

**LA CONSTRUCTION DE BANQUES
DE DONNÉES DOCUMENTAIRES
AU CENTRE DE RECHERCHE TAAC
(CNRS et Université de Paris X)**

Anne-Marie GUIMIER-SORBETS

Le Centre de Recherche sur les Traitements Automatisés en Archéologie Classique (TAAC) est l'une des composantes du laboratoire associé 375, CNRS et Université de Paris X, dirigé par le professeur René Ginouvès. Son activité, située à la rencontre de la réflexion théorique et des applications pratiques, porte sur les possibilités d'exploitation informatisée des données archéologiques, soit pour la gestion ou les recherches comparatives dans les banques de données documentaires, soit pour des constructions mettant en jeu le calcul et la statistique dans les banques de données de structuration, visant à la classification ou à la sériation automatique, etc.

Mais, en amont du traitement informatisé en lui-même, il est nécessaire de contrôler la représentation des données archéologiques, et c'est là le problème majeur de la "constitution des données", sur lequel nous avons fait porter dès les débuts une grande part de nos efforts (cf. A.M. Guimier-Sorbets et R. Ginouvès, *La constitution des données en archéologie classique*, CNRS, 1978). Depuis, les chercheurs du Centre participent, en collaboration avec des spécialistes français et étrangers, à la préparation de systèmes de représentations linguistiques contrôlés et multilingues, visant d'abord à normaliser la langue de description traditionnelle, ensuite à fournir des descripteurs pour les traitements automatisés.

Ce travail porte sur plusieurs champs de l'archéologie classique : l'architecture, où R. Ginouvès et R. Martin, en collaboration avec l'Institut de recherche d'architecture antique, dirigé par P. Gros, préparent un *Dictionnaire méthodique d'architecture grecque et romaine*, dont le tome I, *Matériaux, techniques de construction, techniques et formes de décoration*, a paru en 1986 ; deux autres tomes sont en préparation, l'un sur les éléments formels de la construction (murs,

colonnes, voûtes et arcs, toitures), l'autre sur les fonctions des bâtiments, dans l'architecture religieuse, militaire, civique, etc. Le travail de normalisation de la langue porte aussi sur les formes de vases, où l'on prépare les règles du découpage et de caractérisation dans un groupe de travail comportant Fr. Villard, L. Jehasse, P. Rouillard et d'autres spécialistes ; et aussi sur l'analyse du décor de la mosaïque, en collaboration avec le Centre de Recherche sur la mosaïque, qui est une autre composante du laboratoire 375 ; C. Balmelle, M. Blanchard, J. Christophe, J.P. Darmon, A.M. Guimier-Sorbets, H. Lavagne, R. Prudhomme, H. Stern ont fait apparaître, en 1986, le premier tome d'un travail intitulé *Décor géométrique dans la mosaïque romaine*, Paris, consacré aux compositions linéaires et de surface ; un second tome, consacré aux compositions centrées, est en préparation.

La réalisation progressive de ces instruments de travail, et la réflexion sur l'adaptation des systèmes informatiques aux besoins des archéologues, ont permis le lancement de toute une série d'opérations, soit dans le domaine de la structuration (études menées en collaboration avec V. Tran Tam Tinh, Université Laval, Québec, sur les images de Sérapis ; ou, en collaboration avec H. Van Effenterre et l'association internationale du *Corpus de minoischen und mykenischen Siegel*, sur les collections d'intailles minoennes, ou, en collaboration avec J. Coulton, Université d'Oxford, sur le profil des chapiteaux doriques), soit dans le domaine des banques de données documentaires, dont il sera maintenant seulement question ici, car elles sont, à l'heure actuelle, au cœur de l'activité de notre laboratoire. Dans cette perspective, nous travaillons sur des banques de données documentaires soit factuelles (portant sur les documents primaires eux-mêmes, par exemple des mosaïques, des inscriptions...) soit référentielles (portant sur des documents secondaires par rapport aux objets archéologiques, par exemple photographies, dessins et plans, estampages d'inscriptions ou moulages de monnaies, archives de toute sorte) ; certaines de ces banques sont constituées dans notre centre, qui est ainsi producteur de banques de données documentaires archéologiques ; d'autres sont conçues et mises au point chez nous, à la demande d'organismes français et étrangers qui en assurent ensuite la constitution et l'utilisation, et en cela notre Centre joue le rôle de laboratoire-conseil.

C'est ainsi que nous avons constitué, dans notre Centre, - une banque de données factuelles sur la *Mosaïque dans le monde grec, des origines à la fin de l'époque helléniste*, qui regroupe une information très

détaillée sur un nombre assez restreint de mosaïques dispersées entre l'Italie, la Grèce propre, l'Égypte, la Turquie et jusqu'en Afghanistan,

- une banque de données référentielles, *Images de la Mosaïque*, automatisant la consultation de la riche photothèque du Centre de recherche sur la mosaïque, qui appartient lui aussi, comme on l'a vu, à notre laboratoire. Ici les documents, beaucoup plus dispersés encore dans le temps et dans l'espace, sont extrêmement nombreux et leur masse ainsi que la qualité "référentielle" de la banque, nous ont conduits à une analyse beaucoup moins fine que pour la banque factuelle sur la mosaïque grecque.

Mais c'est pour des laboratoires extérieurs à notre Centre que nous avons conçu le plus grand nombre de banques, qu'ensuite nous mettons en place, mais dont la réalisation, et évidemment l'utilisation, sont sous la seule responsabilité des organismes qui ont fait appel à notre aide. On citerait ainsi :

- pour le Centre de Documentation photographique et photogramétrique (CDPP) de l'Université de Paris I et du CNRS, *Images du monde méditerranées antique*, une banque de données référentielle, automatisant la consultation de son importante photothèque : la responsabilité de sa constitution appartient à M. Fourmont.

- pour l'École française d'Athènes, plusieurs banques sont installées ou en cours de mise au point : c'est le cas pour une banque factuelle sur les *Vases conteneurs - amphores et lagynoi - dans le monde grec et romain*, sous la responsabilité de J.Y. Empereur, secrétaire général de l'École ; puis une banque, factuelle ici encore, sur les *Monnaies de Thasos*, sous la responsabilité du Directeur de l'École, O. Picard ; et aussi une grosse banque de données référentielle, *Archives de l'EFA* sur la documentation, très riche, accumulée dans cette école depuis sa fondation, avec plusieurs centaines de milliers de photographies, des estampages d'inscriptions, toute une série d'archives écrites, allant de la correspondance aux carnets de fouilles, etc.

- pour la Fondation nationale hellénique de la recherche scientifique, laboratoire KERA, à Athènes, nous avons installé une banque de données factuelles sur les *Inscriptions de Macédoine*, sous la responsabilité de M. Hatzopoulos. Cette banque a la particularité d'être bilingue, la documentation y étant enregistrée et pouvant être interrogée et éditée indifféremment en grec ou en français, et évidemment avec des caractères latins et grecs.

Toute une série d'autres opérations sont en préparation ou en projet, aussi bien pour l'Université de Strasbourg (banque de données référentielles sur une photothèque de documents byzantins) que pour l'École française de Rome et que pour le Service archéologique de Jordanie, qui nous a demandé notre aide pour gérer son énorme documentation.

Toutes ces banques de données, dont l'information présente une plus ou moins grande finesse d'analyse, selon qu'elles sont factuelles ou référentielles, possèdent un point commun : c'est que leurs analyses ne peuvent être seulement constituées par des listes de descripteurs ; ces derniers doivent pouvoir être reliés, de manière que soit prise en compte la structure interne d'un document. Pour une mosaïque, par exemple, il est nécessaire de rattacher, à chacune des parties de son "découpage" (tapis, panneaux, bordures de chaque tapis ou panneau, etc.), les informations relative à leur décor et à leur technique. Et, d'un autre côté, il est impossible de prédéclarer la structure de l'ensemble des documents d'une banque, en raison de la diversité de composition des objets archéologiques : pour en revenir à l'exemple des mosaïques, il est bien évident que chacune présente son propre type d'emboîtement des parties, sa spécificité dans le nombre des parties à chaque niveau d'emboîtement, etc.

C'est pour résoudre ce problème, essentiel dans l'analyse du document archéologique, que après expérimentation d'autres systèmes, nous avons choisi le logiciel SIGMI, conçu par le Centre d'Automatique et d'Informatique de l'École des Mines de Paris, sous la direction de M. Lenci. En effet ce logiciel, qui offre toutes les fonctions d'un système documentaire classique, présente en outre une syntaxe originale : un jeu de parenthèses sert à exprimer, à l'entrée des documents, l'emboîtement des différentes informations ; et la richesse des opérateurs d'interrogation permet de retrouver les différentes relations ainsi créées. Le système SIGMI fonctionne en traitement par lots, sur la gamme IBM 370 ; c'est avec lui que nous avons exploité les premières banques de données que nous avons mises en place. Mais nous utilisons désormais, produit par le même centre, le logiciel SIGMINI qui, lui, est destiné à la gamme des mini-ordinateurs, et même aux plus puissants des micros ; ce logiciel fonctionne cette fois en conversationnel, avec le système d'exploitation UNIX, et il peut prendre en charge directement toutes les banques préparées pour le système SIGMI. Grâce à ces deux systèmes, nos banques peuvent être implantées sur des gammes de matériels très différents et cela va nous permettre d'accroître leur diffusion. On notera

d'ailleurs que, si les analyses destinées aux différentes banques ont été évidemment conçues en fonction de la nature et des particularités de chaque type de document, elles mettent en jeu, en même temps, des formules compatibles, afin que soient possibles si on les souhaite, des échanges d'information entre les différents centres de recherche archéologiques : c'est le cas, en particulier, pour les trois photothèques du Centre de recherche sur la mosaïque, du Centre de documentation photographique et photogrammétrique et de l'École française d'Athènes. On rend ainsi possible la création de "réseaux", à l'échelle nationale et à l'échelle internationale, et c'est pourquoi le Conseil de l'Europe, groupe PACT, nous a confié la mise en place d'un réseau sur les "banques d'images en archéologie", dont la première fonction sera l'échange d'informations sur les Centres producteurs d'images archéologiques, puis, éventuellement, l'échange des images elles-mêmes.

C'est précisément pour améliorer, d'une part, les conditions de consultations des banques documentaires, et, d'autre part, leur diffusion, que nous avons fait appel à une technique encore exceptionnelle dans notre domaine, celle du vidéodisque. En effet, en avril 1986, nous avons présenté à Paris le premier vidéodisque, "Images de l'archéologie", couplé à trois banques de données que nous avons construites ou aidé à construire. Il s'agit d'un "vidéodisque optique analogique à laser", sur lequel sont enregistrées environ 40 000 images qui illustrent d'une part la banque de données factuelle sur la *Mosaïque dans le monde grec* et, d'autre part, les deux banques référentielles *Images de la Mosaïque* et *Images du monde méditerranéen antique*, trois banques de données conçues, on l'a vu, selon des formules compatibles.

Ainsi, chacune de ces images enregistrées sur le vidéodisques peut être appelée directement et immédiatement par son numéro d'ordre, en utilisant une commande manuelle ; mais l'utilisation la plus riche du système consiste à coupler le vidéodisque avec les banques de données, ce qui permet d'appeler les documents par une question posée à l'ordinateur ; la réponse est constituée d'une part par un texte, qui, apparaissant sur un écran alphanumérique, décrit le document ou les documents pertinents, et d'autre part, sur un écran type télévision, par une ou plusieurs images pour chacun de ces documents. C'est encore l'École des Mines de Paris qui a réalisé le logiciel de couplage, selon nos spécifications, et le micro-ordinateur spécialisé permettant de récupérer l'information des banques, situées à distance sur une machine du CIRCE, Centre de calcul universitaire de la région parisienne, d'afficher les descriptions sur un écran, et de piloter en local, le vidéolecteur qui envoie

des images sur un second écran. Avec SIGMINI, la consultation de la banque et des images se fait entièrement en local et selon les mêmes modalités. Pour chaque document (qui peut occuper plusieurs "pages" d'écran) il existe normalement plusieurs images (plusieurs photographies et plans pour une même mosaïque grecque, plusieurs agrandissements ou cadrages différents réalisés à partir de la même diapositive pour les photothèques); le logiciel permet à l'utilisateur une recherche interactive, avec défilement des images selon un rythme choisi, retour en arrière, etc., la correspondance du texte et de l'illustration restant toujours automatiquement assurée. Ce vidéodisque, qui va permettre d'élargir la diffusion de banques de données archéologiques en les enrichissant de leurs images, constitue en même temps l'aboutissement et le point de départ de toute une série de recherches sur les techniques de l'information scientifique.

Anne-Marie GUIMIER-SORBETS
Responsable du Centre de Recherche
sur les traitements automatisés
en Archéologie Classique
(CNRS. Université de Paris X)