

Les bases de données en égyptologie

Jean WINAND

La dernière Table Ronde "Informatique et Egyptologie" s'est tenue à Paris, dans les locaux de la Sorbonne, du 4 au 6 juillet 1990. Une vingtaine de spécialistes venus de dix pays se sont réunis pour faire part de leur expérience face aux problèmes soulevés par la constitution des bases de données dans leur discipline.

Au début des années quatre-vingts, l'informatique appliquée à l'égyptologie était encore l'apanage de quelques privilégiés qui avaient la chance de travailler sur des gros systèmes, chacun au sein d'une institution de recherche. Depuis cinq ans, l'essor prodigieux de la micro-informatique a soudainement élargi le champ des utilisateurs potentiels. Après une brève période, inévitable, de tâtonnement et de flottement, on peut maintenant affirmer que l'égyptologie est entrée de plain pied dans l'ère informatique.

Assimilé d'abord par beaucoup d'entre nous à une machine à écrire perfectionnée, l'ordinateur s'est vite imposé comme un puissant outil de recherche. La multiplication des bases de données documentaires en est une des preuves les plus éclatantes¹. En égyptologie, celles-ci peuvent se répartir en trois catégories selon leur objet; c'est ainsi qu'on distingue :

- a) les bases de données consacrées à la bibliographie ou au traitement de citations,
- b) celles qui sont réservées à l'analyse des textes ou de la langue,

¹ L'intérêt de procéder rapidement à l'établissement de banques de données recouvrant tous les aspects de l'égyptologie n'est pas neuf. L'idée était déjà dans l'air à la fin des années soixante; pour s'en convaincre, il suffit de se reporter aux actes du colloque organisé à Darmstadt en 1969 (A. SCHWAB-SCHOLTT, éd., *Dokumentation ägyptischer Altertümer, Akten der Arbeitsgemeinschaft Dokumentation in den Historischen Wissenschaften*, 1, Darmstadt, 1970). Dix ans plus tard, Cl. Crozier-Brelot plaide avec passion pour l'introduction des nouvelles techniques en égyptologie, écrivant notamment : "L'égyptologie ne peut plus prendre le risque de rester en dehors de la société informationnelle. Il ne suffit pas de refuser simplement les évolutions rapides liées à l'informatisation pour en maîtriser le cours." (*Egyptologie et Informatique*, dans *CdE*, 56 (1981), p. 53).

- c) celles, enfin, qui sont orientées vers l'étude des documents archéologiques, expression prise ici dans un sens très large.

1. Les bases bibliographiques ou de citations

Le recours à l'informatique pour établir la bibliographie ou dresser des listes de citations ou de références est déjà une habitude bien établie. Le travail est rendu d'autant plus aisé que la matière à traiter permet le plus souvent l'utilisation de logiciels commerciaux largement diffusés. Trois communications ont été présentées sur ce thème.

a) Depuis bientôt trois ans, J. Serdült, sous la direction de N. Grimal, a commencé l'enregistrement des ouvrages contenus au cabinet d'égyptologie du Collège de France et au Centre de Recherches Egyptologiques de la Sorbonne. Les promoteurs du projet ont décidé de faire d'abord porter leur effort sur le dépouillement des numéros de revues les plus récents. L'encodage ne se limite pas aux données bibliographiques traditionnelles; l'adjonction de mots-clés pour chaque article devrait grandement aider à la consultation du fichier. Cette base de données ne peut actuellement être consultée que par le biais de fiches imprimées; toutefois, l'installation d'un terminal dans la salle de lecture du Collège de France a d'ores et déjà été prévue afin de permettre à l'utilisateur de poser directement ses questions. Le travail est actuellement réalisé avec le logiciel ISIS implanté sur un ordinateur BULL fonctionnant sous UNIX.

b) La *Bibliographie égyptologique annuelle* fait depuis un an l'expérience de l'informatisation grâce à l'aide technique apportée par l'équipe de D. van der Plas de l'Université d'Utrecht. Toutes les entrées bibliographiques concernant l'année 1986 ont été encodées, y compris les résumés. C'est le logiciel ISIS qui a été retenu pour accomplir cette tâche. A côté de la diffusion imprimée traditionnelle, l'objectif est de distribuer désormais l'année bibliographique sur un support magnétique. Par ailleurs, les volumes déjà parus (depuis 1947) devraient rapidement s'intégrer à cette nouvelle base de données.

c) Depuis plus de 20 ans, Cl. Crozier-Brelot fait œuvre de pionnier en augmentant inlassablement sa base de données consacrée aux citations des *Textes des Pyramides*. Par delà l'évolution des machines et des programmes (l'auteur utilise pour l'instant un programme personnel écrit en Quick Basic et fonctionnant sur un LOGABAX Persona 1600), la structure du fichier n'a guère varié. Avec un accroissement annuel de quelque 10.000 citations, le but avoué est d'atteindre le chiffre des 200.000 citations à l'aube de l'an deux mille.

2. Les bases de données de textes

En ce qui concerne les textes, la constitution de bases de données importantes a été longtemps subordonnée à la résolution des problèmes soulevés par l'encodage des hiéroglyphes et par la possibilité de les reproduire à l'écran ou sur papier. En ce qui concerne la codification des signes à proprement parler, un large accord semble s'être formé, du moins chez les utilisateurs IBM, autour des propositions faites dans le *Manuel*².

a) D'une manière générale, le stockage des textes peut être envisagé de deux façons différentes, selon les buts poursuivis. C'est ce que J. Winand (Liège, Belgique) a tenté de montrer dans sa communication :

- 1) dans les bases de données à caractère général, le texte entier (ou, à tout le moins, une large portion) forme l'unité de l'enregistrement. Les manipulations que l'on peut faire sur ce type de base sont naturellement fort réduites, puisque les formes ne sont ni lemmatisées, ni analysées; on ne peut guère aller au delà de la recherche simple d'occurrences.
- 2) avec les bases de données de la seconde catégorie, l'utilisateur souhaite pouvoir procéder à des enquêtes philologiques ou linguistiques approfondies. Dans ce but, il faut recourir à un découpage totalement différent de la matière. L'indication, dans ce cas indispensable, du lemme, de l'analyse morpho-syntaxique et de la traduction, nous oblige à consacrer un enregistrement par mot. C'est le mot qui devient ici le noyau de la base de données et non plus le texte tout entier.

Les bases de données sont actuellement traitées avec les logiciels dBase III+ et ISIS tournant sur du matériel Hewlett-Packard, auxquels il faut ajouter une série de programmes spécifiques écrits par l'auteur.

b) Depuis plusieurs années, J. Hallof (Berlin, Von Humboldt Universität) s'occupe des problèmes méthodologiques posés par l'informatisation du *Wörterbuch der ägyptischen Sprache*. La seconde édition de cette œuvre monumentale ne devrait être que la petite partie imprimée d'une banque de données générale comprenant tous les textes qui auront été dépouillés. Les principes majeurs de l'enregistrement sont très proches de ceux présentés par J. Winand. La liste des codes morpho-syntaxiques est toutefois un peu plus réduite, ce qui se comprend parfaitement dans la mesure où elle a été dressée pour répondre aux besoins spécifiques du *Wörterbuch*.

² J. BUURMAN, N. GRIMAL, M. HAINSWORTH, J. HALLOF, D. VAN DER PLAS, *Inventaire des signes hiéroglyphiques en vue de leur saisie informatique, Manuel de codage des textes hiéroglyphiques en vue de leur saisie sur ordinateur*, Paris, 1988.

c) La première édition du *Wörterbuch* fait en ce moment l'objet d'une informatisation complète. Le premier objectif de ce projet, conçu par O. Goelet (New York, USA) et Chr. Sturtewagen (Inst. Pontifical, Italie), est de présenter à l'écran un système interactif permettant non seulement un accès rapide à toutes les parties constitutives du dictionnaire, mais aussi des recherches qu'il aurait été difficile, sinon impossible de mener à bien à partir de l'édition imprimée. L'ambition des promoteurs ne se limite pas à encoder les données contenues dans l'ancien *Wörterbuch*, puisqu'ils ont conçu leur base de données d'une manière ouverte autorisant chaque utilisateur à la compléter et à la modifier selon ses besoins spécifiques. Le logiciel conçu par les auteurs a été écrit en CLIPPER et est destiné aux ordinateurs de la famille IBM.

3) Les bases de données archéologiques

Une des tâches prioritaires de l'égyptologie moderne est de mettre à la disposition des chercheurs les vestiges archéologiques qui se trouvent encore *in situ* ainsi que les innombrables objets contenus dans les musées. Plusieurs communications ont abordé ce sujet important. On notera tout spécialement cette année l'intérêt croissant des concepteurs à intégrer l'image dans les bases de données, le plus souvent sur un support vidéo-disque.

a) Depuis 1975, le musée du Louvre, sous l'impulsion du Ministère de la Culture a entrepris l'informatisation des antiquités égyptiennes. Le travail est réalisé avec le logiciel MISTRAL sur un ordinateur BULL DPS-8 fonctionnant sous MULTICS. A ce jour, quelque 34.000 objets ont été répertoriés, soit plus des trois-quarts de la collection. Pour l'heure, la base n'est consultable que sur place. En raison du coût relativement élevé de la fabrication d'un vidéo-disque, le département a provisoirement abandonné l'idée d'associer une banque d'images à la base de données.

b) Toujours au Louvre et avec le même logiciel, une application particulière a été développée par J.-L. Bovot sur un lot de 1.300 ouchebtis non inventoriés. L'auteur ne prétend pas condenser dans une fiche la totalité des informations concernant un ouchebti, mais espère pouvoir rassembler ainsi rapidement une documentation exhaustive des collections. Dans cette perspective, la méthode d'enregistrement proposée ici pourrait servir de modèle pour établir un bordereau de saisie standard.

c) Les ouchebtis ont aussi été la préoccupation de J.-L. Chappaz (Genève, Suisse). Toutefois l'approche est fort différente : à l'inverse de J.-L. Bovot qui met l'accent sur la description des données archéologiques et iconographiques, J.-L. Chappaz s'intéresse principalement à la prosopographie et à réunir la bibliographie qui paraît chaque année sur les ouchebtis. L'objectif de cette base de données est double : regrouper tout d'abord les mentions relatives aux figurines

funéraires en un seul grand fichier et garantir l'accès à l'ensemble des informations; publier ensuite régulièrement les accroissements du fichier sous forme d'un index annuel. Le travail est effectué sur un IBM-compatible à l'aide des logiciels dBase III+ et ISIS.

d) L'informatique était naturellement désignée pour aider à dresser l'inventaire des magasins de fouilles établis sur les sites mêmes en Egypte. Depuis 1988, Amal Hilal (Service des Antiquités, Egypte) a entrepris l'informatisation des magasins de Saqqarah avec la collaboration technique de A.M. Guimier-Sorbets (CNRS, Paris). L'exploitation des données se fait actuellement à l'aide du logiciel SIGMINI, de l'Ecole des Mines de Paris, fonctionnant sous UNIX ou XENIX.

e) F. Saleh (Le Caire, Egypte) a présenté une structure de base de données destinée principalement à la saisie des objets et à la gestion d'un musée. Un projet pilote a été réalisé à partir des objets provenant de la tombe de Toutankhamon et conservés au musée du Caire. Le travail utilise les logiciels dBase IV et ISIS implantés sur un IBM. L'auteur espère que son système pourrait servir de modèle pour la saisie des données archéologiques. Dans cette perspective, un thésaurus important avec les symboles destinés à la codification est joint à la fin de l'article.

f) Dès la mise en chantier du *Corpus Antiquitatum Aegyptiacarum* en 1976, on avait envisagé l'informatisation des données. Toutefois, les moyens techniques dont on disposait à l'époque n'ont pas permis la réalisation de ce projet. C'est cette idée qu'a reprise D. Bergman (Uppsala, Suède) avec le projet EDOC (Egyptological Documentation on Computer) avec l'appui du Comité international des Egyptologues. Le but d'EDOC est de procurer à la communauté scientifique un système informatisé intégré contenant les textes, les dessins et les photos de la plupart des objets renfermés dans les collections de par le monde.

g) L'archéologie de terrain était également représentée. Avec le système HYPERCARD gérant une base de données relationnelle (Reflex +), N. Strudwick (Cambridge, Royaume Uni) a montré comment on pouvait concevoir la saisie de monuments (en l'occurrence des tombes thébaines) en respectant les différentes parties constitutives. Une fois encore, une banque d'images est associée à la base-mère.

h) Enfin, il faut signaler la venue d'une nouvelle base de données, destinée cette fois à dresser un inventaire de l'iconographie royale pharaonique et entreprise par N. Cherpion et C. Rommelaere sous la direction de Cl. Vandersleyen (Louvain, Belgique). A l'heure actuelle, ce fichier compte 4.000 entrées. Les résultats seront disponibles dans une version imprimée; aucune consultation directe n'est possible pour le moment. Les auteurs ont retenu le logiciel File Maker Plus fonctionnant sur un Macintosh SE/30.

Le foisonnement formidable auquel on assiste dans le domaine informatique, l'effervescence extraordinaire auquel ce bouillonnement donne lieu en égyptologie menacent parfois de déboucher sur un certain chaos. Quel que soit le domaine particulier de l'égyptologie que l'on considère, plusieurs systèmes, parfois fort différents, sont sur le point de voir le jour. Or, la constitution de fichiers informatisés ne trouve réellement son sens que s'il existe une possibilité raisonnable de les rendre accessibles à d'autres chercheurs que ceux qui les ont créés. La mise au point d'une banque de données est une entreprise longue et fastidieuse, coûteuse en moyens matériels mais surtout humains. On ne peut donc plus se donner le luxe de faire deux fois le même travail.

Toutes les possibilités d'échange, de collaboration, de coopération que je viens brièvement de passer en revue tournent autour d'un maître-mot : la **compatibilité**.

Encore faut-il s'entendre sur ce que ce terme implique. Il est susceptible, me semble-t-il, de s'appliquer à quatre domaines relativement différenciés.

a) En règle générale, quand on parle de compatibilité, on pense d'abord au matériel (compatibilité hardware). La plupart d'entre nous, si ce n'est la totalité, utilise désormais des micro-ordinateurs, qu'ils appartiennent, pour faire bref, au monde IBM ou Macintosh. Même si le transfert de données entre ces deux familles relève encore parfois d'une gymnastique un peu compliquée, les problèmes posés sont loin d'être insurmontables.

b) Des complications peuvent encore surgir suite au système d'exploitation utilisé. Sans vouloir ici entrer dans des détails trop techniques, il faut savoir que tous les programmes ne tournent pas indifféremment sous DOS, OS/2, XENIX ou UNIX, pour ne parler ici que des PC.

c) Le troisième écueil est inhérent au choix du logiciel. Là aussi, il peut y avoir de nombreuses surprises. Le transfert des fichiers d'un logiciel vers un autre n'est pas toujours réalisable. Ceci est déjà vrai pour les programmes commerciaux. Je ne dirai rien ici de certains programmes écrits directement par des utilisateurs chevronnés.

d) Enfin, l'utilisation d'un même matériel et d'un logiciel identique n'est pas encore la garantie absolue. Même dans ce cas, *a priori* favorable, l'accès à l'information peut être terriblement perturbé si la structure de la base varie d'un utilisateur à l'autre.

La plupart des projets qui ont été présentés n'envisageaient qu'un domaine bien défini de l'égyptologie (bibliographie, philologie ou archéologie) avec, à chaque fois, un système différent et les problèmes de compatibilité que cela soulève. C'est

pourquoi certains ont imaginé une structure unique permettant d'intégrer toutes les bases de données, quel que soit leur objet.

a) A Utrecht, l'équipe dirigée par D. van der Plas a mis au point un système de bases de données intégré dont l'armature est constitué par le programme CDS/ISIS distribué par l'UNESCO. Ce logiciel fonctionne actuellement sur IBM PC et compatibles, VAX et gros systèmes. Il devrait bientôt être disponible sur Macintosh. Signalons également qu'il en existe une version arabe, ce qui constitue évidemment un point très important pour notre discipline. Les programmes des utilisateurs sont écrits en PASCAL.

Le système proposé consiste en un réseau de bases de données qui rassemble sous une gestion centrale des bases de données séparées regroupées au sein de quatre types : les objets, les textes, l'iconographie et l'archéologie. La colonne vertébrale est constituée par la base de données objets qui sert de lien entre les autres bases de nature plus spécifique. Actuellement ont déjà été intégrées à ce système plusieurs bases de données de nature fort diverse : une base de citations (Egyptian Coffin Word Index), une base de texte (le pLeyde T 2), une base archéologique (des stèles du Moyen Empire) et une base bibliographique (un volume de l'année bibliographique).

b) A. Grimm et D. Wildung ont présenté un projet fort semblable en théorie, même s'il s'en écarte considérablement dans la réalisation technique. Ici aussi, le but est de parvenir à un système documentaire intégré faisant intervenir quatre niveaux de connaissance de valeur identique : les données muséographiques, la description des objets, la description iconographique et les textes.

Le système dispose d'un thésaurus de termes réservés (lexique) auquel sont joints un fichier d'aide et un fichier de gestion. Ces bases sont gérées par dBase III+ ou IV. Le système général des banques de données est régi par le logiciel LARS 3.4 dans lequel l'utilisateur peut définir ses propres programmes en langage C. A cela s'ajoute une banque de données des sources (pour les textes) et une banque d'images dans laquelle sont stockés les dessins et représentations soit par l'intermédiaire d'une table digitale soit par celui d'un scanner. Le projet, qui bénéficie du soutien actif d'IBM, fait appel à une infrastructure assez lourde : un IBM PS/2-80 avec un disque de 200 Mb, un Laufwerk optique IBM 3363, un IBM PS/2 portable P70 386 pouvant être connecté au premier via le réseau IBM LAN.

c) Dû à E. Loring, le système ONTOΣ se présente comme un système de banques de données relationnelles. Son architecture est telle que l'utilisateur peut ajouter des opérations spécifiques à n'importe quel moment, à n'importe quel niveau et à n'importe quel secteur. Actuellement le programme répond surtout aux besoins des musées et des études iconographiques ou archéologiques; rien n'a encore été fait pour l'encodage des textes ou la gestion des images. ONTOΣ fonctionne sur

IBM 386 ou 286 avec des fichiers dBase III+; pour des raisons d'efficacité, les programmes qui assurent les activités de recherche sont compilés en CLIPPER.

Tous ces projets sont très généreux, mais il est peut probable qu'un de ces systèmes puisse un jour s'imposer. Est-ce d'ailleurs souhaitable? Doit-on vraiment imaginer tous les égyptologues travaillant sur un matériel unique avec un logiciel commun sur des fichiers uniformément structuré? Cette hypothèse apparaîtra peut-être séduisante aux yeux de certains. Pour ma part, je voudrais dire tout ce qu'elle a d'illusoire, faisant souvent fi de la situation particulière de chaque pays (institutionnelle notamment), des goûts et des préférences légitimes de chacun. Enfin, le problème est-il bien là? Comme on l'a déjà dit, le transfert des données d'une machine à l'autre est toujours réalisable, à tout le moins sous une forme ASCII, quel que soit le matériel ou le logiciel utilisé. En revanche, la manière dont les données ont été encodées ne peut pas être facilement remaniée. Le vrai combat pour la compatibilité se trouve ici. C'est le seul point sur lequel il est crucial que les égyptologues puissent arriver à un accord; c'est ce que j'appellerai les *normes de codages*. Certaines ont déjà été définies.

a) Les textes ont été les premiers à bénéficier d'une standardisation à la saisie : la translittération et le codage des hiéroglyphes se font désormais suivant une norme définie dans ce que nous désignons déjà d'une manière familière sous le nom de *Manuel* (cité n. 2). En ce qui concerne la codification de l'analyse morphosyntaxique, une solution devrait rapidement être trouvée à partir des travaux présentés par J. Hallof³ et J. Winand⁴.

b) Pour la référencement des sites, des monuments et des scènes, l'utilisation du système mis au point dans l'ouvrage de B. Porter et R. Moss est largement répandu.

c) Il est aujourd'hui de la plus haute importance d'arriver à un consensus analogue pour la description des objets. La situation est peut-être moins aisée dans la mesure où le nombre des mots réservés qui devraient figurer dans un thésaurus est beaucoup plus important que pour les textes; enfin, il faut également tenir compte du fait que plusieurs institutions ont déjà derrière elles une longue tradition sur laquelle il est souvent difficile de revenir. Le département des antiquités égyptiennes du musée

³ J. HALLOF, *Wörterbuch und Datenbanken*, dans *Informatique et Egyptologie*, 7 (1990), p. 90-92; ID., *Codierungslisten und Analyse ägyptischer Hieroglyphen*, dans *Meroitica*, 12 (sous presse).

⁴ Les premiers essais ont été consacrés au néo-égyptien : J. WINAND, *Late Egyptian Analysis by Computer*, dans *Crossroad, Chaos or the Beginning of a New Paradigm*, Copenhague, 1986, p. 389-400; ID., *Constitution de fichiers-textes en néo-égyptien : lemmatisation et analyse automatiques*, dans *Revue*, 22 (1986), p. 179-190; ces positions ont été résumées dans *Le Voyage d'Oumamon*, Liège, 1987, p. XXVIII-XXXII. Dans *Les bases de données de texte en égyptien*, (*Informatique et Egyptologie*, 7 (1990)), p. 164-169), des normes de codage valables pour toute l'histoire de la langue pharaonique ont été proposées.

du Louvre travaille depuis plusieurs années à l'aide d'un lexique qui comprend plusieurs milliers d'entrées hiérarchisées. De même, au musée de Munich, l'équipe dirigée par D. Wildung a mis au point un thésaurus qui ne compte pas moins de 9.000 termes. Un travail similaire a été entrepris il y a quelque temps par le British Museum. Enfin, F. Saleh a proposé une nouvelle liste de termes destinés à décrire les objets conservés au musée du Caire. On peut espérer que de toutes ces listes sortira un jour un manuel de codage multilingue qui assurerait aux bases de données archéologiques la même compatibilité que celle des bases textuelles.

Enfin, je voudrais terminer cette revue consacrée aux bases de données en signalant une approche fort différente. Tous les projets qui ont été présentés ont été conçus pour fonctionner avec des logiciels dont les différences ne sont, en fait, que superficielles. Bien sûr, certains sont plus efficaces, plus souples ou mieux développés que d'autres; les performances peuvent également varier d'une manière très sensible d'un programme à l'autre. Il n'en reste pas moins que la logique qui est à la base de ces logiciels est partout fondamentalement la même.

Dans le cadre de son travail sur les talatates du IX^e pylône à Karnak, R. Vergnien s'est intéressé aux possibilités offertes par les systèmes experts. Dans ce type de structure, la division très nette entre la base de faits et le groupe de règles agissant sur les faits l'a obligé à revoir la façon dont les données sont introduites afin d'avoir une description univoque et minimale pour chaque élément. A moyen terme, c'est là que se joue le problème de la compatibilité : pas tant sur le choix des matériels ou des logiciels, mais bien sur la manière de décrire les données et de les encoder. Le degré de compatibilité d'une base de données se mesurerait ainsi dans la faculté qu'elle offrirait à n'importe quel utilisateur de l'utiliser d'une autre manière que celle pour laquelle elle a été initialement prévue.