

## INFORMATIQUE ET ÉTUDE DE TEXTES MÉDIÉVAUX

Lors du colloque de Louvain-la-Neuve, la Commission VII a traité les points suivants :

1. Compte rendu du congrès de Bonn en 1977.
2. Informations concernant
  - a) les projets de recherches,
  - b) les colloques,
  - c) l'évolution du matériel informatique.
3. Problèmes de méthodologie.

J'adopterai un schéma identique. Je rappelle, en effet, que le rôle de cette Commission est de faire le point sur les projets de recherches en cours ou terminés depuis la dernière réunion, sur les publications en la matière, sur les colloques, sur l'évolution de l'informatique en ce qui concerne nos études, et enfin sur ce qui semble devoir être mis en évidence du point de vue méthodologique.

### 1. Compte rendu de la session du congrès de Louvain-la-Neuve en 1982

Qu'il nous suffise de renvoyer au texte publié dans les Actes de ce congrès, tome I, aux pages 174 à 186.

### 2. Projets de recherches

Depuis le congrès de Louvain-la-Neuve, on a assisté à une véritable explosion informatique, due principalement au développement de la micro-informatique. Nous aurions d'ailleurs l'occasion d'en reparler quand nous aborderons les questions d'évolution du matériel et les problèmes méthodologiques, la micro-informatique, comme toutes choses, ayant son caractère positif et négatif.

L'informatique s'introduit peu à peu partout, et c'est tout à fait normal puisque nous traitons des données qui correspondent toujours à des fichiers, dont l'étude est automatisable dans une mesure souvent très large.

Etre informé des projets en cours est assurément une chose très importante : il faut profiter de ce qui existe, de tout ce qui est utilisable, éviter les doubles ou les triples emplois, bref il faut organiser au mieux la recherche, pour des questions de temps, d'argent et d'efficacité. Cette Commission peut ainsi jouer un rôle important.

Je distinguerai les projets bibliographiques, les projets codicologiques, les projets d'études textuelles, lexicales et autres. Le relevé en a été fait grâce à ce qui nous a été communiqué, grâce au dépouillement de revues, et spécialement de *Computers and Medieval Data Processing* (CAMDAP) et de la revue *Le Médiéviste et l'Ordinateur*. Signalons que CAMDAP a réalisé en 1983 l'index des volumes 1 à 13, index qu'il faut évidemment compléter pour les travaux annoncés par ailleurs. Notons cependant qu'il n'est pas toujours aisé de savoir si les chercheurs ont utilisé ou non l'ordinateur, ni dans quelle mesure. Ainsi dans la bibliographie

de CAMDAP, il y a plusieurs références d'articles et d'ouvrages qui n'ont rien à voir avec l'utilisation de l'ordinateur. Dans les informations données dans le *Bulletin* de la SIEPM en ce qui concerne les éditions, rien n'est dit sur une informatisation éventuelle. Or ce point est très important - nous y reviendrons d'ailleurs au paragraphe méthodologique -. En effet, tel texte critique mis sur disquettes pourrait être intégré dans une banque de données. De plus, comme j'ai pu en faire l'étonnante expérience, certains chercheurs qui utilisent des machines de traitement de texte, ignorent tout ce qui peut être élaboré à partir de l'enregistrement effectué. Je propose dès lors que, dans le questionnaire que reçoivent les membres de la SIEPM, on ajoute une question relative à l'utilisation ou non de supports informatiques et à la spécification de ces supports.

#### a) Projets bibliographiques

Ceux-ci se développent de plus en plus. On connaît les outils usuels, tel le *Bulletin signalétique*, dont le traitement informatique est réalisé par le "Centre de Documentation Sciences Humaines" de Paris : des recherches bibliographiques sont aujourd'hui possible en mode conversationnel, bien qu'il faille mentionner que le coût d'une telle interrogation s'élève très rapidement. Des logiciels ont été mis au point un peu partout pour réaliser l'informatisation des données bibliographiques, notamment à l'aide d'un micro-ordinateur, tel le logiciel bibliographique présenté dans le volume 20 de la revue *Documentaliste*, de juillet-octobre 1983. Toute bibliographie d'ensemble devrait désormais être automatisée. Il est à souhaiter qu'on ne se contente pas de simples listes de titres, mais qu'on insère quelques notes critiques : que de publications qu'il est inutile de consulter, et, par ailleurs, que de publications que nous devrions consulter, mais qui se trouvent noyées dans la masse et que l'on risque dès lors d'ignorer ! Sans doute, ce souhait risque-t-il de demeurer souvent un vœu pieux, mais il me paraît néanmoins important de le formuler.

Je signalerai ici le projet *Rhetor* qui a pour but de donner une entrée de type encyclopédie "for every rhetorician and his works", et ce du V<sup>e</sup> s. avant Jésus-Christ. jusqu'en 1914. Quand on sait l'importance fondamentale que revête la rhétorique dans le développement de la pensée patristique et médiévale - lesquelles doivent sans cesse être associées -, on devine l'utilité d'un tel projet. Celui-ci est dirigé par un maître en la matière, James Murphy de l'Université de Californie. Un premier produit - modeste, certes, mais il fallait démarrer - est le résultat du traitement par ordinateur des informations contenues dans l'ouvrage de James Murphy, *Renaissance Rhetoric. A short title Catalogue*, publié en 1981. Le catalogue des ouvrages rhétoriques diffusés par incunables est en préparation. Pour les années 400 à 1455, 750 auteurs se trouvent ainsi répertoriés.

#### b) Projets codicologiques

Là aussi on constate, heureusement, l'utilisation croissante de l'emploi de l'ordinateur. L'important est d'arriver à une coordination dans la description des manuscrits décrits et à une standardisation permettant des index cumulés et des interrogations sur de grands ensembles.

Il faut ici donner la première place à un institut maître en la matière, l'Institut de Recherche et d'Histoire des Textes. Le projet de l'I.R.H.T. a été présenté par M.J. Beaud, A. Guillaumont et J.L. Minel, sous le titre *A Medieval Manuscript Data Base*, dans le volume d'actes

publié par Robert Allen, que je mentionnerai également sous la rubrique des colloques : *Data Bases in the Humanities and Social Sciences*, 1985. Cette base de données sera interrogeable par réseau. Une première expérimentation de ce réseau est en cours entre Montréal et Paris.

Pour la bibliothèque de Saint-Victor de Paris, V. Gerz-Von Buren, H.R. Hubschmid, G. Ouy et C. Regnier ont publié en 1983 *Le catalogue de la bibliothèque de Saint-Victor de Paris de Claude de Grandrue 1514*.

A Pittsburgh, Alison Stones et Elizabeth Peterson ont entrepris un *Index to French liturgical and Devotional Manuscripts*. Ce projet comprend le catalogue informatisé de manuscrits de textes latins et français.

Julian Plante à Collegeville a entrepris *A Computer-based Catalogue of the Alcobaca manuscripts* conservés à la Bibliothèque Nationale de Lisbonne.

Dans le domaine des textes scientifiques, deux projets sont à mentionner :

- celui de Nan L. Hahn de New Brunswick : *Benjamin Data Bank of Medieval Scientific Manuscripts in Latin* : il s'agit là de manuscrits rassemblés par Francis S. Benjamin pour la période 500-1600.
- Karen Reeds à Berkeley a un projet similaire pour les *Medieval and Scientific Books in Medieval Libraries*, pour la période allant du VII<sup>e</sup> siècle jusqu'en 1520.

Doivent être signalées les entreprises menées par le Professeur A. Gruys et d'autres, à Nimègue : *Producing codicological catalogues with the aid of computers*.

Un projet voudrait démarrer : c'est celui annoncé par le Père Bougerol qui fait appel aux collaborateurs (dans le *Bulletin* de la SIEPM de 1985), afin de réaliser une *Codicographie de saint Bonaventure*.

Un projet démarre, sous l'égide de l'ACCT (Agence de Coopération Culturelle et Technique), et sous la présidence de Jacqueline Hamesse : *Répertoire international d'incipit des manuscrits médiévaux*.

Dans le domaine juridique, mentionnons les travaux de G. Dolezalek, et de M. Huglo pour les manuscrits relatifs à la musique médiévale. Quant au catalogue des manuscrits hébraïques datés, dont le répertoire est constitué à l'I.R.H.T., leur mise en ordinateur a été faite à l'Université de Jérusalem.

#### c) Projets d'études textuelles, lexicales et autres

Je vous propose de faire connaître, ou de rappeler, d'abord les projets qui concernent un auteur, ensuite les projets d'ensemble, enfin des études particulières.

- A) Pour les auteurs, le plus simple est d'adapter un ordre alphabétique.
- Abélard : nous avons parlé des travaux de David Luscombe de l'Université de Sheffield au colloque de Louvain-la-Neuve. Celui-ci continue à travailler dans ce domaine, d'autant plus qu'un projet d'ensemble pour une édition des oeuvres d'Abélard a été mis sur pied dans le cadre du *Corpus Christianorum, Continuatio Mediaevalis*, projet que dirige D. Luscombe et L.J. Engels, de l'Université de Groningue. Ce projet

connaît, d'autre part, un prolongement dans le cadre des *Instrumenta Lexicologica Latina*, dont il sera question plus loin.

- Anselme de Canterbury : la concordance d'Anselme, annoncée précédemment, a été publiée en 4 volumes, en 1984.
- Avicenne : Mademoiselle S. Van Riet continue l'élaboration de lexiques arabo-latins et latino-arabes. On peut lire à ce sujet ce qu'a publié Madame M.C. Lambrechts, dans la *Revue Philosophique de Louvain*, tome 81, 1983, sous le titre : *Les "lexiques" de la Métaphysique*.
- Bonaventure : on n'a pas, hélas, procédé à de nouveaux enregistrements de textes de Bonaventure. Par contre, une première lemmatisation des *Sermones dominicales* a été effectuée. Signalons, par ailleurs, qu'on a continué au Cetedoc l'élaboration du *Corpus des Sources franciscaines*. Celui-ci se voit enrichi des *Actus sancti Francisci*, le texte qui est à l'origine des Fioretti : la concordance lemmatisée de cette oeuvre est sortie de presse en septembre 1987. Un septième volume paraîtra en 1988, lequel sera suivi par la *concordantia concordantiarum* de tout le corpus franciscain. Si je signale ce projet de publication, c'est précisément parce qu'il permettra notamment de déterminer la spécificité d'un Bonaventure par rapport à l'ensemble de ces textes relevant de l'ordre dont il fut le général.
- Giordano da Pisa : les sermons de cet auteur ont été étudiés à l'Université de Lausanne par Cristina Murillo-Bianchi, mais je ne dispose pas d'information supplémentaire à cet égard.
- Jean Scot Érigène : en rendant compte de la publication des index du *Periphyseon*, assumée par G.H. Allard à l'Université de Montréal, le Père Goulven Madec a fait le point en ce qui concerne - et c'est le titre qu'il utilise - *Jean Scot à l'ordinateur*. Voyez la *Revue des Études Augustiniennes*, 30 (1984), pp. 303-307. Goulven Madec note qu'en tenant compte de ce qui a été fait à Montréal et à Louvain-la-Neuve, il resterait "à enregistrer l'homélie et le commentaire de saint Jean, ainsi que les traductions d'oeuvres de Grégoire de Nysse et de Maxime le Confesseur... et l'on pourrait espérer voir paraître, sans trop tarder, une véritable concordance complète des oeuvres de Jean Scot." Depuis que Madec a écrit ces lignes, je préciserai que certaines traductions de Maxime le Confesseur sont dans la banque de données du Cetedoc, dont je parlerai ultérieurement.
- Physiologus : Karen C. Kossuth de Claremont a annoncé en 1982 un projet "*Lexical and syntactical concordances for the early German Physiologus along with its Latin counterpart*". Les textes parallèles latins seraient donc englobés dans ce projet.
- Pierre Lombard : les concordances de formes du *Livre des Sentences*, auxquelles travaille Jacqueline Hamesse, vont paraître incessamment. Cette publication, annoncée d'abord dans la collection *Informatique et étude de textes*, trouvera sa place dans le cadre des *Instrumenta Lexicologica Latina* dont nous parlerons ultérieurement.
- Robert Grosseteste : ses traductions du Pseudo-Denys seront citées *infra*. Le projet que je voudrais rappeler ici est celui dont il a déjà été fait mention au colloque de Louvain-la-Neuve : les commentaires de l'Éthique à Nicomaque dans la traduction latine de Robert Grosseteste. Ce projet de Paul Mercken qui, pour plusieurs circonstances, a connu

son temps de léthargie, a été remis sur les rails : on a procédé à une juxtaposition informatisée des textes latins et des textes grecs pour les passages cités d'Aristote. La lemmatisation du texte latin est en cours. On désire aboutir de la sorte à une concordance latino-grecque. (Dans ce cadre, je signalerai aussi, qu'après la mort de Suzanne Mansion et divers avatars, on a poursuivi à Louvain-la-Neuve le traitement informatique des diverses éthiques d'Aristote. Les résultats pourront théoriquement en être publiés prochainement. Ceux-ci s'ajouteraient ainsi aux travaux complémentaires poursuivis à Liège par Christian Rutten sur la Métaphysique.)

- Thomas d'Aquin : si le projet Thomas d'Aquin qui a abouti à la monumentale publication de l'*Index Thomisticus*, n'a pas connu de développement nouveau, le Père Busa n'en a pas moins continué à publier plusieurs études. Celles-ci se trouvent mentionnées dans les *Studies in honour of Roberto Busa S.J.*, publiées en 1987 dans la revue *Linguistica Computazionale* en tant que volumes IV et V. Précisons que ces mélanges contiennent plusieurs contributions qui intéressent directement les médiévistes (notamment, en ce qui concerne des projets précis, la contribution de Jacqueline Hamesse sur *Le traitement automatique du Livre des Sentences de Pierre Lombard* - pp. 71-78 -, et celle de Paul Tombeur sur les *Banques de données constituées au Cetedoc pour l'étude de la tradition occidentale* - pp. 259-278 -).
- Thomas de Buckingham : les *Quaestiones super sententias*, second quart du XIV<sup>e</sup> siècle, projet de Arthur R. Lee, de Starke (Floride). Nous nous trouvons ici face à un projet d'ensemble pour une édition critique d'une oeuvre : transcription des textes sur des supports informatiques, comparaison des textes, application de la "cluster analysis" pour situer les variantes, sélection d'un texte de base, édition du texte, puis exploitations diverses, notamment du type concordance.
- Vincent de Beauvais : le *Speculum maius*. Maître d'oeuvre du projet : Madame M. Paulmier. Il faut renvoyer ici à la publication *Spicae*, ainsi qu'à un article de Madame Paulmier, dont le titre est suffisamment évocateur : *Ut diligens et intelligens lector requirat : Une banque de données du Speculum maius de Vincent de Beauvais*. Cet article a paru dans le numéro spécial de *Verbum* de 1985 qui s'intitule : *De la plume d'oie à l'ordinateur. Études de philologie et de linguistique offertes à Hélène Nais*. On trouvera, sous la rubrique "colloques", le rappel de l'intéressante table ronde qui fut organisée sur ce sujet à Nancy en 1985.

A la suite de cette liste d'auteurs, je mentionnerai un projet qui concerne une étude informatique portant sur les commentaires aux Sentences. Sous la direction du Père Busa, avec la consultance de M.T. Fumagalli (Università Statale de Milan) et de A. Ghisalberti (Università Cattolica de Milan), N. de Fernex, M. di Stefano, E. Colombo et L. Mappelli ont soumis à un traitement électronique les distinctions 42, 43, 44 du premier livre des commentaires aux Sentences de quatre auteurs : Jean Quidort, Duns Scot, Guillaume d'Ockham et Thomas de Strasbourg. Cette recherche a comme point de départ l'étude de E. Randi sur l'omnipotence divine. Le projet voudrait continuer ce travail en intégrant le plus grand nombre possible de textes écrits entre 1277 et 1347 et portant sur l'omnipotence divine.

### B) Les grands projets d'ensembles

1° Je citerai d'abord les travaux du *Lessico Intellettuale Europeo* (LIE) que dirige à Rome Tullio Gregory. Plusieurs articles ont présenté les activités du LIE ; je signale notamment pour les anglophones un article de A. Lamarra, *Computers and Philosophical Lexicography : the Activities of the Lessico Intellettuale Europeo*, paru dans le volume 16 de *Computers and the Humanities* de 1982.

Les travaux du LIE aboutissent notamment à la publication d'une prestigieuse collection. A la suite des publications que j'ai mises en évidence dans le rapport de Louvain-la-Neuve, je citerai ici les *Note di lessicografia ippocratica* de A. Bozzi, l'important *Lessico delle Novellae di Giustiniano* dû à A.M. Bartoletti Colombo, deux volumes parus, respectivement en 1983 et 1986, qui comprennent les lettres A à M, les *Indices chemicorum graecorum* de R. Halleux, paru en 1983, le colloque *Spiritus*, paru en 1984, le colloque *Francis Bacon. Terminologia e fortuna nel XVII secolo*, publié en 1984, le colloque *Grafia e interpunzione del Latino nel Medioevo*, dont les actes ont paru en 1987 grâce à A. Maieru ; cette collection comprend encore bien d'autres remarquables ouvrages, mais qui n'ont pas leur place sous cette rubrique.

Dans ce même mouvement, le LIE a fondé une nouvelle revue : *Lexicon philosophicum. Quaderni di terminologia filosofica e storia delle idee*, dans laquelle la part de l'informatique sera assurément croissante.

Le LIE continue d'oeuvrer pour un *Thesaurus mediae et recentioris latinitatis*, en accordant notamment une attention particulière au lexique de l'*Aristoteles latinus*.

Il apparaît que la volonté du LIE est d'aller, en partant des enregistrements de textes, des index et des concordances, des listes diverses de lemmes, vers la constitution de lexiques au sens propre du terme, décrivant le contenu sémantique et philosophique de chacun des termes retenus. Et c'est assurément ainsi qu'il faut progresser en la matière.

2° Lors du colloque de Louvain-la-Neuve, j'ai exposé le développement qu'avait pris le projet majeur du Cetedoc, celui qui s'intitule : *Thesaurus Patrum Latinorum*. Dans les *Studi Medievali* de 1984, Tullio Gregory nous a fait l'honneur de présenter ce projet, en notant la complémentarité des projets du LIE que je viens de citer, dans un article intitulé : *Instrumenta Lexicologica Latina. Verso un "Thesaurus Patrum Latinorum"*.

Ce projet comprend le traitement systématique par ordinateur des écrits des Pères latins et d'oeuvres du moyen âge :

- constitution d'une banque de données de l'ensemble des textes retenus ;
- constitution de diverses bases de données permettant l'interrogation en accès direct selon les procédures booléennes du "et" , du "ou" et du "non" ;
- constitution et publication des *Instrumenta Lexicologica Latina* et de *Thesauri* : *Series A - Formae* (*Enumeratio formarum, Concordantia formarum, Index formarum a tergo ordinatorum*) et *Series B - Lemmata* (*Enumeratio lemmatum, Concordantia lemmatum et formarum, Index formarum et lemmatum, Index lemmatum a tergo ordinatorum, Tabula frequentiarum*).

- publication de la collection *Informatique et étude de textes* aux éditions du Cetedoc.

En voici l'état automne 1987 :

\* *Thesauri*

- *Thesaurus Augustinianus* : relevé des formes, index inverse et concordance complète de toutes les oeuvres de saint Augustin (collaboration Augustinus Lexikon de Würzburg et Cetedoc). Publication de la *Series A* : début 1988. J'ajouterai ici que c'est notamment sur la base de l'enregistrement complet sur ordinateur des textes d'Augustin, qu'à commencé, depuis 1986, la publication de l'*Augustinus Lexikon*, que dirige, entouré de toute une équipe, le Père Cornelius Mayer.
- *Thesaurus sancti Gregorii Magni* : relevé des formes, index inverse et concordance complète de toutes les oeuvres de saint Grégoire le Grand répertoriées dans la *Clavis Patrum Latinorum*, n° 1708 à 1714 et n° 1719. Publication de la *Series A* : tome I : *Enumeratio formarum, Index formarum a tergo ordinatarum* ; tome II : *Concordantia formarum*, Brepols, 1986. Une première lemmatisation est dès à présent achevée.
- *Thesaurus Tertulliani* : tout le corpus des oeuvres est en mémoire.
- *Thesaurus sancti Ambrosii* : douze oeuvres sont actuellement mises en mémoire (collaboration Centre de littérature et de spiritualité de l'Université de Metz et Cetedoc).
- *Thesaurus sancti Hieronymi* : I. *Opera scripturistica* : tous les commentaires bibliques sont en mémoire. Actuellement on enregistre les lettres.
- *Thesaurus Pseudo-Dionysii Areopagitae* : le texte grec et toutes les traductions latines jusqu'à celle d'Ambrosio Traversari y comprise sont en mémoire. Les *Series A* et *B* seront publiées en 1988.
- *Thesaurus sancti Bernardi Claraevallensis* : la *Series A* de toutes les oeuvres de Bernard de Clairvaux vient de sortir de presse.

\* *Instrumenta Lexicologica Latina (ILL)*

Publication de la *Series A*, fasc. 1 à 41, publication de la *Series B*, fasc. 1 à 4, 6 et 11. Dernière publication : la *Series A* et la *Series B* des Sermons de saint Léon le Grand. Grâce à ces publications, on dispose notamment de toute une série de concordances pour diverses oeuvres de Raymond Lulle. D'autre part, doit paraître incessamment la concordance comparative des divers états de la *Theologia* d'Abélard, laquelle, comme on sait, n'a cessé d'évoluer au cours de sa vie : *Theologia scholarium*, *Theologia summi boni* et *Theologia christiana*.

\* *Corpus Christianorum, Series Latina*

Achèvement de la mise en ordinateur de toutes les oeuvres parues dans la *Series latina*, sauf les tomes 162 à 162c qui comprennent d'ailleurs des textes qui nous intéressent moins ici : le *Corpus benedictionum pontificalium*. De nombreuses concordances ont été réalisées au Cetedoc. Aux oeuvres du CC, il faut ajouter la mise en ordinateur des traductions latines du Pseudo-Dionysius Areopagita et des *Etymologiae* d'Isidore de Séville, de même que les oeuvres traitées dans la série

*Littérature arienne latine*, par R. Gryson, dans la collection *Informatique et étude de textes*.

\* *Corpus Christianorum, Continuatio Mediaevalis*

Mise en ordinateur de toutes les oeuvres parues. Au CC, CM, il faut ajouter particulièrement le traitement informatique de l'*Avicenna latinus*, des *opera omnia* de Bernard de Clairvaux et des *Libri sententiarum* de Pierre Lombard.

\* *Opera latino-belgica medii aevi*

Mise en ordinateur et concordance de toutes les oeuvres recensées dans l'*Index scriptorum operumque latino-belgicorum medii aevi*. *Nouveau répertoire des oeuvres médiolatines belges*, publié sous la direction de L. Genicot et P. Tombeur, Académie Royale de Belgique, t. I à III (VII<sup>e</sup> - XII<sup>e</sup> siècle), Bruxelles, 4 volumes, 1973-1979, y compris toute une série d'*addenda*.

Publication du *Thesaurus linguae scriptorum operumque latino-belgicorum medii aevi*. Première partie : *Le vocabulaire des origines à l'an mil*, 5 volumes publiés sous la direction de P. Tombeur, Académie Royale de Belgique, Bruxelles, 1986. Ce *Thesaurus* est le résultat de la lemmatisation intégrale de tous les textes de cette période.

Cette banque de données est par ailleurs enrichie de toutes les oeuvres traitées et publiées dans le cadre de la collection du Cetedoc *Informatique et étude de textes* : conciles oecuméniques, *Thesaurus Bonaventurianus*, *Corpus des sources franciscaines*, *Auctoritates Aristotelis*, etc.

\* *Traitements informatiques et études lexicologiques grecques*

Plusieurs oeuvres grecques ont été traitées ou sont en cours de traitement informatique (notamment les *opera omnia* de Grégoire de Nazianze, l'oeuvre du Pseudo-Denys, certaines oeuvres ayant paru dans la *Series graeca* du CC, ainsi que le corpus des Pseudépigraphes grecs d'Ancien Testament).

La réalisation de ces divers travaux comprend, outre l'élaboration des logiciels, la constitution et l'enrichissement progressif de dictionnaires automatiques permettant la lemmatisation de textes latins et grecs.

En dehors de ce cadre, se situent en outre les travaux arabo-latins de Mademoiselle Van Riet.

3° En ce qui concerne les textes enregistrés sur ordinateur et qui intéressent les médiévistes, je renvoie par ailleurs à l'*Oxford Text Archive*, qui publie un état annuel, et, pour le grec, au *Thesaurus Linguae Graecae* que dirige le Professeur Theodore F. Brunner à l'Université de Californie à Irvine. Le *Thesaurus linguae graecae* peut être actuellement acquis en CD-ROM - support dont nous parlerons *infra*.

c) *Études particulières*, où l'on parle de l'utilisation future de l'ordinateur, sans en préciser le rôle exact :

CIVICIMA (Comité International du Vocabulaire des Institutions et de la Communication Intellectuelle au Moyen Age) dont la secrétaire est Olga Weijers : ce comité désire établir un fichier informatisé du vocabulaire des écoles, des universités, du livre et de l'écriture, des méthodes, instruments et produits du travail intellectuel, des appellations des disciplines et de leurs étudiants. La limite chronologique est 1520. Signalons



qu'O. Weijers a publié en 1987 dans la collection du LIE un ouvrage intitulé *Terminologie des universités au XIII<sup>e</sup> siècle*.

### 3. Colloques et publication d'actes de colloques

Signalons d'abord les grands colloques généraux : ALLC (*Association for Literary and Linguistic Computing*) et ICCH (*International Conference Computers and the Humanities*). Il est utile, en effet, d'examiner les applications informatiques réalisées dans le domaine de l'étude des textes et de l'étude quantitative, quels que soient la période, la langue et le genre littéraire, et de se rendre compte dans quelle mesure on peut appliquer les mêmes programmes à des textes philosophiques médiévaux. Si beaucoup de colloques voient la publication de textes médiocres, voire insignifiants, si effectivement on assiste dans ce domaine, comme en d'autres, à une pollution étonnante, il importe néanmoins de voir ce qui se fait, ne fût-ce, parfois, que pour prendre le contre-pied.

- Citons d'abord le colloque de Liège de 1981, dont les actes ont été publiés en 1983 sous le titre : *Actes du Congrès International Informatique et Sciences humaines*.
- Juin 1982, colloque ALLC de Pise dont les actes ont été publiés en 1983 comme supplément au volume III de la revue *Linguistica Computazionale*, sous le titre *Computers in Literary and Linguistic Research*.
- Juin 1983, colloque de Metz dont les actes ont été édités chez Slatkine en 1983 : *La recherche française par ordinateur en langue et littérature* (publiés par C. Charpentier et J. David).
- Colloque ALLC 1984, Louvain-la-Neuve, publication des actes en 1985 par J. Hamesse et A. Zampolli : *Computers in literary and linguistic Computing. L'ordinateur et les recherches littéraires et linguistiques*.
- Victoria, Colombie britannique, 1984. Actes publiés à Montréal en 1985 par B. Delval et M. Lenoble : *L'ordinateur et la critique littéraire. Literary Criticism and the Computer*.
- ALLC 1985, colloque de Nice : *Méthodes quantitatives et informatiques dans l'étude de textes*, publié chez Slatkine par Etienne Brunet.

En dehors des colloques de l'association ICCH, j'attirerai particulièrement l'attention sur les colloques généraux consacrés aux Data-Bases :

#### *Conference in Data Bases in the Humanities and Social Sciences :*

- juin 1983, New-Brunswick,
  - juin 1985, Grinnel College, Iowa,
  - juillet 1987, Auburn University, Montgomery.
- Les actes du colloque de 1983 ont été publiés en 1985 par Robert Allen sous le titre *Data-Bases in the Humanities and Social Sciences*.
- Colloques généraux ou sections de colloques consacrés au thème moyen âge et informatique :
    - En ce qui concerne les symposiums de Kalamazoo, on a vu la publication en 1984, assurée par Anne Gilmour-Bryson, d'un volume *Computer Applications to Medieval Studies*.
    - En mai 1985, l'Université de Californie à Davis a organisé une *Conference on technological Aids to Medieval Studies*.

Dans un certain nombre de grands colloques ou de tables rondes, il fut question d'informatique : au colloque Guillaume d'Ockham en juin 1985 qui eut lieu à St. Bonaventure University, ainsi qu'à divers colloques organisés à Rome par le *Lessico Intellettuale Europeo* : le colloque *Spiritus* organisé en 1983 et publié en 1984, le colloque *Grafia e interruzione del latino nel Medioevo* organisé en 1984 et publié en 1987, le colloque *Imaginatio et Phantasia* organisé en janvier 1986 ; la réalisation et la publication des actes de ces colloques sont un des fleurons du LIE. L'informatique est intervenue également au grand colloque augustinien de septembre 1986, dont les actes viennent d'être publiés (*Congresso Internazionale su S. Agostino nel XVI centenario della conversione*).

Enfin, il faut signaler la table ronde organisée à Nancy en juillet 1985 sur *La Banque de données Vincent de Beauvais*.

On ne peut passer ici sous silence le colloque informatique consacré à ce qui sous-tend toute la pensée médiévale : l'Écriture. En 1985 eut lieu à Louvain-la-Neuve le 1<sup>er</sup> colloque *Bible et informatique*. Le prochain congrès consacré au même thème aura lieu à Jérusalem en 1988, la fin de la seconde semaine de juin, après le colloque ALLC qui aura lieu dans la même ville.

Une table ronde, dont la lecture des actes est particulièrement à recommander du point de vue méthodologique, fut celle organisée à Pise par l'*European Science Foundation* - la *Fondation Européenne de la Science* en 1981 : les actes ont été publiés dans le vol. III, année 1983 de la revue *Linguistica Computazionale*, sous le titre *The possibilities and Limits of the Computer in producing and publishing Dictionaries*. La même association a organisé en 1986 un colloque à Saarbrücken sur le thème *Standardization in computerized Lexicography*.

Sans doute n'est-il pas inutile d'attirer également l'attention sur des colloques informatiques organisés par des historiens ; plusieurs des sujets abordés sont en effet susceptibles d'intéresser plusieurs d'entre nous dans le cadre de leurs recherches.

Je mentionnerai d'abord les tables rondes consacrées à la création, la connexion et l'utilisation de banques de données dans les disciplines historiques : *International Workshop on the creation, connection and usage of large-scale interdisciplinary source banks in the historical disciplines*. L'animateur de ces tables rondes est Manfred Thaller du Max-Planck-Institut für Geschichte à Göttingen : 1985 : Göttingen ; 1986 : Graz ; 1987 : Paris. Cette troisième table ronde organisée par Jean-Philippe Genet, avait pour titre : *Troisième Table Ronde Internationale sur la Standardisation et l'Échange de Données Informatisées dans les Disciplines Historiques*. Si beaucoup de projets présentés se situent en dehors de l'étude des textes et de ce qui s'y rattache - notamment de nombreux projets en matière démographique -, il est une question importante qui fut discutée lors de ces réunions, c'est le problème des échanges de données. A Paris, particulièrement, on aborda, mais trop brièvement, faute de temps, un aspect fondamental : l'aspect juridique concernant les banques de données textuelles et leur reproduction.

Je signalerai également la création d'une association nouvelle *History and Computing*. En mars 1986 eut lieu au Westfield College à Londres la conférence inaugurale de cette association, et un second colloque fut organisé en mars 1987.

Un domaine d'application qui nous intéresse particulièrement, afin de pouvoir en utiliser les résultats pour nos recherches, est celui de la prosopographie.

En décembre 1982 eut lieu à Bielefeld un colloque dont les actes ont été publiés à Kalamazoo en 1986 sous le titre : *Medieval Lives and the Historians. Studies in Medieval Prosopography. Proceedings of the first International Interdisciplinary Conference on Medieval Prosopography*. On y trouve notamment, dû à la plume de Ralph Evans, *The Analysis by Computer of A.B. Emden's Biographical Registers of the Universities of Oxford and Cambridge*.

A Paris eut lieu, en octobre 1984, un colloque dont les actes ont été publiés en 1985 par Hélène Millet sous le titre : *Informatique et prosopographie*.

Signalons également ici les travaux de J. Wollasch et de ses collaborateurs. Citons la publication par J. Wollasch, en 1982, de la *Synopse der Cluniacensischen Necrologien*.

#### 4. Évolution du matériel informatique

##### 1) La micro-informatique

Depuis le dernier congrès, on a assisté à un développement extraordinaire des micro-ordinateurs. Leur taille mémoire et leur capacité en périphériques n'ont cessé de se développer. Vient d'apparaître sur le marché une nouvelle gamme de micro-ordinateurs, comme l'IBM PS/2 : dans la distinction qu'il importe d'opérer entre les maxi-ordinateurs (ou *main-frames*), les mini-ordinateurs et les micro-ordinateurs, on peut se demander dans le cas des PS/2 s'il s'agit de supermicros ou de mini-ordinateurs.

Le problème des chercheurs et des centres de recherches est de savoir ce qu'il faut acquérir aujourd'hui. Contentons-nous de souligner qu'il faut être attentif en tout premier lieu à la compatibilité et songer que l'avenir (et déjà le présent) réside dans la télématique : dans un système de réseaux d'ordinateurs connectés les uns aux autres.

##### 2) L'évolution des systèmes d'entrée, de sortie et de mémorisation

###### A. Entrée de données

On dispose aujourd'hui de la possibilité de procéder à l'entrée manuelle de n'importe quel type de caractère. Pour les micro-ordinateurs, existe à cette fin la "carte Hercules".

Une firme comme COS ("Computer Optimierungs-Systeme") a conçu des claviers "multiples" : les caractères pouvant apparaître sur les touches du clavier correspondent aux différents jeux de caractères qui ont été mis en mémoire. On peut ainsi disposer de touches affichant tantôt l'alphabet grec, par exemple, tantôt l'alphabet latin ou tout autre alphabet mis en mémoire.

La lecture optique des données se répand de plus en plus. On connaît depuis longtemps le système Kurzweil. Actuellement, de nombreux systèmes de lecture optique sont proposés pour des micro-ordinateurs, ain-

si, par exemple, pour citer un système parmi beaucoup d'autres, celui mis au point par la société française Inovatic. L'appareil lui-même coûte de l'ordre de 30.000 francs français, mais il faut acquérir en outre le logiciel de reconnaissance des caractères ; selon le développement de ces logiciels, les prix varient de 26 à 72.000 francs français. Il faut cependant mettre en garde les chercheurs qui auraient tendance à croire en une saisie optique presque infaillible. L'homme a de la peine à imaginer la différence existante entre "les yeux" du lecteur optique et ses yeux à lui : ce qui est clair à ses yeux - parce qu'il lit en situant sémantiquement les caractères lus - peut être mal lu par la machine qui, elle, n'a pas d'intelligence. Nous avons pu constater que dans certains cas le taux d'erreurs était incroyablement élevé. Le nombre d'erreurs de lecture dépend grandement de la compétence de l'opérateur et de sa capacité d'attention. De plus, jusqu'il y a peu, il fallait pouvoir disposer de deux exemplaires du texte à lire : l'un soumis à la machine, l'autre se trouvant dans les mains de l'opérateur, lui permettant ainsi de donner à la machine les instructions adéquates, sans travailler à l'aveugle. Dans les versions récentes, le scanner optique travaille en plusieurs phases distinctes : au cours de la première, il met le document en mémoire ; à partir de ce moment, l'opérateur peut reprendre en main le document enregistré et faire procéder à l'identification des caractères - les logiciels actuels travaillent sans police de caractères préétablie et permettent dès lors l'identification de n'importe quel caractère - ; au cours de la troisième phase a lieu la reconnaissance automatique. La qualité typographique des documents soumis est déterminante pour une résolution efficace et productive. Ajoutons cependant que, lors de la reconnaissance de caractères, il suffit que l'opérateur se trompe de touche - comme lors de toute saisie manuelle - pour qu'une erreur devienne systématique. La publicité en la matière est une chose, l'expérience en est une autre...

Pour faciliter la lecture de manuscrits ainsi que la reconnaissance des tracés de lettres, signalons l'aide que peuvent apporter de simples scanners optiques connectés à un micro-ordinateur disposant en outre d'une imprimante au laser. On peut ainsi, par exemple, avec un système Canon, faire apparaître sur l'écran d'un terminal la copie d'une page d'un manuscrit, isoler, par exemple, un mot, une lettre ou un groupe de lettres, les agrandir autant que l'on veut, puis imprimer le résultat sur papier.

#### *b. Sortie des données*

Les imprimantes au laser se sont développées et diversifiées, particulièrement en ce qui concerne les systèmes d'impression connectables à des micro-ordinateurs. Ces imprimantes permettent une qualité d'impression remarquable. Le coût en est cependant relativement élevé.

Il faut noter en outre le système COM, "Computer Output Microfiche". Ce système, qui existe depuis de nombreuses années, permet de réaliser automatiquement des microfiches à partir d'une bande magnétique. On peut ainsi rassembler de vastes concordances sur microfiches et permettre l'accès à l'information, sans aucune contrainte informatique. Je signalerai ici une nouveauté : la réalisation automatique de microfiches contenant des textes grecs. Le Cetedoc a réalisé ainsi une concordance et des index pour les *Pseudépigraphes grecs d'Ancien Testament*.

### c. Mémorisation

Le problème pour l'étude informatique de vastes ensembles de données textuelles a toujours été celui de la place mémoire nécessaire. Se développe de plus en plus aujourd'hui le système des mémoires de masse. Les micro-ordinateurs disposent, quant à eux, de disques durs qui peuvent comprendre plusieurs dizaines de millions de caractères.

Un nouveau type de mémoire se développe de plus en plus : le CD-ROM ; le "Compact Disk-Read Only Memory". La capacité d'un tel disque est de l'ordre de 550 millions de caractères. Comme le nom l'indique, ce disque peut être uniquement lu : on ne peut y écrire et donc y modifier une information qui serait, par exemple, erronée. Pour utiliser un CD-ROM, il faut avoir un micro-ordinateur de 526K au moins, un lecteur de Compact Disk et un interface qui relie l'un et l'autre. Chaque CD-ROM est livré avec un logiciel d'interrogation des données. Les capacités de ce logiciel varient d'un type à l'autre. Plusieurs encyclopédies et dictionnaires se trouvent déjà sur CD-ROM, de même, notamment, qu'un vaste corpus de textes grecs (le *Thesaurus Linguae Graecae* de Théodore Brunner).

Se développe, d'autre part, le système "WORM", "Write once read multiple" : il s'agit d'un disque optique d'une contenance de 200 millions de caractères que l'utilisateur peut créer lui-même.

A partir de nouvelles formes de mémorisation, on entrevoit à quel point les instruments de recherches textuels et lexicaux sont appelés à se modifier dans les années à venir au grand bénéfice des études médiévales.

Quant à l'évolution du logiciel, il n'y a pas lieu d'en parler ici. Qu'il nous suffise d'attirer l'attention sur certains problèmes méthodologiques.

## 5. Problèmes méthodologiques

Dégageons tout d'abord les grandes perspectives qui se présentent : la constitution, d'une part, de banques et de bases de données textuelles, et celle, d'autre part, de banques et de bases de données lexicales. Les banques de données rassemblent un ensemble d'informations dont on peut tirer ce que l'on désire, selon les programmes auxquels on fait appel ; les bases de données sont des ensembles d'informations structurées de telle manière qu'on peut les interroger directement selon les procédures booléennes du "et", du "ou" et du "non" afin d'obtenir, en accès direct, les réponses souhaitées.

L'ensemble des perspectives est clairement présenté par Bernard Quemada dans *The possibilities and Limits of the Computer in producing and publishing Dictionaries*, dont le texte a été publié dans le troisième volume de la revue *Linguistica Computazionale* (1983).

D'un point de vue plus général, je signale l'ouvrage de Mario Borillo, *Informatique pour les sciences de l'homme*, Bruxelles, 1984.

Dans les Actes du congrès de Stockholm, d'août 1984, *The Editing of Theological and Philosophical Texts from the Middle Ages*, parus en 1986 dans les *Studia Latina Stockholmiensia*, je me permets de signaler

ma contribution *Latinité et informatique. : travaux réalisés par le CETE-DOC. Perspectives et implications méthodologiques.*

### 1) Les éditions critiques

Les implications méthodologiques ont notamment trait à tout ce qui touche à l'établissement et à la publication d'éditions critiques. Notons tout d'abord le nouveau type de relation qui s'est instauré entre le chercheur et une maison d'édition. Il faut souhaiter que, dès le point de départ, le chercheur réalise une copie du manuscrit de base sur un support magnétique. Cet enregistrement sera modifié au fur et à mesure des collations. En finale, il enverra à l'éditeur non plus des pages dactylographiées, mais des disquettes, ou éventuellement une bande magnétique qui rassemble l'ensemble des disquettes. Ces supports magnétiques constituent l'entrée permettant la photocomposition du texte, ou, éventuellement, une impression au laser qui sera reproduite par offset.

Pour les problèmes de critique textuelle, rappelons le volume d'actes paru en 1979 aux éditions du C.N.R.S. sous le titre *La pratique des ordinateurs dans la critique des textes*. En 1984 a paru à Rome, sous la direction d'A. D'Agostino, *La critica dei testi latini medievali e umanistici*. Il s'agit là d'une traduction italienne d'un intéressant volume paru précédemment en allemand, où on trouve notamment, du point de vue qui nous occupe, des contributions de J. Mau et de W. Ott. Je renvoie également au volume publié à Bruxelles en 1986 par Ghislaine Viré, *Informatique et classement des manuscrits. Essai méthodologique sur le De astronomia d'Hygin*.

D'un point de vue tout à fait pratique, quantité de réalisations voient le jour dans le domaine de l'utilisation de micro-ordinateurs pour l'établissement d'éditions critiques. Voyez notamment Yves Chartier dans le n° 16 de la revue *Le Médiéviste et l'Ordinateur* (automne 1986). Le numéro de juillet 1986 de la revue *PC World* est consacré à l'édition réalisée à l'aide d'un micro-ordinateur. Signalons également la parution d'une nouvelle revue *Publish!* qui désire traiter de tout ce qui est matériel et logiciel en matière de publication (B.P. 55400, Boulder, CO 80322 - 5400, USA). J'ajouterai que le Cetedoc a également effectué plusieurs réalisations dans ce domaine, en continuant dans la ligne des travaux présentés dans le volume du C.N.R.S. paru en 1979.

### 2) L'analyse lexicale

En ce qui concerne l'analyse lexicale, je me contenterai d'insister, une fois de plus, sur la nécessité d'établir un métalangage dans ce domaine, afin de sortir de l'état d'anarchie que nous connaissons, chacun établissant ses propres règles, ce qui empêche toute élaboration d'ensemble. Le Cetedoc s'est tout particulièrement attaché à ce travail et je me permets de signaler les règles d'analyse lexicale que j'ai présentées dans l'introduction méthodologique du *Thesaurus Linguae Scriptorum Operumque Latino-Belgicorum Medii Aevi. Le vocabulaire des origines à l'an mil*, paru en 1986. Le type de lemmatisation qui est présenté est systématiquement appliqué à toute la banque de données constituée et nous constatons que d'autres centres, tel l'A.R.T.E.M. de Nancy (Atelier de Recherche sur les

Textes médiévaux et leur Traitement assisté), ont décidé de s'y conformer pour l'analyse lexicale des textes médiévaux traités. A partir des règles définies, des dictionnaires automatiques ont été constitués, lesquels permettent d'automatiser le plus possible la réalisation de la lemmatisation des textes.

Le système de lemmatisation mis en place pour le latin - lequel a dès à présent été transposé pour d'autres langues, tels le grec et le français -, est basé sur la structure linguistique du latin et demeure valable pour toutes les époques de la *latinitas*, ce qui permet de faire des comparaisons entre les diverses époques. Un lemme rassemble toutes les variantes orthographiques d'une même forme ainsi que toutes les formes grammaticales dérivées d'un même vocable. Une telle lemmatisation n'opère dès lors pas de distinction selon les critères morphologico-syntaxiques ou sémantiques, hormis, bien entendu, le cas des homographes, dont la réalité est fixée selon l'étymologie. Les règles mises au point constituent le seul moyen d'avoir un langage d'analyse commun. Toute distinction morphologico-syntaxique ou sémantique pour un même lemme est reléguée au second plan : ces distinctions peuvent ainsi constituer autant de sous-lemmes.

### 3) Les analyses statistiques

Dernier point qu'il faut évoquer : celui des analyses statistiques. Si les mots sont "dangereux", les nombres le sont encore bien davantage. Le problème n'est pas seulement celui des éléments servant de base à une analyse statistique, mais également celui du choix des outils adéquats. Je signalerai à ce propos les volumes 27 et 29 de la collection du *Lessico Intellettuale Europeo*, le premier consacré, en 1982, à *L'analisi delle frequenze. Problemi di lessicologia*, publié par les soins de M. Fattori et de M. Bianchi et contenant des contributions de U. Berni Canani, E. Brunet, R. Busa, G.Th. Guilbaud et J. Hamesse, le second édité en 1982 par R. Busa et contenant les actes d'un séminaire qui eut lieu à Gallarate en 1981 sous le titre *Global Linguistic Statistical Methods to locate style identities*. Notons en outre les travaux de Ch. Muller et la collection qu'il dirige chez Slatkine sous le titre *Travaux de Linguistique quantitative*. Signalons enfin l'apparition en 1986 d'une nouvelle revue parisienne : *Histoire et mesure*.

On peut construire de merveilleuses grandes orgues : encore faut-il que l'on dispose de musiciens capables de les jouer vraiment ! C'est bien là le souhait que je formulerai à la fin de ce rapport : puissions-nous allier aux moyens techniques dont nous disposons, la lecture et la connaissance des textes, l'érudition des Du Cange, des Mauristes et de la plupart des hommes de science du XIX<sup>e</sup> siècle ! Alors nous pourrons faire merveille et oeuvrer pour une science exacte des textes qui permette de mieux cerner faits et hypothèses.

Paul TOMBEUR (Louvain-la-Neuve)