

A propos de deux expériences de banques de données sur ordinateur en archéologie

par

Alain GALLAY

Les deux expériences dont il est question ici sont de nature très différente. La première (de Lumley et al. 1974)¹ n'est qu'un projet très ambitieux portant sur l'information récoltée lors de fouilles de sites préhistoriques, la seconde (F. Digard et al. 1975)² réunit la présentation et le mode d'emploi d'une banque de données existante portant sur un thème spécifique, les cylindres-sceaux orientaux.

Avant de les aborder nous voudrions présenter certaines options méthodologiques (proposées entre autres dans le second travail) que nous considérons comme acquises et qui ne feront pas, par conséquent, l'objet d'une discussion ou d'une remise en question. Cette position est bien entendu personnelle; nous savons en effet que l'accord au sujet de ces quelques principes est encore bien loin de se réaliser dans la communauté des archéologues. Certaines de ces options serviront de repères dans l'évaluation des deux entreprises que nous considérons.

1. Il est indispensable de distinguer clairement les démarches de l'érudition qui sont à la base des compilations les plus diverses en archéologie (corpus, catalogues, maintenant banques de données sur ordinateur) des *productions scientifiques* proprement dites (théories explicatives, modèles, etc.) dans un « *souci d'un partage plus clair et surtout plus avoué entre les constructions porteuses de « connaissances » scientifiques nouvelles et celles qui se réduisent à la présentation de « données documentaires inédites »* (Digard, vol. 1, p. 240). J.-C. Gardin montre clairement (Digard, vol. 1, chap. 6, effets logiques des bases de données sur les disciplines d'érudition) les conséquences de cette position sur le plan institutionnel (gestion collective des instruments de travail de l'érudition) et pédagogique (dévalorisation prochaine du savoir « factuel », compilatoire, et valorisation du savoir « opératoire »; nécessité d'une preuve expérimentale de l'utilité du discours pédagogique, etc.).

2. Il est nécessaire de subordonner la pratique de la description compilatoire à des *objectifs explicites*. Le temps où la description des objets archéologiques pouvait être un but en soi est révolu. Cette option fondamentale montre qu'il n'existe pas une description possible, « exhaustive », d'un objet archéologique mais autant de descriptions que de projets scientifiques.

3. Les projets compilatoires et scientifiques ne peuvent se passer d'un support linguistique ou métalinguistique spécifique, langage documentaire (LD) d'une part, langage scien-

¹ LUMLEY-WOODYEAR, Henry DE. 1974. *Exploitation des données d'un site préhistorique* (en collaboration avec Louis BOURRELY et Henriette CAMPS-FABRER). In: *Banques de données archéologiques*. Colloques nationaux du Centre national de la Recherche scientifique, n° 932, Marseille 12-14 juin 1972. Paris, 39-80. Cette présentation comporte 3 articles: LUMLEY-WOODYEAR, H. DE. I. *Présentation générale du système*. BOURRELY, L. II. *L'analyse et la programmation*. CAMPS-FABRER, H. III. *Projet de code pour l'analyse des objets en os*. Nous avons choisi de parler essentiellement de ces trois études qui paraissent significatives de la tendance générale des travaux présentés lors de ce colloque.

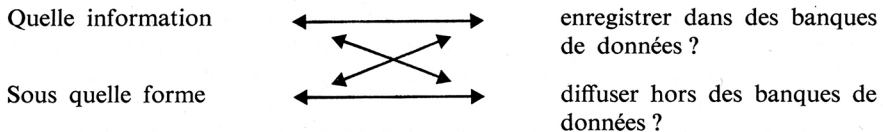
² DIGARD, F. *Répertoire analytique des cylindres orientaux publiés dans des sources bibliographiques éparses (sur ordinateur)* (en collaboration avec C. ABELLARD, L. BOURRELY, J. DESHAYES, J.-C. GARDIN, J. LE MAÎTRE et M.-R. SALOMÉ). Centre national de la Recherche scientifique, Paris 1975, 3 vol. (1. *Principes et résultats*, 320 p., 4 fig.; 2. *Code*, 349 p., 31 pl., 4 cartes; 3. *Commentaire*, 217 p., 29 fig.), + un catalogue sur fiches 75 × 125 mm (4719 fiches) + un fichier électronique sur disque Pyral PL 20. Les références données dans le texte seront signalées sous Digard, bien que plusieurs chapitres aient été écrits par les autres collaborateurs, J.-C. Gardin en particulier.

tifique (LS) d'autre part, le langage naturel étant impropre. S'agissant des langages documentaires, ce passage du langage naturel à un outil adéquat recouvre « dans un seul et même mouvement trois ordres de réductions différents : (a) la réduction de la diversité des langues naturelles à une seule, « artificielle », (b) la réduction de la diversité des dialectes d'école, voire des idiolectes de chacun, à un langage « normalisé », (c) enfin la réduction du discours linéaire pratiqué par les unes et les autres aux expressions discontinues d'un langage « analytique » (Digard, vol. 1, p. 33). D'autre part il y a lieu de distinguer (malgré leurs relations évidentes) les langages documentaires (LD) utilisés dans les compilations pour la recherche des documents signifiants des langages scientifiques (LS) proposés par la suite dans les constructions explicatives porteuses de connaissance scientifique.

4. Conséquence des points 2 et 3, on postule la rétroaction nécessaire du domaine de la connaissance scientifique (LS) en direction des projets compilatoires (LD). La description des objets doit tenir compte à la fois des objectifs de la recherche (point 2) et de l'état présent des connaissances scientifiques.

5. Enfin l'ordinateur, qui se généralise dans tous les domaines de la recherche, est susceptible d'apporter une aide incomparable dès que la quantité d'information atteint un certain seuil, ou que les procédures de « calcul » deviennent trop longues pour être effectuées à la main. Les exigences du travail en machine contribuent également grandement à l'explicitation des organisations logico-sémantiques mises à contribution.

Ces différents points acquis, posons que l'examen critique des deux expériences précitées doit nous permettre de répondre à la question: dans le cadre des compilations, quelle information enregistrer, sous quelle forme, quelle information diffuser, et sous quelle forme ?



Une remarque tout d'abord. Le projet appliqué aux données recueillies lors de la fouille paraît à première vue plus proche de la phase compilatoire de la recherche; l'expérience appliquée aux cylindres orientaux semble par contre s'insérer dans une phase plus avancée de la recherche scientifique. Si l'on y regarde de près pourtant les objectifs des deux entreprises sont moins éloignés l'un de l'autre qu'il ne paraît. La même problématique d'analyse peut donc s'appliquer à l'un et à l'autre.

1. L'ENREGISTREMENT DES DONNÉES DE FOUILLES

1.1. Définition des objectifs

Signalons tout d'abord que la position de l'archéologie paraît singulière quand on la compare à la situation existant dans les sciences expérimentales (fig. 1).

En sciences expérimentales le *même matériel* d'étude (virus, rats, grenouilles) est constamment disponible; il est donc toujours possible de s'y référer et de monter une nouvelle expérience sur ce même matériel pour contester ou vérifier les résultats publiés en retournant aux conditions de l'expérience. Cette situation est identique pour certaines sciences d'observation (astronomie, géologie, etc.).

En archéologie la fouille détruit un matériel d'étude auquel on ne pourra plus revenir par la suite; l'enregistrement des données joue donc un rôle essentiel. La publication (au sens très large) joue donc un double rôle puisqu'elle doit, à la fois, rendre compte de la

nature du matériel d'étude dans des protocoles d'observation (auxquels il doit être possible de se référer en cas de contestation) et des résultats des expériences (construction de typologies, C^t , ou d'explications, C^e) faites sur ce matériel. Cette situation justifie donc la présence de banques de données quelle qu'en soit la forme.

Nous rejetons par contre l'argument de l'« unicité » de chaque site préhistorique, souvent avancé pour justifier l'accumulation incohérente de n'importe quelles données d'observation au sein des monographies de sites. Ces dernières risquent de ne jamais voir le jour du fait de l'énormité de la masse d'information que l'on croit nécessaire de rassembler (il se trouve que nous nous sentons le premier visé par ce genre d'argumentation!). Nous rejetons donc cet argument au nom du principe 2 évoqué ci-dessus et aussi parce qu'il n'est

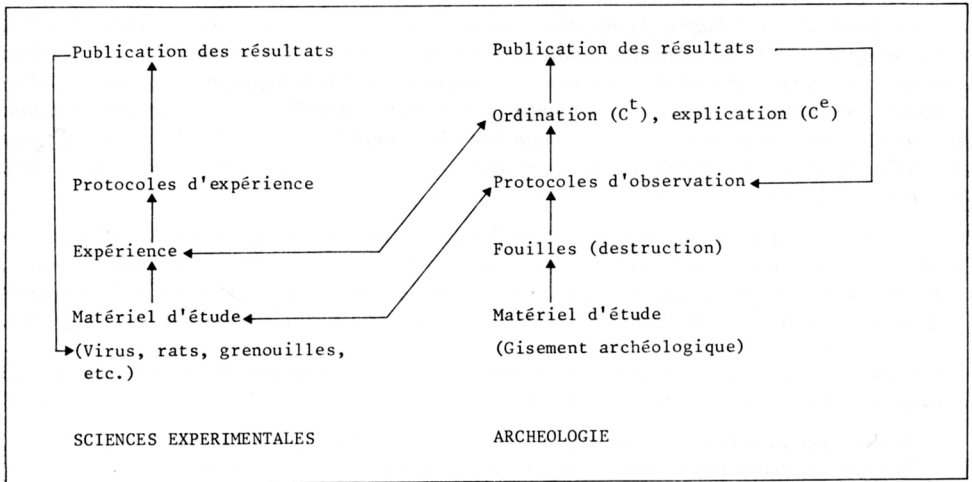


FIG. 1. — Déroulement de la recherche en sciences expérimentales et en archéologie. Les boucles de rétroaction s'appliquent aux procédures de vérification *interne* de l'expérience.

de sciences que du général. L'unicité ou la diversité d'un phénomène dépend de la définition qu'on en donne. A l'archéologue de définir le niveau conceptuel des phénomènes qu'il étudie. Une réduction de ce genre permet immédiatement de rassembler plusieurs sites comparables auxquels peuvent s'appliquer des procédures de validation.

Ceci dit, les objectifs que fixe la recherche actuelle, notamment en archéologie préhistorique (et probablement bientôt en archéologie tout court), pour l'étude et la publication d'un site, restent singulièrement lourds. Ils sont connus de tous; nous ne ferons que résumer les principaux:

1. Aspects historiques et culturels. Définition des composantes culturelles des diverses civilisations représentées et mise en place dans un cadre géohistorique élargi.
2. Aspects ethnologiques. Définition du cadre technologique, économique, social et parfois même idéologique.
3. Aspects écologiques. Reconstitution du cadre naturel (climat, faune, flore) et insertion de l'homme dans ce cadre (territoire, influence de l'homme sur le milieu, etc.).
4. Aspects biologiques. Reconstitution de l'aspect physique des hommes d'autrefois et étude des relations avec les aspects développés en 1, 2 et 3.

Etant donné l'action destructrice de la fouille et la raréfaction de fait de certains types de sites dans les régions fortement urbanisées on admet communément la nécessité de retenir l'ensemble de ces objectifs à propos de chaque site. Nous ne sommes pas certain que cette position (qui du fait des fouilles de sauvetage est très rarement tenue) soit totalement défendable. Nous laisserons pourtant ici la question en suspens et nous ferons comme si cela était vrai. Malgré cela nous verrons que la position de de Lumley n'en est guère plus défendable.

1.2. *Le projet d'exploitation des données d'un site préhistorique de H. de Lumley.*

Précisons tout d'abord que l'entreprise est un projet et référons-nous à de Lumley même pour en exposer les grandes lignes :

« L'étude des outillages, la meilleure connaissance possible des sols d'habitat fossiles, la reconstitution de l'environnement quaternaire, bénéficieront de la confrontation des résultats d'analyse de chaque spécialiste : géologue, archéologue, anthropologue, paléontologue, paléobotaniste. A l'issue des recherches le nombre des données recueillies est considérable d'autant plus que les précisions apportées par chaque chercheur sont de plus en plus nombreuses. Traiter une telle masse de documents est un long travail, fastidieux, voire même impossible, avec des moyens classiques » (p. 41).

« Souhaitant mettre au point une « méthode universelle » (nous soulignons) d'investigation et de traitement automatique des informations recueillies sur un site préhistorique, notre programme lui-même est plus ambitieux » (que les entreprises répondant à des objectifs limités mises en œuvre à Bylany en Tchécoslovaquie et à Teotihuacan au Mexique) (p. 42).

Le projet tend ainsi à regrouper et à développer les nombreux codes mis en chantier ces dernières années et à les articuler dans les secteurs suivants (p. 42) :

- Lexique des caractéristiques géographiques d'un site archéologique.
- Lexique des références et des caractéristiques topographiques d'un objet.
- Lexique des caractéristiques de la faune.
- Lexique des caractéristiques de l'industrie lithique.
- Lexique des caractéristiques de l'outillage en os.
- Lexique des caractéristiques de l'outillage en métal.
- Lexique des caractéristiques de la céramique.
- Lexique des caractéristiques des pierres et des blocs.
- Lexique des caractéristiques des macrorestes paléobotaniques, des charbons de bois, des cendres et des traces fugaces.
- Lexique des structures.
- Lexique des caractéristiques des pollens.
- Lexique des données stratigraphiques et des caractéristiques sédimentologiques.

En apparence le projet répond aux objectifs que nous avons posés précédemment ce qui explique probablement l'écho rencontré dans la sphère des préhistoriens. Il nous paraît pourtant outrancier et utopique pour deux raisons : il tend à enregistrer *tous les caractères d'un objet*, il est en principe destiné à *tous les objets d'une fouille*¹. Expliquons-nous sur ces deux points.

Tous les caractères d'un objet. « Notre système est ouvert car nous voulons que chaque chercheur puisse ajouter au lexique les données dont il a besoin. Toutes les listes typologiques seront donc incluses dans la deuxième version du code. Ceux qui voudront se référer seulement

¹ Ajoutons que, apparemment, rien n'est dit sur l'aire géographique concernée ni sur la période de temps couverte par le projet. Faut-il comprendre que le champ géographique est également universel et que le projet englobe toute la préhistoire et la protohistoire (puisqu'il est question de céramique et d'objets métalliques) ?

aux listes typologiques le pourront, mais la description des objets sera donnée elle aussi. Toutes les caractéristiques de l'objet seront prises en compte » (de Lumley, p. 80). Un premier point peut être écarté, qui doit être évident aux auteurs du projet : l'exhaustivité d'une description ne veut rien dire, il n'est pas possible de donner une description complète d'un objet car, dans ce cas, la description se confondrait avec l'objet lui-même. Il s'agit donc d'une tendance asymptotique vers l'enregistrement d'un *maximum de données*. Cette tendance n'apparaît peut-être pas clairement dans les textes publiés (voir pourtant le projet de code des instruments en os retenant 122 points de description. Camps-Fabrer, p. 79); elle devient évidente quand on écoute les auteurs exposer le projet.

L'accumulation des données permet ainsi d'éviter de se poser la question de l'adéquation de la description aux objectifs posés (points 2 et 4 de nos remarques préliminaires). On remet ainsi à plus tard le moment où il faut choisir entre les critères jugés utiles et inutiles. Cette manière de voir entraîne obligatoirement une accumulation de données non significatives, ce qui ne saurait en soi être trop grave si cela n'avait pas de conséquences immédiates sur le plan pratique (rentabilité du système), institutionnel (mise en place d'une infrastructure capable d'assumer les différentes tâches nécessaires à l'enregistrement et à la gestion du système) et humain (hypertrophie de certaines tâches ingrates dénuées de finalité).

Tous les objets d'une fouille. Le second point est, pour de nombreux préhistoriens, moins évident. Nous pensons pourtant que le problème est essentiel. H. de Lumley écrit : « *Nous référant à notre expérience personnelle, l'étude et la mise au point du matériel recueilli sur le dernier sol d'habitat du Lazaret, qui comprenait plus de 3000 objets, ont nécessité près d'un an de travail et celles du matériel recueilli dans la grotte de l'Hortus au cours de cinq mois de fouilles, qui comprenait près de 15000 objets, ont nécessité trois ans de recherches.*

Était-il possible dans ces conditions d'analyser le matériel recueilli à Terra Amata (plus de 80.000 objets), à la Baume Bonne (plus de 100.000 objets) et à la Caune de l'Arago (plus de 300.000 objets dans l'état actuel des fouilles) » (p. 41).

Le second problème éludé est donc celui de l'échantillonnage et concerne également le point 2 de nos remarques préliminaires (quels objets choisir pour répondre aux questions que l'on pose ?). Cette position est en fait quasi générale chez les préhistoriens. L'archéologie préhistorique procède par accumulation d'indices, il est donc communément admis de tout enregistrer. On peut pourtant se demander si cette attitude n'entraîne pas un certain gaspillage (de sites, d'énergie, etc.). Nous reviendrons sur cette question dans notre conclusion.

En résumé, nous nous demandons si, par ses outrances mêmes, le système proposé ici ne se condamne pas lui-même à rester au stade de projet. Nous voyons mal en effet dans quel cadre institutionnel actuel une telle machine pourrait prendre place. Il n'en reste pas moins que l'énorme effort entrepris sur le plan des langages d'analyses et des codes risque de porter ses fruits, mais probablement sur un autre registre que prévu. Il adviendra des codes proposés ce qui est advenu de la plupart des codes mis en chantier auparavant (mais ces derniers étaient conçus pour cela); ils serviront à générer d'autres codes plus spécifiques utilisés dans des entreprises ponctuelles aux objectifs bien définis et limités. Cet aspect des choses nous semble le point le plus positif de l'entreprise, mais il s'agit d'une retombée secondaire.

2. LES CORPUS THÉMATIQUES.

2.1. Définition des objectifs.

Dans ce second cas les objectifs sont plus limités. Le corpus porte sur une seule catégorie d'objets, mais ces derniers proviennent de sites et de fouilles diverses dispersés dans une aire géographique déterminée. Il n'en reste pas moins qu'une même catégorie d'objets

peut donner lieu à des discours très divers. L'adéquation de la description aux objectifs implique donc qu'on choisisse certains champs de recherche pour en éliminer d'autres.

2.2. *Le répertoire analytique des cylindres orientaux de F. Digard.*

Ce corpus correspond à une entreprise réalisée avec « *le désir de dépasser le plan des innovations de méthode, abordé il y a une vingtaine d'années, pour se placer maintenant sur celui des mœurs, qui en est d'ailleurs le corollaire, et où l'archéologie n'est d'ailleurs pas seule en cause* » (Digard, p. 7). Il est donné comme une entreprise finie (ce qui est contraire à la notion même de banque de données), soit comme une expérience destinée à tester la portée pratique d'une réalisation de ce genre et à induire de nouvelles habitudes de recherche dans la communauté des archéologues.

Les différences existant par rapport au projet précédent sont résumées dans le tableau 1.

L'objectif est double, méthodologique d'abord : étendre à l'iconographie les expériences précédemment développées à propos d'objets archéologiques d'autres types (outillage de

TABLEAU 1. — *Caractéristiques différenciant les deux expériences de banques de données envisagées.*

| | H. de Lumley | F. Digard |
|-------------|---|---|
| Objectif | multiple (historique, ethnologique, écologique, etc.) | unique (iconographique) |
| Corpus | sites | objets (cylindres) |
| Visée | documentaire | scientifique |
| Conception | outil de travail effectif (corpus illimité) | expérience limitée (projet pilote, corpus fini) |
| Réalisation | projet | réalisation effective |

métal, monnaies, etc.), mais choisir un exemple relativement simple où l'iconographie « *plus complexe que celle de la numismatique, et cependant moins difficile à décrire, semblait-il, que ne pourraient l'être des représentations figurées plus riches ou plus raffinées* » (p. 10). Objectif archéologique ensuite : rassembler des documents totalement épars dans la « diaspora » des périodiques, rapports de fouilles, catalogues de vente, etc.

La banque de données présente de ce fait la structure suivante.

1. Un catalogue sur fiches diffusé avec les volumes consacrés au mode d'emploi. Ces fiches (75 × 125 mm) portent, outre le numéro d'inventaire, la photographie du thème iconographique, des indications sur l'origine de l'objet, la traduction des inscriptions éventuelles, les dimensions et les sources bibliographiques. Il s'agit donc de données signalétiques.
2. Trois volumes consacrés au système, au code et à son commentaire, soit :
 - Volume 1. Exposé des principes et exemple de résultats permettant de se faire une idée des performances du système.
 - Volume 2. Exposé du code analytique.
 - Volume 3. Commentaire permettant de rendre l'apprentissage du code plus facile.

3. La banque de données proprement dite sur support électronique (fichier inversé sur disque), réduite à un exemplaire conservé au Centre de Calcul universitaire du Pharo à Marseille (demandes d'exploitation à adresser à Monsieur J.-Cl. Gardin, CRA, 23, rue du Maroc, F-75019 Paris). Cette banque de données contient donc le support analytique du système et permet le tri des fiches du catalogue en fonction des questions posées.

Il s'agit donc d'un système volontairement hybride où les informations signalétiques ne sont pas incorporées dans la mémoire centrale, choix judicieux, pensons-nous, dans cette période de transition où les anciennes habitudes portant sur le contact « direct » avec les objets prédominent encore largement.

Enfin, au-delà du caractère spécifique de l'expérience, on trouvera dans le volume 1 des commentaires un exposé particulièrement circonstancié et critique sur les problèmes posés par l'introduction des banques de données sur ordinateur en archéologie, qui dépasse de très loin, le projet ponctuel présenté. N'étant pas spécialiste de la glyptique orientale c'est essentiellement à cette partie que nous consacrerons nos commentaires.

Mais avant d'aborder cette question, débarrassons-nous de quelques remarques plus spécifiques. L'apprentissage d'un langage analytique est une opération longue et complexe, fait justement souligné par les auteurs du projet-pilote; c'est également une opération nécessaire que nous ne remettrons pas en cause (cf. point 3 de l'introduction). Un premier examen du code (probablement trop rapide) montre pourtant certaines lacunes dans les dispositifs destinés à faciliter cet apprentissage. Si le lexique est parfaitement explicité à l'aide de nombreuses références iconographiques, nous avouons avoir eu du mal à trouver les éléments nécessaires à la description syntaxique et nous aurions aimé trouver des exemples concrets de codage de scènes (pourtant annoncés, vol. 1, p. 21, § 3). Il ne s'agit en effet pas d'une simple curiosité mais bien d'un élément fondamental du système car l'apprentissage du codage (donc de l'analyse) est indispensable à l'apprentissage de l'interrogation. A ce titre le langage de l'interrogation (vol. 1, chapitre 3) paraît mieux explicité que le langage d'analyse, d'où une question: pourquoi n'avoir pas joint au catalogue sur fiches la description codée des scènes ?

Venons-en maintenant aux problèmes de fond. Jean-Claude Gardin développe tout au long du chapitre 1 (L. Digard, vol. 1, p. 10-16) une argumentation très serrée en faveur de l'utilisation des banques de données et résume sa position dans un schéma, que nous pensons utile de reproduire ici (fig. 2), souhaitant « *vérifier s'il est maintenant possible (...) d'engager une discussion technique sur un canevas de ce genre (résumé sous forme d'organigramme de décision)* ». (p. 15).

Nous voudrions aborder ici les problèmes soulevés par les points 1 (corpus utile) et 2 (forme imprimée inadéquate) car la suite des options stratégiques (points 3, 4 et 5) et tactiques ne soulève à notre avis aucune objection fondamentale (cf. nos principes de départ.).

Postulat 1. Les corpus sont utiles. Posée sous cette forme, la question de l'utilité des corpus ne peut qu'entraîner une réponse positive. Evidemment les corpus sont utiles et même nécessaires puisqu'ils forment la matière même de toute recherche en archéologie. Un problème n'est pourtant pas posé: quel corpus, ou plutôt, un corpus de quelle étendue ? Toute l'argumentation part en effet du principe qu'il est non seulement utile mais nécessaire d'utiliser tous les documents disponibles. Autrement dit le problème de l'échantillonnage n'est pas posé et nous retrouvons l'un des deux travers que nous avons signalés à propos de l'entreprise précédente. Il est évident que cette question est essentielle puisque le recours aux ordinateurs est basé, entre autres, sur la masse considérable de l'information à traiter.

Postulat 2. La forme imprimée est inadéquate. Le point que nous voulons soulever est celui de la mise à jour. Il est clair que la forme imprimée se prête mal à la mise à jour et

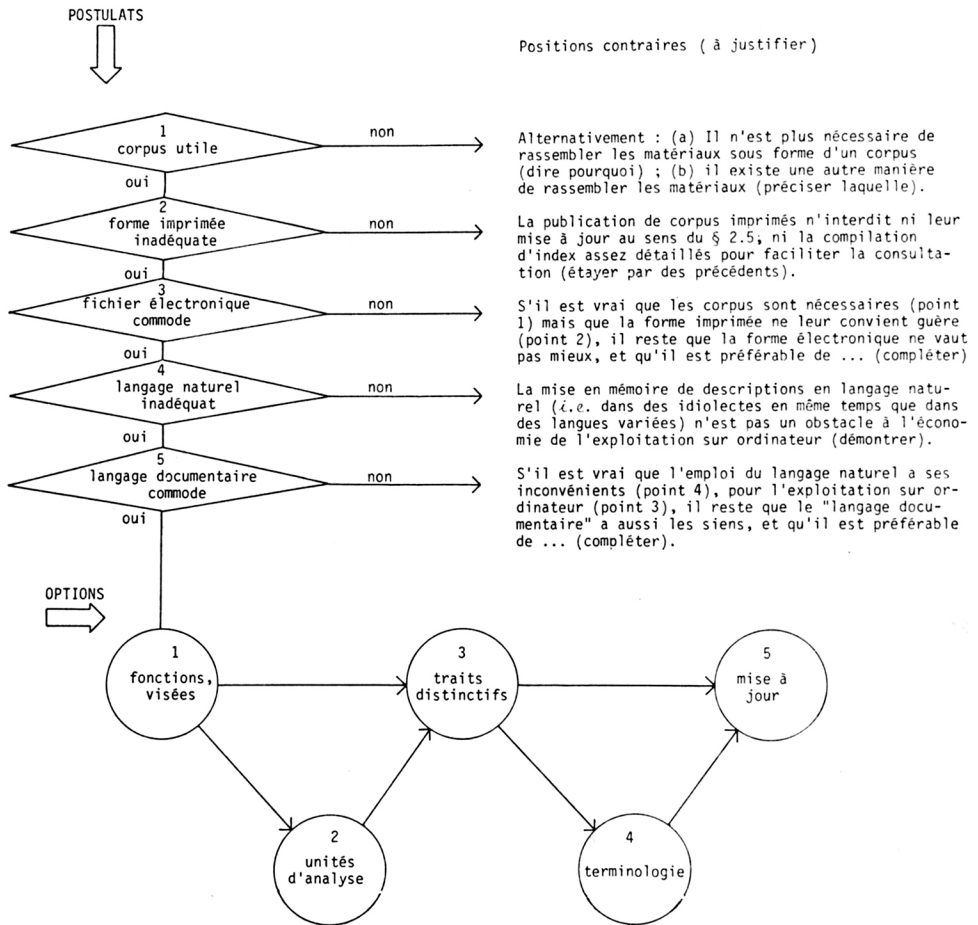


FIG. 2. — Résumé de l'argumentation développée en faveur de la création de banques de données sur ordinateur. Partie supérieure, options stratégiques; partie inférieure, conséquences tactiques de l'acceptation des propositions précédentes (repris de F. Digard, *Répertoire analytique des cylindres orientaux*, vol. 1, p. 16).

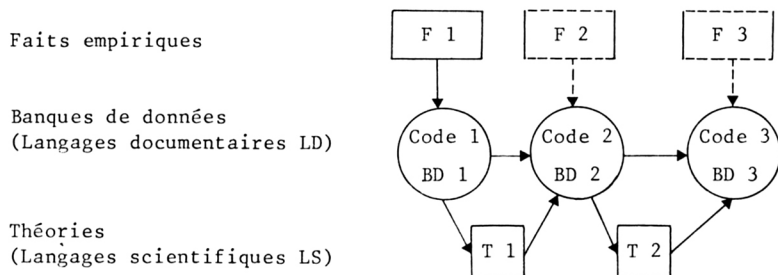
que le recours aux fichiers électroniques facilite cette tâche. L'avantage est évident au niveau du contenu des fichiers; l'est-il autant au niveau de la forme du langage analytique? Les auteurs semblent le penser:

« Cette manière de postuler et de limiter à l'avance le champ de recherches admises peut choquer (...); et l'on doit prévoir que toute option de cet ordre donnera lieu à des critiques, qu'il est facile de résumer en une seule: « votre répertoire ne permet pas de poser les questions qui m'intéressent ». La fatalité même de ce verdict peut éveiller quelque soupçon; la sagesse est toutefois de n'en rien montrer et de rétorquer seulement: « ajoutons donc au répertoire les données de votre goût ». Rien n'est plus facile en effet, dans le cas d'un fichier électronique. » (p. 14); ou bien « les fichiers électroniques sont essentiellement malléables; pour chacun des vices dénoncés, le remède est en principe des plus simples: ajouter ou retrancher des traits distinctifs, en dissocier ou en amalgamer d'autres, etc. » (p. 14); ou bien « Les

mêmes considérations de pertinence ou de commodité qui conduiront demain à supprimer, ajouter, dissocier ou amalgamer certains traits distinctifs, pour passer d'un Code 1 à un Code 2, conduiront après-demain à des opérations analogues pour passer du Code 2, au Code 3, etc.» (p. 33).

Toute l'argumentation repose en effet sur une situation où l'on fait *comme si* (nous soulignons, car ce problème est bien évidemment présent à l'esprit des promoteurs de cette expérience) une banque des données pouvait d'elle-même engendrer sa propre mise à jour sous la pression exercée par les théories (cf. schémas fig. 3). Il est évident que le cheminement que nous présentons à la partie supérieure de la figure 3 est irréalisable puisqu'il

L'UTOPIE



LA REALITE

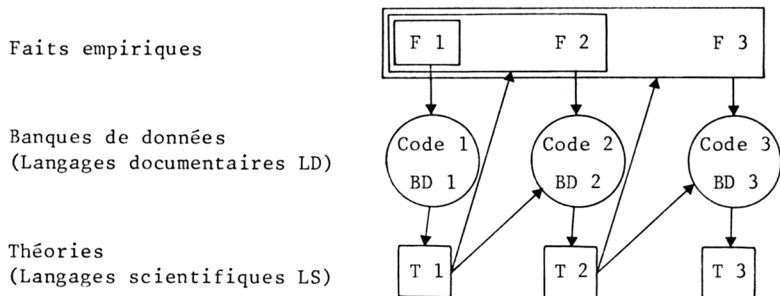


FIG. 3. — Schéma pour méditer du bon usage de la relation entre les théories et les faits empiriques.

ne peut entraîner qu'une *dégradation de l'information* contenue dans le corpus d'origine (par suppression ou amalgame) mais en tout cas pas un enrichissement (par dissociation ou adjonction). Tout ceci pour dire que le retour à la *totalité* des faits empiriques est toujours nécessaire, ce qui limite évidemment la rapidité et l'efficacité de la mise à jour, donc la souplesse de l'instrument.

Ce problème existe également au niveau de la publication imprimée, mais on aurait tort de le sous-estimer au niveau des banques de données électroniques, *surtout si celles-ci sont de grandes dimensions* (ce qui est une autre manière de retrouver notre première objection au sujet de l'utilité des corpus).

Ces diverses remarques nous amènent à poser une alternative. Est-il plus profitable :

1. de travailler sur des corpus très étendus contenus dans des banques de données institutionnalisées en sachant que l'on aura obligatoirement les mains liées par le niveau d'analyse des données ? La stratégie proposée, jointe aux difficultés de mise à jour, *oriente* en effet nécessairement la recherche.
2. de travailler sur des corpus de plus petites dimensions (dont nous ne voyons pas pourquoi ils ne seraient pas électroniques) en adoptant une stratégie nouvelle sur deux plans :
 - relations plus étroites entre les objectifs posés et les faits empiriques mobilisés (procédures d'échantillonnage);
 - va-et-vient plus fréquent entre les théories (établies sur des faits limités) et d'autres faits (également limités) (procédures de validation) ?

Plus simplement vaut-il mieux se donner les moyens de la guerre ou pratiquer l'art de la guérilla ?

3. CONSÉQUENCES SUR L'ENREGISTREMENT ET LA DIFFUSION DES DONNÉES ARCHÉOLOGIQUES.

Le moment est venu de tirer la leçon des deux expériences analysées, sur le plan de l'enregistrement des données d'une part (que nous avons largement abordé), sur le plan de leur diffusion d'autre part (dont nous avons moins parlé).

3.1. *L'enregistrement des données*

Le défaut commun aux deux banques de données réside dans le postulat implicite admettant un recours nécessaire au *plus grand nombre de données possible*. Il tire son origine de la méthode même de la recherche archéologique qui procède par accumulation d'indices.

Grâce aux progrès faits dans l'épistémologie de notre discipline on enseigne désormais qu'il est nécessaire de subordonner les langages descriptifs à des projets scientifiques définis et qu'il ne sert à rien de décrire pour décrire. Cette manière de voir constitue une première révolution épistémologique de notre discipline.

Parallèlement ne pourrait-on pas concevoir une adéquation identique des objets mobilisés aux objectifs visés ?

C'est le problème de l'échantillonnage que nous avons évoqué. Cela pourrait être une seconde révolution dans l'archéologie. Cette manière de voir entraîne naturellement une tout autre conception des banques de données : insertion dans des projets de recherche limités, cadre analytique plus spécifique, données moins nombreuses, mise à jour plus facile, contrôle plus efficace. C'est ce que nous avons tenté de faire à propos de l'étude du site préhistorique du Petit-Chasseur à Sion ¹, d'abord par répugnance instinctive envers les grands projets technocratiques et par « plaisir », maintenant par conviction délibérée et, nous l'espérons, logiquement et explicitement fondée. Cette position entraîne du reste d'autres conséquences sur le plan institutionnel et humain qui dépassent nettement le cadre strict de l'archéologie.

J.-C. Gardin compare à juste titre (Digard, vol. 1, p. 247) les conséquences entraînées par la création de grandes bases de données aux changements provoqués par la révolution industrielle.

« L'on peut craindre en effet que le prix de la rationalité promise dans les voies de la connaissance ne soit ici comme ailleurs fort lourd, et qu'à vouloir s'en prendre aux travers

¹ A. GALLAY. *Stèles néolithiques et problématique archéologique* (à paraître).

d'une activité agréable au moins à quelques-uns, on ne finisse par imposer à un plus grand nombre l'ennui de besognes impersonnelles et répétitives (...) ».

Que le mouvement annoncé en direction des grands centres de gestion de l'information « soit une manifestation de plus de ce que certains ont appelé la « révolution cybernétique » de l'an 2000, rien n'est plus clair ; que celle-ci à son tour ne soit finalement que l'extension aux activités de l'esprit de ce qu'aura été la révolution industrielle pour les travaux de la main, beaucoup l'ont dit aussi » ; ou encore (du fait de la séparation entre gestion de l'information et recherche) : « On risque de voir se développer, dans l'univers ainsi transformé de la recherche historique, une espèce de prolétariat de la mécanologie peu conforme à l'imagerie égalitaire et individualiste de l'érudition traditionnelle » (p. 248).

Les lendemains qu'on nous promet ne sont donc pas roses car il est évident que toute la problématique des luttes sociales se posera (et se pose déjà) à ce niveau de la recherche scientifique. A nous donc de choisir la voie qui permettra d'éviter dans ce domaine le renouvellement des sanglantes expériences découlant de la révolution industrielle, en proposant des objectifs mieux pensés et une répartition des tâches plus humaine. Cette position n'est pas une attitude de refus de toute technologie avancée, bien au contraire, mais une recherche d'un nouvel équilibre, ce qui nous entraîne au cœur d'un débat très actuel dans lequel les positions d'Ivan Illich auraient plutôt notre sympathie.

Soulevons pour terminer une dernière question d'ordre tout à fait général. Nous nous demandons si les « failles » que nous avons trouvées dans les deux expériences analysées ne proviennent pas aussi d'une certaine confusion entre projets compilatoires et projets scientifiques. Dans les deux cas (mais surtout dans le premier) les langages utilisés tendent à se rapprocher de langages scientifiques (LS) très complexes. Cette situation entraîne une lourdeur certaine de l'appareil et en limite son emploi. D'où notre suggestion : les grands projets de banques de données ne devraient-ils pas se limiter à des langages documentaires plus simples (LD) orientés vers la recherche de l'information mais non vers le traitement de cette dernière dans le cadre de projets scientifiques ? La distinction peut paraître futile, elle est en fait essentielle. L'avenir des banques de données en archéologie en dépend largement.

3.2. *La diffusion des données.*

On admet actuellement que le développement des banques de données permet de débarrasser la littérature imprimée des données descriptives pour la consacrer entièrement aux produits des activités « supérieures » de l'esprit, c'est-à-dire les théories scientifiques. Cette position est naturellement valable entièrement dans le cadre des grands corpus sur ordinateur présentés par quelques-uns. Il nous reste à examiner maintenant l'influence des positions que nous avons défendues ici (à propos de l'enregistrement) sur les modalités de la diffusion imprimée des connaissances. Deux faits nous incitent à nuancer la position « officielle » des tenants des banques de données.

1. Le premier résulte des commentaires provoqués par le schéma de la figure 1. Il ne concerne que la publication des matériaux des sites.
2. Le second résulte de la limitation voulue des faits mobilisés dans les constructions de l'archéologie. Il concerne à la fois les sites et les recherches thématiques.

Dans les deux cas il nous paraît indispensable de rendre compte des faits mobilisés dans une construction, ne serait-ce que sommairement, afin de rendre objectif le lien réunissant les données aux interprétations. Il n'en reste pas moins que la forme ne saurait être celle de la littérature traditionnelle, dont on a bien montré l'inadéquation totale¹ : ordre

¹ J.-Cl. GARDIN. *Mérites et limites de la formalisation en archéologie*. Cours donné au Département d'Anthropologie de l'Université de Genève en 1976.

de présentation des explications suivant l'ordre de genèse ¹, fondements logicosémantiques non explicites, résultats exposés de manière diffuse, objectifs non explicites, etc., et qu'une réflexion nouvelle doit avoir lieu quant aux modalités de présentation des données et des résultats. Mais d'autres ont traité, mieux que nous, ce sujet qui dépasse d'autre part le cadre de cette réflexion.

RÉSUMÉ

On présente un compte rendu et une critique de deux expériences de banques de données en archéologie. La première est le projet d'H. de Lumley consacré à l'enregistrement des données récoltées sur les sites préhistoriques; la seconde (F. Digard) est une expérience pilote ponctuelle consacrée aux cylindres orientaux.

Dans les deux cas les auteurs admettent implicitement que la recherche doit s'appuyer sur un maximum de faits. Cette position est discutée et l'on propose une alternative permettant d'éviter les inconvénients des énormes corpus institutionnalisés (difficultés de mise à jour du système analytique, industrialisation de l'activité intellectuelle, etc.): la création de banques de données plus limitées liées à des procédures d'échantillonnage et de validation dont il reste à définir les modalités.

*Département d'Anthropologie
de l'Université de Genève
12 rue Gustave-Revilliod,
CH 1227 Carouge-Genève.*

¹ Pour un bon exemple de cette situation lire: A. GALLAY, *Le Néolithique moyen du Jura et des Plaines de la Saône*, Antiqua, 6. Société suisse de Préhistoire, Frauenfeld, Huber (à paraître).