981): 11-15. <https://doi.org/10.3406/medio.1981.952>.
- . « Éditorial ». *Le médiéviste et l'ordinateur* 11, n° 1 (1984): 1-2.
- . « Automatic text processing and factorial analysis: a method for determining the lexicographical horizon of expectation ». In *Computer Applications to Medieval Studies, Kalamazoo, Western Michigan University*, édité par Anne Gilmour-Bryson, 147-75. Kalamazoo, États-Unis d'Amérique: Medieval Institute Publications, Western Michigan University, 1984.
- . « Histoire, Informatique, Mesure ». *Histoire & Mesure* 1, n° 1 (1986): 7-18. <https://doi.org/10.3406/hism.1986.904>.
- . « L'informatique au service de la prosopographie : PROSOP ». *Mélanges de l'école française de Rome* 100, n°1 (1988): 247-63. <https://doi.org/10.3406/mefr.1988.2972>.

- . « La formation des historiens à l'informatique en France : espoir ou désespoir ? » *Le médiéviste et l'ordinateur* 31, n° 1 (1995): 32-36. <https://doi.org/10.3406/medio.1995.1423>.
- . « Les idées sociales et politiques en Angleterre : du début du XIV^e siècle au milieu du XVI^e siècle ». Thèse de doctorat d'État, Université Paris 1 Panthéon-Sorbonne, 1997.
- . *La genèse de l'État moderne : culture et société politique en Angleterre*. Paris, France: Presses Universitaires de France, 2003.
- Genet, Jean-Philippe, Francois Hucher, Jacques Mondelli, et Edouard Valensi. « Un programme de traitement automatique des textes : ALINE ». *Bulletin du centre d'analyse du discours de l'université de Lille III* (1974): 96-121.
- Genet, Jean-Philippe, Andrea Zorzi, eds., *Les historiens et l'informatique : un métier à réinventer*. Rome, Italie: École française de Rome, 2011.
- Genicot, Léopold, éd., *Typologie des sources du Moyen Âge occidental*. Turnhout, Belgique: Brepols, 1972.
- . « La typologie des sources du Moyen Âge occidental ». *Annales* 27, n° 6 (1972): 1257-63. <https://doi.org/10.3406/ahess.1972.422543>.
- . « Pour une organisation de la recherche en histoire médiévale ». *Francia* 1 (1973): 692-98.
- . « Index Scriptorum Operumque Latino-Belgicorum Medii Aevi. Nouveau répertoire des œuvres médiolatines belges publié sous la direction de L. Genicot et P. Tombeur. Première partie : VII^e-X^e siècles par A. Stainier (= Académie royale de Belgique. Comité national du dictionnaire du latin médiéval), Bruxelles, 1973 ». *Bulletins de l'Académie Royale de Belgique* 60, n° 1 (1974): 29-30.
- . « Pour une organisation de la recherche en histoire médiévale, II. L'information sur les recours aux ordinateurs ». *Francia* 3 (1975): 643-47.
- . « Le traitement électronique des textes diplomatiques belges antérieurs à 1200 ». *Publications de l'École Française de Rome* 31, n° 1 (1977): 97-104.
- Ginzburg, Carlo. *Il Formaggio e i vermi: il cosmo di un mugnaio del'500*. Torino, Italie: G. Einaudi, 1976.
- Gjesdal, Anje Müller. « The Influence of Genre Constraints on Author Representation in Medical Research Articles. The French Indefinite Pronoun On in IMRAD

- Research Articles ». *Discours. Revue de Linguistique, Psycholinguistique et Informatique. A Journal of Linguistics, Psycholinguistics and Computational Linguistics*, n° 12 (10 juillet 2013). <https://doi.org/10.4000/discours.8770>.
- Glénisson, Jean, et Edmond-René Labande. « Formation et destin de l'Institut de Recherche et d'Histoire des Textes [Réponses de Jean Glénisson, directeur de l'I.R.H.T., aux questions posées par E.R. Labande au cours d'un entretien, le 11 juin 1971.] ». *Cahiers de Civilisation Médiévale* 15, n° 57 (1972): 53-60. <https://doi.org/10.3406/ccmed.1972.2023>.
- Glénisson, Jean. « Prosopographie et informatique ». *Publications de l'École Française de Rome* 31, n° 1 (1977): 227-29.
- Glorieux, Frédéric. « Pourquoi informatiser un vieux glossaire ? » *ELA. Études de linguistique appliquée* 156, n° 4 (2009): 417-28.
- Goldthwaite, Richard A. « Les Maîtres Du Marbre : Carrare, 1300–1600. By Christiane Klapisch-Zuber. Paris, S.E.V.P.E.N., 1969. Pp. 364. 43F. » *Business History Review* 45, n° 1 (ed 1971): 93-94. <https://doi.org/10.2307/3113310>.
- Goody, Jack. *La raison graphique. La domestication de la pensée sauvage*. Paris, France: Les Editions de Minuit, 1979.
- . *The domestication of the savage mind*. Cambridge, Royaume-Uni: Cambridge University Press, 1977.
- Goubert, Pierre, et Guy Arbellot (directeur). « Beauvais et le Beauvaisis de 1600 à 1730 : contribution à l'histoire sociale de la France du XVII^e siècle ». Thèse de doctorat d'État, Université de Paris, 1958.
- Greimas, Algirdas Julien. « Histoire et Linguistique ». *Annales* 13, n° 1 (1958): 110-14. <https://doi.org/10.3406/ahess.1958.2715>.
- Grenier, Jean-Yves, et Bernard Lepetit. « L'expérience historique. À propos de C.-E. Labrousse ». *Annales* 44, n° 6 (1989): 1337-60. <https://doi.org/10.3406/ahess.1989.283659>.
- Grossetti, Michel, et Pierre-Eric Mounier-Kuhn. « Les débuts de l'informatique dans les universités ». *Revue française de sociologie* XXXVI, n°2 (1995): 295-324.
- Grote, Mathias. « Total Knowledge? Encyclopedic Handbooks in the Twentieth-Century Chemical and Life Sciences ». *BJHS Themes* 5 (2020): 187-203. <https://doi.org/10.1017/bjt.2020.11>.

- Guerreau, Alain. « Analyse factorielle au moyen d'une calculatrice de poche programmable ». *Le médiéviste et l'ordinateur* 2, n° 1 (1979): 10-12. <https://doi.org/10.3406/medio.1979.905>.
- . *Le Féodalisme, un horizon théorique*. Paris, France: le Sycomore, 1980.
- . « Analyse factorielle et analyses statistiques classiques : le cas des ordres mendiants dans la France médiévale ». *Annales* 36, n° 5 (1981): 869-912. <https://doi.org/10.3406/ahess.1981.282792>.
- . « Le Haut Moyen Âge factorialisé ». *Le médiéviste et l'ordinateur* 5, n° 1 (1981): 7-11. <https://doi.org/10.3406/medio.1981.951>.
- Guilhaumou, Jacques. « Le corpus en analyse de discours : perspective historique ». *Corpus*, n° 1 (15 novembre 2002). <https://doi.org/10.4000/corpus.8>.
- . « L'idéologie du Père Duchesne : les forces adjuvantes (14 juillet-6 septembre 1793) ». *Le Mouvement social*, n° 85 (1973): 81-116. <https://doi.org/10.2307/3806994>.
- Guillaumont, Antoine. « Henri Irénée Marrou. *De la connaissance historique* ». *Revue de l'histoire des religions* 149, n°1 (1956): 115-16.
- Guiral, Jacqueline. « II. Informatique et péages du royaume de Valence (1494) ». *Mélanges de la Casa de Velázquez* 12, n° 1 (1976): 146-63. <https://doi.org/10.3406/casa.1976.2791>.
- Gutmann, Myron P. « Demographic and Economic Issues in Les Toscans et Leurs Familles ». *The Journal of Interdisciplinary History* 11, n° 3 (1981): 487-93. <https://doi.org/10.2307/203630>.
- Guyotjeannin, Olivier. « La diplomatie médiévale et l'élargissement de son champ ». *Gazette des archives* 172, n° 1 (1996): 12-18. <https://doi.org/10.3406/gazar.1996.3386>.
- . « Les sources de l'histoire économique et sociale du Moyen Âge par Robert-Henri Bautier et Janine Soknay, deuxième série. Les États de la maison de Bourgogne, vol I, Archives centrales de l'État bourguignon (1384-1500), Archives des principautés territoriales - Les principautés du sud - Les principautés du nord supplément) comtés de Hollande et Zélande et duché de Gueidre par Michel van Gent mise à jour du fasc. additions et corrections ». *Bibliothèque de l'École des chartes* 160, n° 1 (2002): 343-44.

- . *Les sources de l'histoire médiévale*. Paris, France: Librairie générale française, 1998.
- Hackney, Sheldon. « Power to the computers: a revolution in history? » In *Proceedings of the May 5-7, 1970, spring joint computer conference*, 275-79. AFIPS '70 (Spring). New York, NY, USA: Association for Computing Machinery, 1970. <https://doi.org/10.1145/1476936.1476982>.
- Haffner, Emmylou, 2018. « L'édition des oeuvres mathématiques au XIX^e siècle en Allemagne. l'exemple des *Gesammelte Werke und wissenschaftlicher Nachlass* de Bernhard Riemann », *Philosophia Scientiæ* 22, n° 2 (2018): 115-135.
- Haigh, Thomas, Mark Priestley, et Crispin Rope. *ENIAC in action: making and remaking the modern computer*. Cambridge, États-Unis d'Amérique: The MIT Press, 2016.
- Hainsworth, Michael. « Dans un centre de calcul : le LISH ». *Le médiéviste et l'ordinateur* 9, n° 1 (1983): 3-5. <https://doi.org/10.3406/medio.1983.1019>.
- Hajnal, John. « Age at Marriage and Proportions Marrying ». *Population Studies* 7, n° 2 (1953): 111-36. <https://doi.org/10.2307/2172028>.
- . « European Marriage Patterns in Perspective. » In *Population in History*, édité par D. V. Glass et D. E. C. Eversley, 101-43. Londres, Royaume-Uni: Routledge, 1965. <https://doi.org/10.4324/9781315127019-7>.
- . « Two Kinds of Preindustrial Household Formation System ». *Population and Development Review* 8, n° 3 (1982): 449-94. <https://doi.org/10.2307/1972376>.
- Haroche, Claudine, Paul Henry, et Michel Pêcheux. « La sémantique et la coupure saussurienne : langue, langage, discours ». *Langages* 6, n° 24 (1971): 93-106. <https://doi.org/10.3406/lgge.1971.2608>.
- Heffer, Jean, Jean-Louis Robert, et Pierre Saly. *Outils statistiques pour les historiens*. Paris, France: Publications de la Sorbonne, 1981.
- Heide, Lars. *Punched-card systems and the early information explosion, 1880-1945*. Baltimore, États-Unis d'Amérique: Johns Hopkins University Press, 2009.
- Heimburger, Franziska, et Émilien Ruiz. « Faire de l'histoire à l'ère numérique : retours d'expériences ». *Revue d'histoire moderne contemporaine* 58-4bis, n° 5 (2011): 70-89.
- Heimendinger, Vincent. « Genèses en train de se faire : écrire l'histoire d'une revue scientifique ». *Zilsel* 5, n° 1 (2019): 271-314.

- Henry, Louis, et Paul Vincent. « Rythme maximum d'accroissement d'une population stable ». *Population* 2, n° 4 (1947): 663-80. <https://doi.org/10.2307/1524112>.
- Henry, Louis. « Simulation d'une reconstitution de familles par ordinateur ». *Annales de Démographie Historique* 1972, n° 1 (1972): 303-9. <https://doi.org/10.3406/adh.1972.2130>.
- Herlihy, David. *Medieval and Renaissance Pistoia: The Social History of an Italian Town, 1200-1430*. New Haven, États-Unis d'Amérique: Yale University Press, 1967.
- . « Vieillir au Quattrocento ». *Annales* 24, n° 6 (1969): 1338-52. <https://doi.org/10.3406/ahess.1969.422170>.
- . « Some Psychological and Social Roots of Violence in the Tuscan Cities ». In *Violence and Civil Disorder in Italian Cities, 1200-1500*, édité par Lauro Martines, 129-55. Berkeley, États-Unis d'Amérique: University of California Press, 1972.
- . « The American Medievalist: A Social and Professional Profile ». *Speculum* 58, n° 4 (octobre 1983): 881-90. <https://doi.org/10.2307/2853788>.
- Herlihy, David, et Christiane Klapisch-Zuber. *Les Toscans et leurs familles : Une étude du catasto florentin de 1427*. Paris, France: Fondation nationale des sciences politiques, 1978.
- . *Tuscans and Their Families: A Study of the Florentine Catasto of 1427*. New Haven, États-Unis d'Amérique: Yale Univ. Press, 1985.
- Herlihy, David, Christiane Klapisch-Zuber, et Mario Bensi. *I toscani e le loro famiglie: uno studio sul catasto fiorentino del 1427*. Bologne, Italie: Mulino, 1988.
- Higounet, Charles, et Charles Samaran (préfacier). *Paysages et villages neufs du Moyen âge : recueil d'articles*. Bordeaux, France: Fédération historique du Sud-Ouest, 1975.
- Hockey, Susan. « The History of Humanities Computing ». In *A Companion to Digital Humanities*, édité par Susan Schreibman, Ray Siemens et John Unsworth, 1-19. Oxford, Royaume-Uni: Blackwell, 2004. <https://doi.org/10.1002/9780470999875.ch1>.
- Holtz, Louis. « L'institut de recherche et d'histoire des textes (IRHT). Premier laboratoire d'histoire au Centre National de la Recherche Scientifique ». *Les Cahiers du*

- Centre de Recherches Historiques. Archives*, n° 36 (30 juin 2005).
<https://doi.org/10.4000/ccrh.3046>.
- Hotelling, Harold. « Analysis of a complex of statistical variables into principal components. » *Journal of educational psychology* 24, n° 6 (1933): 417.
- Huard, Raymond, et Philippe Gut. « Louis Girard 1911-2003 ». *Revue d'histoire du XIX^e siècle* 26/27, (1 décembre 2003): 445-46.
- Hyland, Ken. « What Do They Mean? Questions in Academic Writing ». *Text - Interdisciplinary Journal for the Study of Discourse* 22, n° 4 (2002).
https://www.academia.edu/7567243/What_do_they_mean_Questions_in_academic_writing.
- Hymes, Dell. *Foundations in Sociolinguistics: An Ethnographic Approach*. Philadelphie, États-Unis d'Amérique: University of Pennsylvania Press, 1974.
- Institut d'histoire moderne et contemporaine et Institut d'histoire du temps présent, édés.,
Table ronde « Histoire et informatique », École normale supérieure, 8 mars 1980.
 Paris, France: Éditions du CNRS, 1980.
- Jacquart, Jean. « M. Couturier, Recherches sur les structures sociales de Châteaudun : 1525-1789 ». *Annales de Démographie Historique* 1970, n° 1 (1971): 410-13.
- Jauss, Hans Robert, Claude Maillard, et Jean Starobinski, édés., *Pour une esthétique de la réception*. Paris, France: Gallimard, 1978.
- Jauss, Hans Robert. *Literaturgeschichte als Provokation der Literaturwissenschaft*. Konstanz, Allemagne: Im Verlag der Druckerei und Verlagsanstalt Konstanz, Universitätsverlag, 1967.
- Johrendt, Johann. *"Milites" und "Militia" im elften Jahrhundert: Untersuchung zur Frühgeschichte des Rittertums in Frankreich und Deutschland*. Erlangen, Allemagne : Universität Erlangen-Nürnberg, 1971.
- Jones, Steven Edward. *Roberto Busa, S. J., and the emergence of humanities computing: the priest and the punched cards*. New York, États-Unis d'Amérique : Routledge, 2016.
- Jónsson, Einar Már. « Les « miroirs aux princes » sont-ils un genre littéraire ? » *Médiévales. Langues, Textes, Histoire*, n° 51 (1 décembre 2006): 153-66. <https://doi.org/10.4000/medievales.1461>.

- Joris, André. « Léopold Genicot (1914-1993) ». *Cahiers de Civilisation Médiévale* 39, n° 153 (1996): 173-75.
- Kansteiner, Wulf. « Hayden White's Critique of the Writing of History ». *History and Theory* 32, n° 3 (1993): 273-95. <https://doi.org/10.2307/2505526>.
- Kaplan, Frédéric, Mélanie Fournier, et Marc-Antoine Nuessli. « L'historien et l'algorithme ». *Le Temps des Humanités Digitales. La mutation des sciences humaines et sociales*, édité par Olivier Le Deuff, 49-85. Limoges, France: FYP éditions, 2014.
- Kelty, Christopher M. « This Is Not an Article: Model Organism Newsletters and the Question of 'Open Science' ». *BioSocieties* 7, n° 2 (1 juin 2012): 140-68. <https://doi.org/10.1057/biosoc.2012.8>.
- Klapisch-Zuber, Christiane. « Carrare et ses marbres, 1300-1600 ». Thèse de doctorat de 3^e cycle, École Pratique des Hautes Études (VI^e section), 1966.
- . « Fiscalité et démographie en Toscane (1427-1430) ». *Annales* 24, n° 6 (1969): 1313-37. <https://doi.org/10.3406/ahess.1969.422169>.
- . « Sources et méthodes de la démographie médiévale, le Catasto Florentin de 1427-30 ». *Actes des congrès de la Société des historiens médiévistes de l'enseignement supérieur public* 1, n° 1 (1970): 53-61. <https://doi.org/10.3406/shmes.1970.1174>.
- . « A Correspondence Analysis of a XVth Century Census: The Florentine Catasto of 1427. » *Journal of European Economic History* 4, n° 2 (1975): 415-28.
- . « Exploitation démographique et anthroponymique du catasto florentin de 1427 ». *Publications de l'École Française de Rome* 31, n° 1 (1977): 215-23.
- . « Le catasto florentin et son environnement ». *Les Cahiers du Centre de Recherches Historiques. Archives*, n°14-15 (14 octobre 1995). <https://doi.org/10.4000/ccrh.2659>.
- Klapisch-Zuber, Christiane, Béatrice Marin, et Nicolas Veysset. « Le catasto florentin de 1427-1430 : présentation générale de l'enquête et du code ». *L'Atelier du Centre de recherches historiques. Revue électronique du CRH*, 30 novembre 2016. <http://journals.openedition.org/acrh/7458>.

- Klapisch-Zuber, Christiane, et Michel Demonet. « A uno pane e uno vino : la famille rurale toscane au début du XV^e siècle ». *Annales* 27, n° 4 (1972): 873-901. <https://doi.org/10.3406/ahess.1972.422571>.
- Kłoczowski, Jerzy. « Les pays de l'Est de la chrétienté occidentale dans les Archives du Vatican (période médiévale) ». *Publications de l'École Française de Rome* 31, n° 1 (1977): 37-41.
- Knorr-Cetina, Karin. *Epistemic cultures: how the sciences make knowledge*. Cambridge, États-Unis d'Amérique: Harvard University Press, 1999.
- Koyré, Alexandre. *Études d'histoire de la pensée scientifique*. Paris, France: Presses Universitaires de France, 1966.
- . « Les origines de la science moderne ». *Diogène*, n° 16 (1956): 14.
- Kuhn, Thomas Samuel. *The structure of scientific revolutions*. Chicago, États-Unis d'Amérique: The University of Chicago Press, 1962.
- de La Roncière, Charles-Marie. *Un changeur florentin du trecento*. S.E.V.P.E.N., 1973.
- Lachmann, Richard. *States and Power*. Cambridge, Royaume-Uni: Polity Press, 2010.
- Lassus, François. « Toponymie et microtoponymie ». *Le médiéviste et l'ordinateur* 8, n° 1 (1982): 4-8. <https://doi.org/10.3406/medio.1982.1002>.
- Launay, Michel. « Review of *La Scission syndicale de 1921. Essai de reconnaissance des formes*, par Jean-Louis Robert ». *Le Mouvement social*, n° 122 (1983): 125-28. <https://doi.org/10.2307/3777777>.
- Lauvernier, Julie. « Mettre en ordre les archives des départements : genèse et élaboration du cadre de classement des Archives départementales ». *Gazette des archives* 229, n° 1 (2013): 23-40. <https://doi.org/10.3406/gazar.2013.5187>.
- Le Coq, Aurélien. « Compte-rendu de la journée d'études « Microhistoire et pratiques historiennes. Échelles, acteurs, formes narratives », Marne la Vallée, 11 mai 2012 ». *Memini. Travaux et documents*, n° 15 (31 décembre 2011). <http://journals.openedition.org/memini/419>.
- Le Goff, Jacques. « Avant-propos ». *Publications de l'École Française de Rome* 30, n° 1 (1977): 5-5.
- Leblond, Hervé. « Recherches métrologiques sur des plans de bastides médiévales ». *Histoire & Mesure* 2, n° 3 (1987): 55-87. <https://doi.org/10.3406/hism.1987.1325>.

- Lefort, Jacques. « Analyse automatique des documents fiscaux byzantins ». *Publications de l'École Française de Rome* 31, n° 1 (1977): 277-89.
- « Le médiéviste et l'ordinateur, N°5, printemps 1981. Introduction à l'analyse factorielle. » 5, n° 1 (1981). https://www.persee.fr/issue/medio_0223-3843_1981_num_5_1.
- Lemercier, Claire. « Analyse de réseaux et histoire ». *Revue d'histoire moderne contemporaine* 52-2, n° 2 (2005): 88-112.
- Lemercier, Claire, et Claire Zalc. « II. Face aux sources : corpus et échantillons ». *Repères*, 2008, 19-33.
- . *Méthodes quantitatives pour l'historien*. Paris: La Découverte, 2010.
- Léon, Jacqueline. « Aux sources de la « Corpus Linguistics » : Firth et la London School ». *Langages* n° 171, n° 3 (2008): 12-33.
- . « Automatisation-mathématisation de la linguistique en France dans les années 1960. Un cas de réception externe ». *2ème Congrès Mondial de Linguistique Française*, 2010, 055.
- . *Histoire de l'automatisation des sciences du langage*. Lyon, France: ENS Éditions, 2015.
- . « The Statistical Studies of Vocabulary in the 1950-60s in France. Theoretical and Institutional Issues ». In *History of Quantitative linguistics in France*, édité par Jacqueline Léon et Sylvain Loiseau, 9-28. Lündenscheid, Allemagne: RAM-Verlag, 2016.
- . « Jean Dubois, un passeur ». *Linx*, n° 80 (juillet 2020). <https://doi.org/10.4000/linx.5561>.
- Lepetit, Bernard. « L'histoire quantitative : deux ou trois choses que je sais d'elle ». *Histoire & Mesure* 4, n° 3 (1989): 191-99. <https://doi.org/10.3406/hism.1989.1355>.
- éd., *Les formes de l'expérience : une autre histoire sociale*. Paris, France: Albin Michel, 1995.
- Lett, Didier, et Christiane Klapisch-Zuber. « Entretien avec Christiane Klapisch-Zuber ». *Genre & Histoire*, n° 19 (25 août 2017). <http://journals.openedition.org/genrehistoire/2719>.

- Litchfield, R. Burr. *Emergence of a bureaucracy: the Florentine patricians (1530-1790)*. Princeton, États-Unis d'Amérique: Princeton University Press, 1986.
- Loiseau, Sylvain, et Jacqueline Léon, eds., *History of quantitative linguistics in France*. Lündenscheid, Allemagne: RAM-Verlag, 2016. <https://halshs.archives-ouvertes.fr/halshs-01443700>.
- « L'utilisation de l'informatique par les historiens d'aujourd'hui : bilan et perspectives ». *Cahiers de la Méditerranée* 53, n° 1 (1996): 211-17.
- Luzzati, Michèle. « L'identification des individus dans les registres de baptême pisans ». *Le médiéviste et l'ordinateur* 4, n° 1 (1980): 3-6. <https://doi.org/10.3406/medio.1980.935>.
- Machiavelli, Niccolò. *Machiavelli: The Chief Works and Others, Vol. I*. Durham, États-Unis d'Amérique: Duke University Press, 1989.
- Magnani, Eliana. « Qu'est-ce qu'un corpus ? Compte-rendu de la journée d'études ». *Les carnets de l'IRHT*, 19 octobre 2017. <https://irht.hypotheses.org/3187>.
- Magnou-Nortier, Élisabeth. « La société laïque et l'Église dans la province ecclésiastique de Narbonne: zone cispyrénéenne, de la fin du VIII^e à la fin du XI^e siècle ». Toulouse, France: Association des publications de l'université de Toulouse-Le Mirail, 1974.
- Maingueneau, Dominique. « La situation d'énonciation entre langue et discours ». In *Dix ans de S.D.U.*, 197-210. Craiova, Roumanie: Editura Universitarai Craiova, 2004.
- Malanima, Paolo. « The Long Decline of a Leading Economy: GDP in Central and Northern Italy, 1300–1913 ». *European Review of Economic History* 15, n° 2 (août 2011): 169-219. <https://doi.org/10.1017/S136149161000016X>.
- Marcotte-Chénard, Sophie. « Le contextualisme de Quentin Skinner à l'épreuve du cas Machiavel ». *Methodos. Savoirs et textes*, n° 13 (23 avril 2013). <https://doi.org/10.4000/methodos.3168>.
- Margaret, Masterman. « Freeing the Mind: Articles and Letters from The Times Literary Supplement during March-June, 1962 », *Times*, 1962.
- Marie, Alain. « Histoire de la famille et outil informatique : la contribution de l'Église de Jésus-Christ des Saints des Derniers Jours ». *Histoire & Mesure* 4, n° 1 (1989): 63-79. <https://doi.org/10.3406/hism.1989.876>.

- Marrou, Henri-Irénée. *De la connaissance historique*. Paris, France: Éditions du Seuil, 1954.
- . *La civilisation de l'antiquité tardive*. Rome, Italie: Accademia nazionale dei Lincei, 1968.
- (réédition), *De la connaissance historique*. Paris, France: Éditions du Seuil, 1975.
- Martin, Olivier. « Aux origines des idées factorielles [Des théories aux méthodes statistiques] ». *Histoire & Mesure* 12, n° 3 (1997): 197-249. <https://doi.org/10.3406/hism.1997.1544>.
- Martines, Lauro. *The social world of the Florentine humanists, 1390-1460*. Princeton, États-Unis d'Amérique: Princeton University Press, 1963.
- « Matériaux pour l'histoire de notre temps 2006/2 (N° 82) ». Consulté le 5 janvier 2021. <https://www.cairn.info/revue-materiaux-pour-l-histoire-de-notre-temps-2006-2.htm>.
- Mayaffre, Damon. « Vers une herméneutique matérielle numérique. Corpus textuels, Logométrie et Langage politique ». Habilitation à Diriger des Recherches, Université Nice Sophia Antipolis, 2010. <https://tel.archives-ouvertes.fr/tel-00655380>.
- Mazon, Brigitte. *Aux origines de l'École des hautes études en sciences sociales : le rôle du mécénat américain (1920-1960)*. Paris, France: Éditions du Cerf, 1988.
- McCarty, Willard. « A Telescope for the Mind? » In *Debates in the Digital Humanities*, édité par Matthew K. Gold, 113-36. Minneapolis, États-Unis d'Amérique: University of Minnesota Press, 2012. <https://www.jstor.org/stable/10.5749/j.ctttv8hq.11>.
- McCracken, Daniel D. *A Simplified Guide to FORTRAN Programming*. New York, États-Unis d'Amérique: Wiley, 1974.
- Mémoire vive*. Paris, France: Association française histoire et informatique, 1989.
- Mercoureff, Wladimir. « Témoignage : l'informatique dans l'enseignement et au CNRS au début des années soixante-dix. » *Histoire de la recherche contemporaine. La revue du Comité pour l'histoire du CNRS* IV, n°2 (15 décembre 2015): 180-85. <https://doi.org/10.4000/hrc.1154>.

- Metman, Josette, et Bernadette Auzary. « Les archives judiciaires et leur traitement informatique ». *Publications de l'École Française de Rome* 31, n° 1 (1977): 361-77.
- Micheau, Françoise, et Jean-Pierre Molénat. « Création d'alphabets sur Macintosh ». *Le médiéviste et l'ordinateur* 16, n° 1 (1986): 13-18. <https://doi.org/10.3406/medio.1986.1140>.
- Millet, Hélène. « Les chanoines du chapitre cathédral de Laon (1272-1412) ». Thèse 3^e cycle, Université Paris 1 Panthéno-Sorbonne, 1977.
- . « La composition du chapitre cathédral de Laon au XIV^e siècle : résultats d'une analyse factorielle ». *Annales* 36, n° 1 (1981): 117-38. <https://doi.org/10.3406/ahess.1981.282719>.
- . « Pour l'analyse factorielle ». *Le médiéviste et l'ordinateur* 5, n° 1 (1981): 18-19. <https://doi.org/10.3406/medio.1981.955>.
- . « Une expérience : essai de classification des chanoines de Laon ». *Le médiéviste et l'ordinateur* 7, n° 1 (1982): 14-17. <https://doi.org/10.3406/medio.1982.986>.
- Minard, Philippe. « Éditorial. Les passeurs de frontières ». *Le Mouvement Social* 259, n° 2 (21 juillet 2017): 3-11.
- Minel, Jean-Luc. « Le micro-ordinateur est-il une panacée ? » *Le médiéviste et l'ordinateur* 9, n° 1 (1983): 2-3. <https://doi.org/10.3406/medio.1983.1017>.
- « Mittelalterforschung in Münster in Der Nachfolge Des Sonderforschungsbereichs 7 ». *Frühmittelalterliche Studien* 24, n° 1 (31 décembre 1990): 380-429. <https://doi.org/10.1515/9783110242225.380>.
- Mollat du Jourdin, Michel. « Comptes portuaires et informatique ». *Publications de l'École Française de Rome* 31, n° 1 (1977): 149-56.
- Mornet, Élisabeth. « Standardisation et échange des bases de données historiques. Actes de la troisième Table Ronde internationale tenue au L.I.S.H. (C.N.R.S.), Paris, 15-16 mai 1987. Jean-Philippe Genet éd.- Paris, C.N.R.S., 1988 ». *Le médiéviste et l'ordinateur* 20, n° 1 (1988): 4-5.
- Mounier-Kuhn, Pierre. *L'informatique en France de la seconde guerre mondiale au Plan Calcul: l'émergence d'une science*. Paris, France: Presses de l'université Paris-Sorbonne, 2010.

- Muir, Edward. « New Light on Old Numbers: The Political and Cultural Implications of Les Toscans et Leurs Familles ». *The Journal of Interdisciplinary History* 11, n° 3 (1981): 477-85. <https://doi.org/10.2307/203629>.
- Muller, Charles. *Initiation à la statistique linguistique*. Paris, France: Librairie Larousse, 1968.
- Mulon, Marianne. « Une réalisation expérimentale du fichier automatisé des formes anciennes. Le répertoire de l'Ain ». *Publications de l'École Française de Rome* 31, n° 1 (1977): 199-204.
- Murphy, George G. S. « Historical Investigation and Automatic Data Processing Equipment ». *Computers and the Humanities* 3, n° 1 (1 septembre 1968): 1-13. <https://doi.org/10.1007/BF02395443>.
- Naïs, Hélène. « Présentation des travaux du Centre de recherche et d'applications linguistiques (CRAL) ». *Revue du LASLA*, n° 1 (1970): 71-79.
- Naïs, Hélène, et Jean Schneider. « L'atelier Vincent de Beauvais ». *Revue d'Histoire des Textes* 4, n° 1974 (1975): 438-43. <https://doi.org/10.3406/rht.1975.1120>.
- Netter, Marie-Laurence, et Christiane Klapisch-Zuber. « Cinquante ans d'histoire au Centre de Recherches Historiques ». *Les Cahiers du Centre de Recherches Historiques. Archives*, n° 36 (30 juin 2005). <https://doi.org/10.4000/ccrh.3048>.
- Nichols, Stephen G. « A Poetics of Historicism? Recent Trends in Medieval Literary Study in Transformation and Continuity ». *Medievalia et Humanistica*, n° 8 (1977): 77-101.
- Noiret, Serge. « La digital history : histoire et mémoire à la portée de tous », 2011. <https://cadmus.eui.eu/handle/1814/17576>.
- Nyhan, Julianne, et Melissa Terras. « Father Busa's Female Punch Card Operatives ». In *Debates in the Digital Humanities*, édité par Matthew K. Gold. Minneapolis, États-Unis d'Amérique: University of Minnesota Press, 2016. <http://dx.doi.org/10.5749/j.ctt1cn6thb.9>.
- Orain, Olivier. « Les Écoles en sciences de l'homme : usages indigènes et catégories analytiques ». *Revue d'histoire des sciences humaines*, n° 32 (1 juillet 2018): 7-38. <https://doi.org/10.4000/rhsh.288>.

- Orain, Olivier, et Marie-Pierre Sol. « Les géographes et le travail collectif. La recherche coopérative sur programme à l'œuvre ». *La revue pour l'histoire du CNRS*, n° 18 (3 octobre 2007). <https://doi.org/10.4000/histoire-cnrs.4061>.
- Owsley, Frank Lawrence, et Harriet Owsley. *Plain Folk of the Old South*. Nashville, États-Unis d'Amérique: Vanderbilt University Press, 1949.
- Padoa Schioppa, Antonio. « Fonti giuridiche e informatica ». *Publications de l'École Française de Rome* 31, n° 1 (1977): 121-23.
- Pages, J.-P., F. Cailliez, et Y. Escoufier. « Analyse factorielle : un peu d'histoire et de géométrie ». *Revue de Statistique Appliquée* 27, n° 1 (1979): 5-28.
- Parinet, Élisabeth. *Une histoire de l'édition à l'époque contemporaine, XIX^e-XX^e siècle*. Paris, France: Éditions du Seuil, 2004.
- Parisse, Michel. « Traitement des documents diplomatiques : exploitation de l' « Index Verborum » ». *Publications de l'École Française de Rome* 31, n° 1 (1977): 335-45.
- Patel, Kiran Klaus. « La scienza al tempo di Wikileaks. Riflessioni sulla storia contemporanea nel ventunesimo secolo ». *Memoria e ricerca* 37 (2011): 155-71.
- Pearson, Karl. « LIII. On lines and planes of closest fit to systems of points in space ». *The London, Edinburgh, and Dublin philosophical magazine and journal of science* 2, n° 11 (1901): 559-72.
- Pêcheux, Michel, Jacqueline Léon, Simone Bonnafous, et Jean-Marie Marandin. « Présentation de l'analyse automatique du discours (AAD69): théories, procédures, résultats, perspectives ». *Mots. Les langages du politique* 4, n° 1 (1982): 95-123. <https://doi.org/10.3406/mots.1982.1053>.
- Pellen, René. « « CAMDAP » (« Computer and Medieval Data Processing »), revue s'intéressant à tout ce qui concerne le traitement par ordinateur des documents médiévaux. Vol. I, n° 1, février 1971 à vol. VIII. n° 1, 1978. » *Cahiers de Civilisation Médiévale* 22, n° 86 (1979): 215-16.
- Piaget, Jean. *Le structuralisme*. Paris, France: Presses universitaires de France, 1968.
- Plutniak, Sébastien. « L'opération archéologique. Sociologie historique d'une discipline aux prises avec l'automatique et les mathématiques. France, Espagne, Italie, 2e moitié du XX^e siècle ». Thèse de doctorat, EHESS, 2017. <http://www.theses.fr/2017EHES0045>.

- . « Aux prémices des humanités numériques ? La première analyse automatisée d'un réseau économique ancien (Gardin & Garelli, 1961). Réalisation, conceptualisation, réception ». *ARCS - Analyse de réseaux pour les sciences sociales / Network analysis for social sciences* 1 (septembre 2018). <https://hal.archives-ouvertes.fr/hal-01870945>.
- . « Le fichier comme format de publication en sciences sociales : les expériences mécanographiques du CADA à Marseille ». In *Marseille en capitales*, édité par Christelle Rabier et Marie-Luce Rauzy. Marseille: Éditions Gaussien, à paraître.
- Polo de Beaulieu, Marie-Anne. « Étude statistique de la structure lexicale de l'exemplum médiéval (d'après la Scala Coeli de Jean Gobi) ». *Histoire & Mesure* 1, n° 3 (1986): 47-80. <https://doi.org/10.3406/hism.1986.1530>.
- Poly, Jean-Pierre. *La Provence et la société féodale : 879-1166*. Paris, France: Bordas, 1976.
- Poly, Jean-Pierre, et Éric Bournazel. *La mutation féodale : X^e-XII^e siècles*. Paris, France: Presses universitaires de France, 1980.
- Potin, Yann, et Julien Théry. « L'histoire médiévale et la « nouvelle érudition ». L'exemple de la diplomatie ». *Labyrinthe*, n° 4 (1 octobre 1999): 35-39.
- Pouille, Emmanuel, et Owen Gingerich. « Les positions des planètes au Moyen Âge : application du calcul électronique aux tables alphonsines ». *Comptes rendus des séances de l'Académie des Inscriptions et Belles-Lettres* 111, n° 4 (1967): 531-48. <https://doi.org/10.3406/crai.1967.12173>.
- Prost, Antoine. *Vocabulaire des proclamations électorales de 1881, 1885 et 1889*. Paris, France: Presses Universitaires de France, 1974.
- . « La collecte et l'analyse des données ». In *Table Ronde Histoire et Informatique*. Paris, France: Éditions du CNRS, 1981.
- Rabb, Theodore K. « The Development of Quantification in Historical Research ». *The Journal of Interdisciplinary History* 13, n° 4 (1983): 591-601. <https://doi.org/10.2307/203882>.
- Rédaction. « Trente ans d'*Histoire & Mesure*. Entretiens croisés avec Gérard Béaur, Jean-Philippe Genet et Jean Heffer ». *Histoire & mesure* 31, n° 2 (2016): 3-9.
- Reinert, Max. « Un logiciel d'analyse lexicale ». *Cahiers de l'analyse des données* 11, n° 4 (1986): 471-81.

- Renisio, Yann, et Rémi Sinthon. « L'analyse des correspondances multiples au service de l'enquête de terrain ». *Genèses* n° 97, n° 4 (27 novembre 2014): 109-25.
- Renouard, Yves. « Destin d'une grande métropole médiévale : Pise ». *Annales* 17, n° 1 (1962): 137-45. <https://doi.org/10.3406/ahess.1962.420802>.
- Revel, Jacques, éd., *Jeux d'échelles : la micro-analyse à l'expérience*. Paris, France: Gallimard et Seuil, 1996.
- Rey-Coyrehourcq, Sébastien. « Une plateforme intégrée pour la construction et l'évaluation de modèles de simulation en géographie ». Thèse de doctorat, Université Paris 1 Panthéon-Sorbonne, 2015. <https://hal.archives-ouvertes.fr/tel-01652092>.
- Reynolds, Susan. *Kingdoms and communities in Western Europe, 900-1300*. Oxford, Royaume-Uni: Clarendon Press, 1984.
- Rippe, Gérard. « L'onomastique dans la Vénétie du XIII^e siècle : l'exemple de Monselice ». *Publications de l'École Française de Rome* 120, n° 1 (1989): 269-85.
- Robert, Jean-Louis. *La scission syndicale de 1921 : essai de reconnaissance des formes*. Paris, France: Publications de la Sorbonne, 1980.
- . « L'analyse factorielle des correspondances : ses bases ». *Le médiéviste et l'ordinateur* 5, n° 1 (1981): 3-5. <https://doi.org/10.3406/medio.1981.949>.
- Robin, Régine. *Histoire et linguistique*. Paris, France: Armand Colin, 1973.
- Rocke, Michael. *Forbidden friendships: homosexuality and male culture in Renaissance Florence*. Oxford, Royaume-Uni: Oxford University Press, 1996.
- Rosental, Paul-André. « Introduction. Outil ou fétiche : la laïcisation de l'analyse factorielle dans les sciences sociales ». *Histoire & Mesure* 12, n° 3 (1997): 185-96.
- . « La nouveauté d'un genre ancien : Louis Henry et la fondation de la démographie historique ». *Population* 58, n° 1 (2003): 103-36. <https://doi.org/10.2307/3271296>.
- Rosenzweig, Roy. *Clio wired: the future of the past in the digital age*. New York, États-Unis d'Amérique: Columbia University Press, 2011.

- Rouault, Jacques, et George Montpied. « Reconstitution et cartographie d'un parcellaire à partir de cadastres textuels du XV^e siècle ». *Le Médiéviste et l'ordinateur* 8, n° 1 (1982): 11-13. <https://doi.org/10.3406/medio.1982.1651>.
- Rouche, Michel, et Pierre Toubert. « L'Aquitaine des Wisigoths aux Arabes (418-781) : essai sur le phénomène régional ». Thèse de doctorat d'État, Université de Lille, 1977.
- Rougerie, Jacques. « Faut-il départementaliser l'histoire de France ? » *Annales* 21, n° 1 (1966): 178-93. <https://doi.org/10.3406/ahess.1966.421360>.
- Rouwenhorst, K. Geert, et William N. Goetzmann. *The Origins of Value: The Financial Innovations That Created Modern Capital Markets*. Oxford, Royaume-Uni: Oxford University Press, 2005.
- Rowley-Jolivet, Elizabeth, et Shirley Carter-Thomas. « Raconter comment cela se passe vraiment. Comparaison entre la narration de la recherche dans les cahiers de laboratoire et dans les articles de recherche ». In *La mise en récit dans les discours spécialisés*, édité par Catherine Resche, 86-121. Berne, Suisse: Peter Lang, 2015. <https://hal.archives-ouvertes.fr/hal-01256112>.
- Ruggiero, Guido. *The boundaries of eros: sex crime and sexuality in Renaissance Venice*. Oxford, Royaume-Uni: Oxford University Press, 1985.
- Ruiz, Pablo, et Thierry Poibeau. « Mapping the Bentham Corpus: Concept-based Navigation ». *Journal of Data Mining and Digital Humanities* Special Issue: Digital Humanities between knowledge and know-how (Atelier Digit_Hum) (mars 2019). <https://hal.archives-ouvertes.fr/hal-01915730>.
- Russell, D. B. « COCOA-a word count and concordance generator ». Dernière consultation le 4 septembre 2021. <http://www.chilton-computing.org.uk/acl/applications/cocoa/p001.html>.
- Rygiel, Philippe. *Historien à l'âge numérique*. Villeurbanne, France: Presses de l'ENSSIB, 2019. <http://books.openedition.org/pressesenssib/6303>.
- . « Introduction ». *Historien à l'âge numérique*. Villeurbanne, France: Presses de l'ENSSIB, 2019. <http://books.openedition.org/pressesenssib/6327>.
- . « Les sources de l'historien à l'heure d'Internet ». *Hypothèses* 7, n° 1 (2004): 341-54.

- Sánchez-Albornoz y Menduiña, Claudio. *Estudios sobre las instituciones medievales españolas*. México, Mexique: Universidad Nacional Autónoma de México Instituto de Investigaciones Históricas, 1965.
- Saunier, Georges. « Quelques réflexions sur le concept de Centre et Périphérie ». *Hypothèses* 3, n° 1 (2000): 175-80.
- Sauvy, Alfred. « Vues, travaux, cheminement de Louis Henry ». *Population* 32, n° 1 (1977): 7-11.
- Scalfati, Silio Pietro Paolo. « Problèmes de méthodes, à propos des actes notariés de Pise ». *Publications de l'École Française de Rome* 31, n° 1 (1977): 251-52.
- Scheuch, Erwin K. « Les services de données en sciences sociales : historique et perspectives ». *Revue internationale des sciences sociales* n° 177, n° 3 (2003): 433-49.
- Schreibman, Susan, Ray Siemens, et John Unsworth. « The digital humanities and humanities computing: An introduction ». In *A companion to digital humanities*, édité par Susan Schreibman, Ray Siemens, et John Unsworth, 288-90. Oxford, Royaume-Uni: Blackwell, 2004.
- Seignobos, Charles. *La méthode historique appliquée aux sciences sociales*. Paris, France: Félix Alcan, 1901.
- Selz, Marion. « Procédures d'aide à la reconstitution des généalogies ». *Histoire & Mesure* 4, n° 1 (1989): 3-19. <https://doi.org/10.3406/hism.1989.872>.
- Sfoini, Alexandra. « In Memoriam: Spyros I. Asdrachas (1933-2017) ». *The Historical Review/La Revue Historique* 14, n° 0 (2017): 9-12. <https://doi.org/10.12681/hr.16270>.
- Shaw, Brent D. « The Age of Roman Girls at Marriage: Some Reconsiderations ». *The Journal of Roman Studies* 77 (1987): 30-46. <https://doi.org/10.2307/300573>.
- Shorter, Edward. *Historian and the Computer: A Practical Guide*. New York, États-Unis d'Amérique: WW Norton & Co Inc, 1975.
- Shoups, Inge. « Une grille plus connue : celle de l'historien ». *Le Médiéviste et l'ordinateur* 11, n° 1 (1984): 5-7. <https://doi.org/10.3406/medio.1984.1680>.
- Silbey, Joel H. « Clio and Computers: Moving into Phase II, 1970-1972 ». *Computers and the Humanities* 7, n° 2 (1972): 67-79.

- Skinner, Quentin. *Reason and rhetoric in the philosophy of Hobbes*. Cambridge, Royaume-Uni: Cambridge University Press, 1996.
- Skornicki, Arnault, et Jérôme Tournadre. « I. L' « école de Cambridge » et le contextualisme ». In *La nouvelle histoire des idées politiques*, édité par Arnault Skornicki et Jérôme Tournadre, 9-32. Paris, France : La Découverte, 2015.
- Smith, Jon Michael, Christian Saguez, et Jean-Marc Kerilis. *Méthodes numériques pour calculateur de poche*. Paris, France: Eyrolles, 1976.
- Smith, O. R. « GENDEX : GENERAL inDEX of Words with Context. A Concordance Generator ». *Computer Studies III* (1970): 50-53.
- Sollaci, Luciana B., et Mauricio G. Pereira. « The introduction, methods, results, and discussion (IMRAD) structure: a fifty-year survey ». *Journal of the Medical Library Association* 92, n° 3 (juillet 2004): 364-71.
- Souchal, François. « Christiane Klapisch-Zuber. Les maîtres du marbre. Carrare 1300-1600. Paris, 1969. (École pratique des hautes études, VIe section. Ports, routes, trafics, XXV.) ». *Bibliothèque de l'École des chartes* 130, n° 1 (1972): 306-8.
- Soulet, Jean-François, Eric Castex, Robert Marconis, eds., *L'informatique dans l'enseignement de l'histoire et la formation des historiens : actes du V^e colloque national de l'Association française pour l'histoire et l'informatique, 3 et 4 novembre 1998, université de Toulouse-le-Mirail*. Paris, France: L'Harmattan, 2001.
- Souvay, Gilles, et Jean-Marie Pierrel. « LGeRM Lemmatisation des mots en Moyen Français ». *Revue TAL* 50, n° 2 (2009): 21.
- Spearman, Charles. « "General Intelligence" Objectively Determined and Measured. » *American Journal of Psychology* 15 (1904): 201-93.
- Stevens, L. D. « The Evolution of Magnetic Storage ». *IBM Journal of Research and Development* 25, n° 5 (septembre 1981): 663-76. <https://doi.org/10.1147/rd.255.0663>.
- Stouff, Louis. « Un livre terrier arlésien complété de documents divers ». *Le Médiéviste et l'ordinateur* 3, n° 1 (1980): 11-14. <https://doi.org/10.3406/medio.1980.921>.
- Stuard, Susan Mosher. « Family Confronts the Renaissance Household: A Review of Les Toscans et Leurs Familles ». *The Journal of Interdisciplinary History* 11, n° 3 (1981): 495-501. <https://doi.org/10.2307/203631>.

- Sula, Chris Alen, et Heather V. Hill. « The early history of digital humanities: An analysis of Computers and the Humanities (1966–2004) and Literary and Linguistic Computing (1986–2004) ». *Digital Scholarship in the Humanities* 34, n° Supplement_1 (1 décembre 2019): 190-206. <https://doi.org/10.1093/llc/fqz072>.
- Swales, John M. « Reflections on the Concept of Discourse Community ». *ASp. La Revue Du GERAS*, n° 69 (9 mars 2016): 7-19. <https://doi.org/10.4000/asp.4774>.
- Swales, John M. *Genre analysis: English in academic and research settings*. Cambridge, Royaume-Uni: Cambridge University Press, 1990.
- Swales, John. « Approaching the Concept of Discourse Community », non-publié, 1987. <https://eric.ed.gov/?id=ED286184>.
- Swierenga, Robert P. « Clio and Computers: A Survey of Computerized Research in History ». *Computers and the Humanities* 5, n° 1 (septembre 1970): 1-21. <https://doi.org/10.1007/BF02404252>.
- . « Computers and American History: The Impact of the “New” Generation ». *The Journal of American History* 60, n° 4 (1974): 1045-70. <https://doi.org/10.2307/1901012>.
- Tenenti, Alberto. « Christiane Klapisch-Zuber, *Les Maîtres du marbre. Carrare, 1300-1600* ». *Annales* 25, n° 5 (1970): 1390-92.
- Thaller, Manfred. « Data bases v. critical editions ». *Historical social research/Historische sozialforschung* 13, n° 3 (47 (1988): 129-39.
- Thomas, William G. « Computing and the Historical Imagination ». In *A Companion to Digital Humanities*, édité par Susan Schreibman, Ray Siemens, et John Unsworth, 288-90. Oxford, Royaume-Uni: Blackwell, 2004. <https://doi.org/10.1002/9780470999875.ch5>.
- Timbal, Pierre-Clément. « Une grille de départ : celle du juriste ». *Le Médiéviste et l'ordinateur* 11, n° 1 (1984): 5-5. <https://doi.org/10.3406/medio.1984.1679>.
- Tits-Dieuaide, Marie-Jeanne. « Les campagnes flamandes du XIII^e siècle au XVIII^e siècle, ou les succès d'une agriculture traditionnelle ». *Annales* 39, n° 3 (1984): 590-610. <https://doi.org/10.3406/ahess.1984.283079>.
- Tock, Benoît-Michel, Michèle Courtois, Marie-José Gasse-Grandjean, et Philippe Demonty, édés., *La diplomatie française du Haut Moyen Âge: Inventaire des chartes originales antérieures à 1121 conservées en France. Vol. 4*. Turnhout,

- Belgique: Brepols Publishers, 2001. <https://doi.org/10.1484/M.ARTEM-EB.5.105723>.
- Tollenaere, Félicien de. « Les chartes néerlandaises en langue vulgaire ». *Publications de l'École Française de Rome* 31, n° 1 (1977): 105-6.
- Tombeur, Paul. *Raoul de Saint-Trond : Gesta abbatum Trudonensium*. Hildesheim, Allemagne: G. Olms, 1969.
- . « Le Centre de Traitement Électronique des Documents, au service des disciplines ecclésiastiques ». *Revue Théologique de Louvain* 1, n° 4 (1970): 433-43. <https://doi.org/10.3406/thlou.1970.1057>.
- Toubert, Pierre, et Ottorino Bertolini. *Les structures du Latium médiéval : le Latium méridional et la Sabine du IX^e siècle à la fin du XII^e siècle*. Rome, Italie: Publication de l'École française de Rome, 1973.
- Toubert, Pierre. « Nécrologie Cinzio Violante (1921-2001) ». *Le Moyen-Âge* CVII, n° 2 (2001): 395-97.
- Tournier, Maurice. « Le centre de recherche de lexicologie politique de l'E.N.S. de Saint-Cloud ». *Langue française* 2, n° 1 (1969): 82-86. <https://doi.org/10.3406/lfr.1969.5424>.
- . « Mots et politique, avant et autour de 1980 - Entretien ». *Mots. Les langages du politique*, n° 94 (30 novembre 2010): 211-33. <https://doi.org/10.4000/mots.19889>.
- Tournier, Maurice, R. Arnault, L. Cavaciuti, Annie Geffroy, et Françoise Theuriot. « Le vocabulaire de la Révolution. Pour un inventaire systématique des textes ». *Annales historiques de la Révolution française* 195, n° 1 (1969): 109-24. <https://doi.org/10.3406/ahrf.1969.3721>.
- Trexler, Richard. « La prostitution florentine au XV^e siècle ». *Annales* 36, n° 6 (1981): 983-1015. <https://doi.org/10.3406/ahess.1981.282801>.
- « Un nouveau logiciel ». *Le médiéviste et l'ordinateur* 5, n° 1 (1981): 22-23.
- Ustinov, Valentin A. « Les calculateurs électroniques appliqués à la science historique ». *Annales* 18, n° 2 (1963): 263-94. <https://doi.org/10.3406/ahess.1963.420976>.
- Van Luyn, P. « Les milites dans la France du XI^e siècle : examen des sources narratives ». *Le Moyen-Âge* LXXVII, n° 1 (1971): 5-51.

- Veldman, Donald J. *Fortran Programming for the Behavioral Sciences*. New York, États-Unis d'Amérique: Holt, Rinehart and Winston, 1967.
- Verger, Jacques. *Histoire des universités en France*. Toulouse, France: Privat, 1986.
- Verriest, Léo. « Les faits et la terminologie en matière de condition juridique des personnes au Moyen âge : serfs, nobles, villains, sainteurs ». *Revue du Nord* 25, n° 98 (1939): 101-27. <https://doi.org/10.3406/rnord.1939.1814>.
- Villani, Giovanni. *Cronica di Giovanni Villani: a miglior lezione ridotta coll'ajuto de'testi a penna, con note filologiche di I. Moutier e con appendici storico-geografiche, compilate da Franc. Gherardi Dragomanni*. Firenze, Italie: S. Coen, 1844.
- Vovelle, Michel. « La mémoire d'Ernest Labrousse ». *Annales historiques de la Révolution française* 276, n° 1 (1989): 99-107. <https://doi.org/10.3406/ahrf.1989.1243>.
- Wenin, Christian. « L'informatique au service de la philosophie. Réalisation et projets ». *Revue Philosophique de Louvain* 70, n°6 (1972): 177-211. <https://doi.org/10.3406/phlou.1972.5663>.
- . « Le V^e Congrès international de philosophie médiévale ». *Revue Philosophique de Louvain* 71, n° 9 (1973): 126-33.
- Werner, Karl-Ferdinand. « Liens de parenté et noms de personne. Un problème historique et méthodologique ». *Publications de l'École Française de Rome* 30, n° 1 (1977): 13-18.
- . « Problèmes de l'exploitation des documents textuels concernant les noms et les personnes du monde latin (III^e-XII^e siècles) ». *Publications de l'École Française de Rome* 31, n° 1 (1977): 205-12.
- Werner, Michael. « « Les Mots de l'histoire » et la Begriffsgeschichte / sémantique historique ». *Revue de l'IFHA. Revue de l'Institut français d'histoire en Allemagne*, n° 4 (30 septembre 2012): 187-94. <https://doi.org/10.4000/ifha.460>.
- Wimmer, Mario. « Josephine Miles (1911–1985): Doing Digital Humanism with and without Machines ». *History of Humanities* 4, n° 2 (1 septembre 2019): 329-34. <https://doi.org/10.1086/704850>.
- Wolff, Philippe. *Histoire de Toulouse*. Toulouse, France : Privat, 1958.

- Wollasch, Joachim. « Les obituaires, témoins de la vie clunisienne ». *Cahiers de Civilisation Médiévale* 22, n° 86 (1979): 139-71. <https://doi.org/10.3406/ccmed.1979.2103>.
- Woolf, Stuart. « On University Reform in Italy: Contradictions and Power Relations in Structure and Functio ». *Minerva* 41, n° 4 (2003): 347-63.
- Yates, Frank. « Computers, the second revolution in statistics ». *Biometrics* 22, n° 2 (1966): 233-51.
- Zadora-Rio, Elisabeth. « Archéologie et toponymie : le divorce ». *Les petits cahiers d'Anatole*, n° 8 (2001). <https://halshs.archives-ouvertes.fr/halshs-00591271>.
- Zerner, Monique. « Cadastres du Comtat Venaissin ». *Le Médiéviste et l'ordinateur* 3, n° 1 (1980): 7-9. <https://doi.org/10.3406/medio.1980.1641>.
- Zerner, Monique. « Un type de document répétitif fournissant des données quantitatives ». *Publications de l'École Française de Rome* 31, n° 1 (1977): 157-64.
- Zysberg, André. « Les historiens et l'apprentissage de l'informatique ». In *Table ronde « Histoire et informatique » du 8 mars 1980*, 16-28. Paris, France: Éditions du CNRS, 1980.
- . *Les galériens : vies et destins de 60 000 forçats sur les galères de France, 1680-1748*. Paris, France: Éditions du Seuil, 1987.
- . « Impact de l'informatique sur la recherche historique ». *Matériaux pour l'histoire de notre temps* 10, n° 1 (1987): 35-37. <https://doi.org/10.3406/mat.1987.401466>.

Entretiens.

Nous avons mené, au cours de la première année de notre thèse (2016-2017), une série d'entretiens semi-directifs auprès de chercheurs ayant vécu, de près ou de loin, et à des périodes différentes, l'irruption des ordinateurs en histoire. Cette phase de la recherche nous a été d'une grande aide pour dégager des problématiques, identifier des collectifs de recherche ayant mis en œuvre ce type de pratiques et repérer des sources permettant l'étude de ce type de projet. Il nous apparaissait par ailleurs particulièrement important de collecter ce type de témoignages oraux pour le futur.

Toutefois, nous avons fait le choix de ne mobiliser que très ponctuellement les informations recueillies au fil de ces entretiens dans notre argumentation. Cette décision repose sur le fait que nous avons toujours préféré privilégier les sources écrites, dont nous avons pu constater qu'elles étaient toujours plus sûres et plus précises. Pour cette raison, nous n'avons pas considéré nécessaire de retranscrire les entretiens. Nous nous tenons cependant à la disposition de quiconque aimerait avoir accès aux enregistrements.

Pour avoir une idée lointaine de la teneur de ces 20 moments d'échanges enregistrés, en voici la liste par ordre chronologique (nous indiquons la durée des entretiens entre parenthèses) :

- 2 novembre 2016, Françoise Waquet (1 heure 37 minutes)
- 8 novembre 2016, Jacques Revel (1 heure 28 minutes)
- 22 novembre 2016, Michel Kaplan (1 heure 10 minutes)
- 4 décembre 2016, Christiane Klapisch-Zuber (1 heure 05 minutes)
- 6 décembre 2016, Alain Guerreau (3 heures 06 minutes)
- 18 décembre 2016, Michel Demonet (2 heures 02 minutes)
- 19 décembre 2016, Jean-Philippe Genet (1 heure 16 minutes)
- 13 janvier 2017, Ezio Ornato (1 heure 26 minutes)
- 15 janvier 2017, Gérard Béaur (56 minutes)
- 22 janvier 2017, Hélène Millet (1 heure 43 minutes)
- 26 janvier 2017, Antoine Prost (sans enregistrement)
- 31 janvier 2017, Jean Heffer (1 heure 43 minutes)
- 6 février 2017, André Salem (2 heures 36 minutes)

12 février 2017, Marie-Anne Polo de Beaulieu (1 heures 05 minutes)

19 février 2017, Philippe Cibois (43 minutes)

4 mars 2017, Caroline Bourlet (2 heures 05 minutes)

14 mars 2017, Olivier Mattéoni (1 heure 26 minutes)

24 mars 2017, Joseph Morsel (1 heure 27 minutes)

3 mai 2017, Béatrice Joyeux-Prunel (41 minutes)

Table des matières.

Remerciements.....	7
Liste des principales abréviations.....	11
Sommaire.....	12

Introduction générale.....	20
----------------------------	----

PARTIE I

Chapitre 1 : Une étude assistée par ordinateur du <i>catasto</i> florentin de 1427 (1966-1978).....	45
1. Organisations collectives et ordinateurs.....	53
2. Opérations de constitution de l' « édition » : collecte des données, code et codage.....	78
3. Raisonnements, programmes et documents de sortie.....	98
4. Réutilisations de l' « édition » et des conclusions de l'étude.....	125
Conclusion.....	141

Chapitre 2 : Lexicométrie et histoire médiévale : le programme ALINE de l'Équipe de Recherche Associée 713 (1971-1984).....	151
1. Organisations collectives et ordinateurs.....	162
2. Constitution des corpus et éditions des textes.....	179
3. Raisonnements, programme et statistiques lexicales.....	198
4. Réutilisation et devenir de ALINE et de MEDITEXT.....	231
Conclusion	243

PARTIE II

Chapitre 3 : Une nouvelle ambition collective pour l'histoire médiévale : les calculateurs électroniques (1970-1979)	252
1. Les calculateurs électroniques, nouveau vecteur de collaboration à l'échelle de la discipline (1970-1975).....	259
2. Problématiques communes : les échanges du colloque de Rome.....	275
3. Deux propositions d'organisation méta-collective à l'échelle européenne.....	302
Conclusion	315

Chapitre 4 : Objectifs et stratégies de publication du bulletin de liaison <i>Le Médiéviste et l'ordinateur</i> (1979-1989).....	322
1. L'organisation du <i>Médiéviste et l'ordinateur</i>	328
2. Coordonner la production textuelle d'une communauté discursive.....	346
3. Favoriser la circulation de nouvelles pratiques.....	364

Conclusion	385
-------------------------	------------

PARTIE III

Chapitre 5 : Genre et fonctions des « articles de fond » du <i>Médiéviste et l'ordinateur</i> (1979-1989).....	395
---	------------

1. Une auteure, quatre contextes de publication.....	401
--	-----

2. Parties méthodologiques	414
----------------------------------	-----

3. Les « articles de fond » du <i>Médiéviste et l'ordinateur</i>	448
--	-----

Conclusion	469
-------------------------	------------

Chapitre 6 : Conflits épistémologiques autour d'une utilisation des analyses factorielles des correspondances.....	477
---	------------

1. Le problème ancien du développement du féodalisme à l'aune d'une procédure mathématique nouvelle.....	486
--	-----

2. Simplifier les opérations de synthèse historique.....	511
--	-----

Conclusion.....	530
------------------------	------------

Conclusion générale.....	536
---------------------------------	------------

Bibliographie.....	562
---------------------------	------------

Entretiens	602
-------------------------	------------

Table des matières.....	604
--------------------------------	------------

Table des figures.....	606
-------------------------------	------------

Table des figures.

Figure 1. Carte perforée IBM à 80 colonnes.....	56
Figure 2. Bordereau de perforation.....	73
Figure 3. Exemple de perforatrice utilisée à la fin des années 1960 (ici le modèle IBM 029 introduit sur le marché en 1964).....	74
Figure 4. Exemple de bordereaux complétés dans le cadre de l'enquête.....	86
Figure 5. Liste des abréviations utilisées.....	87
Figure 6. Tableau des professions publié <i>Medieval and Renaissance Pistoia</i> (1967)....	103
Figure 7. Tableau des professions publié dans <i>Les Toscans et leurs familles</i> (1978)....	104
Figure 8. Tableau reprenant les résultats des trois méthodes d'évaluation de l'âge au premier mariage.....	110
Figure 9. Matrice de corrélations publiée par Klapisch-Zuber et Herlihy.....	113
Figure 10. Exemple de carte produite de façon automatique par le laboratoire de cartographie automatique de Jacques Bertin dans le cadre de l'enquête du catasto.....	118
Figure 11. Carte représentant la distribution géographique du rapport de masculinité.....	120
Figure 12. Planche présentant trois cartes de la distribution géographique de professions.....	122
Figure 13. Exemple de dictionnaire de fréquences produit à l'aide de ALINE.....	207
Figure 14. Exemple d'un concordancier produit à l'aide de ALINE.....	212
Figure 15. Schéma d'un lexicogramme.....	217
Figure 16. Graphique d'analyse factorielle basée sur le tableau du nombre d'occurrences (axes 1-2).....	226
Figure 17. Deux graphiques d'analyse factorielle des correspondances basé sur deux types de données (tableau des fréquences relatives à gauche et tableau du nombre d'occurrences à droite).....	227
Figure 18. Nombre d'articles rédigés par les membres du comité éditorial du <i>Médiéviste et l'ordinateur</i> (1979-1989).....	341
Figure 19. Typologie des numéros du <i>Médiéviste et l'ordinateur</i> (1979-1989).....	349
Figure 20. Graphe de cooccurrences des 882 noms communs les plus fréquents du corpus.....	358
Figure 21. Liste des sujets du <i>Médiéviste et l'ordinateur</i> établis sur la base du graphe de cooccurrences (1979-1989).	362
Figure 22. Tableau du nombre d'occurrences des 40 termes les plus fréquents reproduit par Genet.....	377
Figure 23. Tableau des valeurs propres reproduit par Genet.....	378
Figure 24. Graphique d'analyse factorielle des correspondances reproduit par Genet...	379
Figure 25. Exemple d'annotations sur un document reproduit par François Lassus....	381
Figure 26. Graphique d'analyse factorielle de correspondance annoté.....	382

Figure 27. Proportions d'articles comportant des éléments de présentation étape par étape.....	458
Figure 28. Proportions d'articles comportant des mentions à la matérialité des opérations.....	460
Figure 29. Distribution de quelques pronoms personnels (<i>je, moi, m', nous, on</i>).....	461
Figure 30. Proportion des articles comportant des mentions des difficultés ressenties par les auteurs.....	463
Figure 31. Répartition du nombre d'articles de la rubrique « <i>Article de fond</i> » du <i>Médiéviste et l'ordinateur</i> en fonction des sous-genre identifiés.....	469
Figure 32. Tableaux produits par Banniard (à gauche) et Guerreau (à droite).....	479
Figure 33. Légende (à gauche) et tableau (à droite).....	489
Figure 34. Légende du tableau de Banniard.....	490
Figure 35. Tableau synoptique publié par Banniard.....	494
Figure 36. Matrice d'incidence construite par Guerreau sur la base du tableau de Banniard.....	497
Figure 37. Tableau de fréquences construit sur la base du tableau de contingence de Banniard.....	502
Figure 38. Graphique d'analyses factorielles des correspondances des données non-pondérées (plan factoriel 1-2).....	504
Figure 39. Graphique d'analyses factorielles des correspondances des données pondérées (plan factoriel 1-2).....	505
Figure 40. Tableau proposé par Guerreau en réponse à celui de Banniard.....	508