

UN CENTRE DE RECHERCHES SUR LES SYSTÈMES D'INFORMATION EN ARCHÉOLOGIE

C'est en 1969 que R. Ginouvès, nommé l'année précédente professeur à l'Université de Paris X, créait, dans le cadre du Centre de Recherche en archéologie classique de cette Université, un groupe de travail tourné vers les applications de l'informatique à sa discipline: dès cette époque en effet on ressentait de plus en plus vivement le besoin de répondre à l'“explosion documentaire” par une mécanisation de l'information, et d'autre part de pratiquer dans nos “sciences de la culture” une démarche qui les rapproche des sciences de la nature. Les premiers travaux s'ordonnaient en vue de la préparation de banques de données, et assez vite ils étaient soutenus par les responsables de la recherche scientifique en France, et liés au CNRS dans la cadre d'une formation intitulée Centre TAAC, “Centre de recherche sur les Traitements Automatisés en Archéologie Classique”, placé depuis maintenant plusieurs années sous la responsabilité d'A.-M. Guimier-Sorbets.

Il est remarquable que, pendant une assez longue période, la recherche ait porté dans ce Centre essentiellement sur les problèmes, à la fois théoriques et pratiques, de la construction de systèmes descriptifs: il était apparu comme évident, dès les débuts de la recherche, que la mécanisation de l'information n'est pas compatible avec les formes habituelles de la description archéologique, dont on sait qu'elle manque, à tout le moins, de la régularité indispensable. C'est pourquoi nous avons publié, en 1978, un petit livre sur *La constitution des données en archéologie classique*: il expose le résultat de nos premières réflexions sur les deux questions fondamentales, s'agissant d'objets (ou de configurations) matériel(le)s dont il faut donner la traduction verbale sur laquelle se feront les tris et les calculs: d'une part que faut-il dire (délimitation du corpus, détermination de l'unité documentaire, systèmes de segmentation, types et échelles des caractéristiques, finesse de l'analyse, etc.); d'autre part comment faut-il le dire (avec la construction de systèmes linguistiques susceptibles de porter cette information)?

Et effectivement notre laboratoire entreprenait, dès cette époque, la réalisation d'ouvrages destinés à régulariser, même en dehors de toute application informatique, le langage de la description. C'est ainsi que R. Ginouvès et R. Martin mettaient en chantier, avec toute une série de collaborateurs français et étrangers, un *Dictionnaire méthodique de l'architecture grecque et romaine*, dont le premier volume, consacré aux *Matériaux, techniques de la construction, techniques et formes du décor*, paru en 1985, sera suivi d'un volume sur les *Éléments*

constructifs: supports, couvertures, aménagements intérieurs, actuellement sous presse, et d'un troisième sur les *Plans et compositions architecturales*, en préparation. Une particularité de ces ouvrages (comme d'ailleurs des autres dont il va être question) est qu'ils donnent, à la suite des mots avec leur définition, leur traduction en allemand, anglais, italien, grec moderne (et aussi leurs équivalents en grec ancien et en latin, lorsqu'on les connaît): la comparaison des champs couverts par les termes correspondants, d'une langue à l'autre, permet de vérifier la manière dont ils segmentent la réalité, et par là d'en mieux apprécier la place dans leur environnement sémantique. Dans un autre champ de la recherche, notre équipe, avec la collaboration de Fr. Villard et de plusieurs autres céramologues, prépare un ouvrage sur *La forme des vases grecs, principes d'analyse et vocabulaire multilingue*, qui propose un vocabulaire normalisé des parties du vase en diverses langues (parmi lesquelles, cette fois, aussi l'espagnol), avec des règles de découpage (comment reconnaît-on s'il y a un COL, et ses limites?), et une explicitation des critères permettant de distinguer ce que les céramologues appellent des "formes" (l'AMPHORE, le CRATERE) avec les principales variétés qu'on en reconnaît. Beaucoup d'attention aussi a été portée à la description de la mosaïque, en particulier en collaboration avec le Centre de recherche sur la mosaïque appartenant à la même formation Paris X-CNRS; c'est ainsi qu'a été publié en 1985 un volume intitulé *Le décor géométrique de la mosaïque romaine*, préparé par C. Balmelle et l'ensemble de ce Centre avec des collaborateurs étrangers, qui présente en 254 planches un très grand nombre de motifs dont la description normalisée est donnée en cinq langues: il concerne les compositions linéaires et de surface, et un second volume, consacré aux compositions centrées, est en préparation; d'un autre côté, la réalisation de certaines banques de données, dont il va être question, a conduit A.-M. Guimier-Sorbets à mettre au point non seulement un vocabulaire, mais aussi des structures d'analyse adaptées à la mosaïque, et à toute une série d'autres documents.

En même temps qu'à ces problèmes de "constitution des données", qui se situent en quelque sorte en amont de toute application informatique (et aucune informatisation sérieuse n'est imaginable avant leur résolution préalable), notre Centre s'est évidemment intéressé aussi très vite à la réalisation pratique de banques de données, et d'abord à la recherche de logiciels permettant de prendre en compte la documentation archéologique avec au moins le même degré de richesse que la publication traditionnelle, mais avec en plus les avantages qu'apporte la mécanisation, — plasticité de la description qu'il est toujours possible de corriger et d'enrichir, et rapidité de la consultation pour des questions aussi complexes qu'on le souhaite. C'est pourquoi nous nous sommes attaqués dès le début à un problème fondamental: celui de relier les informations, sans risque de croisement, à des parties distinctes du document, parties éventuelle-

ment emboîtées les unes dans les autres (comme, pour une mosaïque, la bordure à bandes multiples d'un panneau inclus dans un autre panneau comportant lui aussi une bordure complexe, et lui-même inclus dans un tapis bordé), alors que chaque document d'un même corpus peut présenter une structure différente. Fort heureusement, nous avons été orientés vers l'Ecole des Mines de Paris, dont les besoins descriptifs sont, malgré les premières apparences, de même nature (car l'analyse d'une coupe géologique implique la reconnaissance d'inclusions qui peuvent en enfermer elles-mêmes, sans qu'un schéma convenant chaque fois puisse être défini a priori). Et c'est ainsi que nous avons été conduits à mettre en œuvre le logiciel SIGMINI, logiciel "autostructurant", permettant d'exprimer par des jeux de parenthèses la structure, chaque fois différente, du document. On comprend l'intérêt de cette technique pour les différents domaines de la description archéologique, car le peintre de vases peut lui aussi disposer différemment ses décors, l'architecte du temple composer différemment ses pièces et ses colonnades. Le logiciel SIGMINI présente encore bien d'autres avantages, et, écrit d'abord pour de gros systèmes, il a reçu une version destinée aux mini-ordinateurs et aux micros fonctionnant sous UNIX/XENIX.

C'est avec ce logiciel qu'A.-M. Guimier-Sorbets a commencé à construire des banques de données, et d'abord une banque factuelle directe sur "La mosaïque dans le monde grec, dès origines à la fin de l'époque hellénistique"; opérationnelle depuis 1985, elle ne comporte qu'un nombre assez restreint de documents (moins de mille mosaïques, découvertes de la Sicile à l'Egypte et à l'Afghanistan), mais décrits d'une manière très détaillée; l'interrogation de la banque, comme l'édition des réponses ou la saisie de nouveaux documents, peuvent être réalisées indifféremment en français, anglais ou grec moderne. Puis a été lancée la construction d'une banque, cette fois référentielle puisque portant sur la Photothèque du Centre de recherche sur la mosaïque (et ici l'unité de description n'est plus la mosaïque, mais une photographie de mosaïque); cette photothèque est riche d'un grand nombre de documents pour lesquels on enregistre une information beaucoup plus légère, comme il est naturel, que pour la banque factuelle précédente. Et maintenant on construit une banque de données factuelle sur la mosaïque d'époque impériale en Grèce, rassemblant de très nombreux documents, parfois trop négligés, jusqu'à présent dispersés dans des publications d'accès souvent difficile. A côté de ces travaux qui lui appartiennent en propre, le laboratoire a très vite aussi été invité à apporter son aide à d'autres institutions qui, souhaitant construire des banques de données, ont voulu profiter de son expérience, étant bien entendu qu'ici nous préparons un système descriptif, essayons les logiciels, mettons au point les modalités de l'opération avec l'organisme consultant, mais que c'est ce dernier qui devra ensuite construire lui-même la banque, dont il gardera la totale responsabilité.

C'est ainsi qu'a été préparée l'informatisation du Centre de documentation photographique et photogrammétrique, CNRS et Université de Paris I, comme celle des photographies du Centre de recherches byzantines de l'Université de Strasbourg; hors de France, nous avons aidé l'Ecole française d'Athènes à informatiser son énorme fonds documentaire (plus de 300.000 photographies, des dizaines de milliers de plans, dessins, estampages, moulages, carnets de fouilles, documents divers), mais aussi à construire des banques factuelles, comme celle sur les monnaies de Thasos sous la responsabilité de O. Picard ou celle sur les vases conteneurs (amphores et lagynoi) sous la responsabilité de J. Y. Empereur, avec pour la dernière un système multilingue français-anglais-grec moderne. Toujours en Grèce, nous avons conçu pour le KERA (à la Fondation hellénique pour la recherche) une banque de données sur l'épigraphie de la Macédoine, qui en prend en compte les divers aspects et permet des recherches par exemple sur les noms propres en grec comme en latin; nous avons aussi collaboré avec le Comité pour la sauvegarde des monuments de l'Acropole d'Athènes, pour informatiser la consultation de l'énorme documentation qu'il amasse; dans tous ces cas, c'est le logiciel SIGMINI qui a été choisi pour gérer une information extrêmement complexe. Nous avons été conduits aussi à travailler pour le Service archéologique de Jordanie et pour le Service des Antiquités de l'Egypte, en vue d'opérations qui ajoutent, aux problèmes de la multiplicité des langues, ceux de la multiplicité des alphabets. Au total, c'est d'au moins une vingtaine de banques qu'A.-M. Guimier-Sorbets a assuré la conception: une réflexion d'ensemble sur ses travaux vient de paraître en 1990 dans un volume du CNRS intitulé *Les bases de données en archéologie, conception et mise en œuvre*.

Un certain nombre des banques réalisées ou simplement conçues par notre laboratoire ne pourraient guère s'imaginer sans documentation figurée: plusieurs banques référentielles sont destinées à faciliter la consultation d'une collection d'images, et la banque sur la mosaïque grecque, par exemple, a dès le début été illustrée par une importante série de photographies et de dessins, d'autant plus intéressants que parfois, même anciens et de mauvaise qualité, ils restent le seul témoignage de documents aujourd'hui disparus. C'est pourquoi notre laboratoire a été choisi en octobre 1984 pour réaliser un vidéodisque, le premier sur l'archéologie, et dont une caractéristique est que ses images ne sont pas accessibles seulement à partir d'un catalogue imprimé, mais aussi à partir de trois banques de données déjà constituées auparavant: ces images sont en effet celles des deux premières banques réalisées dans notre laboratoire et évoquées plus haut, et de la banque du Centre de Documentation photographique et photogrammétrique dont nous avons assuré la conception. Le vidéodisque a été diffusé en 1986, et sa sortie accompagnée par la publication d'un fascicule intitulé *Images de l'archéologie. Vidéodisque*. Désormais, pour chacune des trois

banques, l'interrogation fait apparaître, sur un écran, la description des documents qui y répondent et, sur un autre écran, l'image (ou les images, car il peut y en avoir plusieurs dizaines) correspondant à chacun de ces documents. On imagine l'intérêt d'un tel dispositif, qui permet de diffuser à la fois la banque de données et la banque d'images qui l'illustre; car la description verbale, aussi fine soit-elle, ne peut porter la totalité des informations que donne la moindre photographie; et nous avons mis au point, en collaboration avec le Centre d'automatique et d'informatique de l'Ecole des Mines de Paris, un système permettant non seulement d'appeler les images à partir de la description, mais aussi la description à partir des images, et même de mémoriser, dans des sortes de "blocs-notes", analyses et images d'une sélection de documents, auxquelles l'utilisateur peut ajouter ses indications personnelles.

Lorsque ce premier vidéodisque a été conçu, la technique analogique était la seule qui permettait l'enregistrement, et l'appel quasi instantané, d'un nombre de figures allant jusqu'à cinquante mille par face de disque. Elle est toujours la seule à accorder cette facilité, fondamentale pour l'exploitation et la diffusion de grosses collections d'images. Mais d'autres techniques se développent, avec en particulier la numérisation qui facilite, elle, de nouveaux types de traitement, en même temps qu'elle invite à donner à l'image une qualité comparable à celle de nos meilleures photographies, au prix, il est vrai, d'un volume considérable de mémoire. Notre laboratoire étudie les possibilités d'utiliser ces images numériques dans le domaine de l'archéologie. Il travaille aussi désormais à la conception de systèmes d'information multimedia, gérant à la fois des bases de données, des textes numérisés, des photographies et des dessins, ces différents documents pouvant être enregistrés sur les nouveaux supports d'information (vidéodisque, CD-ROM, DON): il s'agit de réaliser, à partir d'un même ensemble de documents, divers systèmes d'information destinés à des publics différents (chercheurs, étudiants, mais aussi public des musées et des sites), ce qui implique différents types d'exploitation et de consultation mettant en œuvre, outre des logiciels de bases de données, des logiciels "hypermedia" et des logiciels de traitement du langage naturel.

RENÉ GINOUVÈS ANNE-MARIE GUIMIER-SORBETS

Centre de recherche sur les traitements
automatisés en archéologie classique
Université de Paris X, Nanterre Cedex

ABSTRACT

The Authors present the research activity carried out at the "Centre de recherche sur les Traitements Automatisés en Archéologie Classique". This activity can be divided into two main

themes. The purpose of the first is to publish works intended to standardize descriptive archaeological language. The second aims at creating data banks, with particular reference to classical archaeology, and videodisks permitting the association of images to the relative documents.