

L'UTILISATION du CALCUL ELECTRONIQUE  
en HISTOIRE ECONOMIQUE et SOCIALE.

par Marcel MASSARD

Le titre de cette communication surprendra vraisemblablement : le calcul, surtout électronique, ne paraît pas devoir être d'un emploi courant dans les sciences humaines. On s'attendrait plutôt à voir évoquer « l'Informatique ». Le lecteur sera sans doute déçu, mais c'est à dessein que nous avons renoncé à ce terme. En effet, l'usage de ce néologisme, qui n'a d'équivalent dans aucune langue étrangère, suggère trop fréquemment quelques procédés magiques capables de suppléer à l'intelligence ou à la réflexion humaines en recourant aux « cerveaux électroniques ».

La réalité est à la fois plus banale et plus complexe. Le « calculateur électronique », « l'ordinateur » ne supprime pas les tâches traditionnelles de l'historien, pas plus que celles de la plupart des chercheurs des autres disciplines. Le dépouillement des sources historiques, la collecte des renseignements, comme l'interprétation des résultats requerront longtemps les efforts et la sagacité de l'historien. Mais le calcul électronique fournit désormais un puissant outil d'analyse de la matière historique : chaque fois que celle-ci se présente sous une forme abondante, quantifiable ou formalisable mathématiquement, l'ordinateur permet d'organiser la masse de renseignements numériques ou de révéler des relations qui n'étaient pas apparentes au premier abord. Cette technique n'a rien de magique : instrument remarquable, mais complexe, elle nécessite de la part de son utilisateur un apprentissage délicat.

La première caractéristique du calcul électronique est son universalité d'application : chaque fois que le chercheur (et c'est particulièrement le cas du chercheur en Histoire économique et sociale, dont nous évoquons les problèmes) se trouve confronté avec des informations chiffrées (par exemple : prix, salaires, productions, dénombrements démographiques) ou formalisables mathématiquement (par exemple : possibilité de coder numériquement professions, lieux de naissance, sexe des individus) qui entraînent des calculs longs et complexes, l'utilisation de l'ordinateur permet de procéder rapidement, sûrement et efficacement. Certes, cette méthode n'est pas sans inconvénients pour le spécialiste des sciences humaines. Le premier réside dans son coût financier, que nous analyserons plus loin. Le second provient de la nécessité de savoir manier l'outil utilisé : en l'occurrence, l'usage du calcul électronique suppose une certaine familiarité avec la pratique du « software ». Celui-ci implique quelques connaissances mathématiques et l'utilisation rigoureuse du raisonnement logique : disciplines peu habituelles aux historiens et pour eux inquiétantes. A vrai dire, si ces connaissances sont indispensables, elles n'ont ni l'étendue, ni la complexité que supposent les non-initiés.

Un troisième inconvénient a parfois été invoqué : la comparaison des méthodes de la mécanographie électronique avec certaines techniques de traitement graphique de l'information a révélé chez ces dernières (1) une

(1) Dont M. VOGLER nous a donné une brillante démonstration à propos de son étude sur les Pasteurs dans le Palatinat au début du XIXe siècle. (Cf. Bulletin du Centre, 1970, n° 4)

ingéniosité et une originalité, en comparaison desquelles la rigueur et la technicité de l'usage de l'ordinateur apparaissent inhumaines et dépourvues de pittoresque. Il est incontestable que, dans certains cas, des méthodes originales peuvent avantageusement entrer en concurrence avec le calcul électronique. Toutefois, envisagée dans l'ensemble du traitement de l'information, cette technique comporte incontestablement des avantages déterminants.

En premier lieu, l'efficacité : seul le traitement par ordinateur permet la précision, la rapidité et la fiabilité dans le cas d'opérations de calcul un peu longues ou complexes. Au delà d'un certain maximum d'informations (par exemple, en démographie, quelques milliers d'individus), aucune autre méthode n'est concevable. Ses possibilités, d'autre part, sont illimitées, qu'il s'agisse du nombre d'informations à traiter, de la nature des informations (statistiques, économiques, sociales, politiques, sociologiques, linguistiques) ou de la subtilité des conclusions que l'on peut tirer de ces informations, en leur appliquant le traitement mathématique adéquat. Ensuite, le coût financier marginal, ainsi que le temps de travail, est très faible : dès lors qu'un programme a été mis au point pour un problème donné, tout problème semblable se trouve résolu avec une rapidité, une précision et une modicité inégalées. Enfin, contrairement à certaines préventions à l'égard de la rigueur inhumaine de l'ordinateur, il faut souligner le haut intérêt intellectuel de cette technique. « L'ordinateur travaille pendant que le chercheur réfléchit » a-t-on dit : en effet, le calcul électronique nécessite la recherche des processus adéquats pour interroger la machine et traiter l'information dont on dispose ; cela entraîne, de la part du chercheur en sciences humaines, la maîtrise, en plus des techniques d'investigation propres à sa discipline, d'un outil complexe d'appréhension et de traitement des réalités. Apprentissage qui nécessite l'appel à une logique irréfutable, à un incessant travail d'invention, de création et de généralisation, en même temps que cessent les opérations longues répétitives et fastidieuses des méthodes traditionnelles. Par contre, le calcul électronique laisse à l'historien la partie essentielle de son travail : l'interprétation de la documentation qu'il a explorée, l'ordinateur lui ayant fourni l'inventaire total de toutes les relations dont certaines sont inapparentes autrement.

Tous les domaines de l'Histoire, dès lors qu'ils présentent des données numériques ou statistiques d'une certaine importance, sont justiciables d'un traitement par ordinateur. Nous nous bornerons à donner quelques exemples déjà classiques. L'Histoire économique offre, bien sûr, un domaine de choix : les prix, les salaires, les productions ; les transformations automatiques de toutes données en pourcentage, en indices, en prix en monnaie constante ; le calcul des corrélations. Même les éléments qualitatifs, peuvent, après transformations appropriées, fournir la base d'un passage en machine : pour une population donnée, l'origine de la fortune (héritage, acquisition), sa nature (foncière, immobilière, mobilière), sa mobilité (sur une certaine période). L'histoire sociale, soit qu'elle reste limitée à l'étude d'une classe ou catégorie restreinte, soit qu'elle envisage l'étude d'ensembles humains importants, peut trouver, dans le calcul électronique, le remarquable instrument d'analyse dont l'absence limitait jusqu'ici l'étendue des recherches. Ainsi la démographie historique (pyramide des âges, structure d'une population par catégories socio-professionnelles, origines géographiques d'une population, reconstitution des familles où s'égarèrent les modernistes) et l'étude approfondie de certaines catégories sociales (hommes de loi, clergé catholique et protestant) se révèlent des terrains d'élection du calcul électronique.

L'histoire politique, enfin, renouvelle ses thèmes avec le développement de domaines nouveaux, grâce au calcul électronique : sociologie électorale, étude du personnel politique, inventaire du vocabulaire politique propre à certains partis, certains organes de presse, certaines idéologies.

Ainsi le calcul électronique ouvre des perspectives nouvelles et stimule des ambitions. Si nous avons contesté précédemment la réalité de quelques uns de ses inconvénients, nous n'avons qu'évoqué le plus important : son coût financier élevé. Il est malheureusement nécessaire d'entrer dans le détail de ce problème. Le coût du calcul électronique est, en effet, le principal obstacle à son utilisation ; ce coût se répercute sur trois domaines : la préparation des données, l'utilisation de la machine, la programmation.

- a) La préparation des données nécessite l'utilisation de deux séries de fiches ou cartes : des fiches

primaires, comportant imprimés les titres des rubriques, que le chercheur complétera en clair au fur et à mesure de ses investigations, et les cases servant à coder numériquement ces renseignements. Le prix de ces documents est évidemment fonction du format des fiches, de leur nombre, de l'importance des caractères imprimés qu'elles comportent : cette dépense est variable, mais non négligeable. Ces fiches permettent de passer à l'opération suivante : la perforation des codes numériques sur les cartes standard à 80 colonnes. Sauf s'il s'agit d'un nombre très restreint de cartes, cette opération ne peut être réalisée que par une entreprise utilisant des professionnels spécialisés dans ce genre de travail (I.B.M., BULL, SEMA, ou INSEE, CNRS qui acceptent, dans certains cas, des travaux à façon). Cette réalisation est onéreuse (1). Les prix d'autre part, variant de façon très sensible selon les entreprises : il convient de demander plusieurs devis précis à différents fournisseurs. Enfin, la perforation peut être simple ou accompagnée de vérification : cette seconde procédure double à peu près le prix de la seule perforation. Il est néanmoins vivement recommandé de l'utiliser : les erreurs de perforation coûtent beaucoup plus cher postérieurement, lors de l'introduction des cartes dans le calculateur, si elles exigent des passages répétés dont le coût est très élevé.

b) L'utilisation de la machine (l'ordinateur) est, en effet, de loin, l'opération la plus coûteuse de la mécanique électronique. Cette utilisation est tarifée en « temps-calcul » : c'est à dire que le prix est fonction de la durée de fonctionnement de l'ordinateur, pour le travail qu'on lui donne à effectuer. Les tarifs varient également selon la taille de l'ordinateur, mais il est bon de savoir qu'un gros calculateur, d'un prix d'utilisation plus élevé, travaille beaucoup plus vite qu'un petit (2). Le temps de calcul est, d'autre part, fonction de la complexité des résultats que l'on désire obtenir : pour un traitement relativement simple l'ordinateur est capable de « digérer » plusieurs milliers de cartes en quelques minutes. Pour ce traitement, les entreprises privées (I.B.M., SEMA) proposent des prix à peu près équivalents, mais lourds pour les budgets des chercheurs en sciences humaines. Le CNRS, par contre, offre à ces derniers des possibilités d'attributions gratuites de temps-calcul : il faut malheureusement prévoir à temps ses besoins ; les demandes doivent être faites au mois de juin de l'année précédant l'année civile d'utilisation.

c) La programmation (3), enfin, est une opération dont le prix est très variable selon la méthode utilisée par le chercheur. Si l'on est obligé de faire appel à un spécialiste du software (analyste ou programmeur) pour la mise au point du programme, le coût est très variable, mais en général très élevé, l'écriture d'un programme demandant de quelques jours à quelques mois de travail. Une solution économique consiste à utiliser un programme standard (soit vendu par une entreprise privée - soit accessible en bibliothèque, au CNRS) : l'inconvénient est qu'un tel programme n'est pas toujours parfaitement adapté au problème que l'on veut traiter et qu'il comporte des limites qu'il convient de parfaitement respecter. L'élégance suprême pour le chercheur utilisant le calcul électronique est d'écrire ses propres programmes : pratique souhaitable que nous avons évoquée au début de cette communication, mais qui entraîne un apprentissage et un effort initial quelque peu contraignants.

Le problème du coût résolu, l'utilisation du calcul électronique dans les recherches de sciences humaines nécessite une technique opérationnelle rigoureuse. D'une façon générale, une entreprise de cette nature exige dans un premier temps une double opération préliminaire. Il faut d'abord réaliser une investigation précise des données générales du problème, ceci par un sondage des sources qui permettra de la façon la plus rapide pos-

(1) Par exemple : perforation et vérification de 12.000 cartes, sur 20 colonnes, en caractères numériques seulement : 100 à 200 francs selon l'entreprise.

(2) A titre d'indication très générale, le prix d'utilisation d'un « gros calculateur » (I.B.M. 360/50, C.D.C. 3.600) est de l'ordre de 1.000 f. l'heure.

(3) C'est à dire la mise au point (« l'écriture »), d'un « programme » (= séquence d'instructions données au calculateur pour lui permettre de traiter et de résoudre le problème posé).

sible de mesurer assez précisément l'ampleur de l'information que l'on va traiter, le nombre et le type de renseignements que l'on obtiendra et les possibilités de codage numérique qui s'offrent dans ces conditions (1). Ceci réalisé, il convient de passer immédiatement à la phase ultime de l'exploitation, c'est à dire des conditions de la programmation : nature et caractéristiques du programme disponible, nécessité de créer un programme propre à ce problème, type et performances du calculateur qui sera utilisé.

Cette double investigation préliminaire effectuée, l'organisation du travail doit procéder de façon régressive. Il faut remonter d'aval en amont, en fonction des caractéristiques et des nécessités de la programmation, pour plier à celles-ci la formulation ou la codification des données que le flair du chercheur a su localiser. Son travail alors ne différera des méthodes traditionnelles que par l'obligation de donner à sa documentation la forme exigée par l'ordinateur pour la traiter.

Ce dernier, sans rien inventer, sans rien interpréter, mais avec une logique irréfutable, une rapidité inégalée, une précision et une sûreté incroyable rendra à l'historien, conformément à ses interrogations, la matière de ses recherches. Pour celui-ci le travail essentiel commencera : l'interprétation historique. L'outil qu'est le calcul électronique n'aura fait qu'accélérer et rendre plus efficace le travail de l'homme.

(1) Nous nous permettons, pour illustrer la technique, d'emprunter à Monsieur VOGLER l'exemple de l'étude d'une catégorie sociale à un moment donné (« Les Pasteurs protestants dans le Palatinat au début du XIVe siècle »). Le sondage dont nous parlons a pour but un survol de toutes les sources disponibles pour fixer un nombre maximal et minimal d'individus qui feront l'objet de l'étude et pour inventorier les « variables » possibles qui caractérisent les Pasteurs et leur famille (par exemple : lieu de naissance, profession du père, du beau-père, université où les études ont été faites, âge au premier mariage). Même si pour un certain nombre d'individus étudiés, pour certaines sources, pour certaines époques les renseignements manquent, toutes les variables observées doivent être relevées.

## DISCUSSION

Deux types de questions reviennent fréquemment lors du débat qui suit la communication de M. MASSARD : d'une part les auditeurs demandent des précisions sur les conditions matérielles de l'utilisation de l'ordinateur ; d'autre part, ils s'interrogent sur les bénéfices que l'histoire peut tirer de l'informatique, l'histoire économique et sociale en particulier.

Pour permettre à M. LEON et M. GARDEN d'évaluer les coûts d'une opération, M. MASSARD estime le prix d'une machine à perforer à un minimum de 5000 francs, et il en déduit que le chercheur a tout intérêt à utiliser et à payer les services d'un perforateur. Il rappelle que le coût horaire est déterminé non par le nombre de fiches, mais par la complexité des questions posées. Le coût horaire est élevé, certes, mais il faut très peu de temps à l'ordinateur pour travailler. A titre indicatif, M. MASSARD affirme, qu'il lui faut moins d'une heure pour donner les résultats concernant 12 000 fiches.

Ces questions de temps et de coût précisées, M. MASSARD explique que la perforation des fiches relève d'une technique assez simple qui rappelle celle de la dactylographie. Il insiste pourtant sur le fait que toute erreur de frappe est irrémédiable.

A une question de M. GARDEN, M. MASSARD répond que le passage des cartes dans l'ordinateur donne naissance à une bande magnétique, que l'on peut conserver dans des bibliothèques prévues à cet effet. L'informatique n'en étant qu'à ces débuts, on ne voit pas encore si ces bandes sont conservables indéfiniment. On voit seulement qu'elles sont extrêmement fragiles et qu'un orage peut suffire à les détruire. Les cartes perforées, elles, constituent des archives d'un type traditionnel et donc plus aisées à conserver.

M. GARDEN demande aussi s'il est tenu compte lors de l'établissement des cartes, de certaines variables dont les données sont susceptibles de modifications. M. MASSARD répond par l'affirmative ; des cases en blanc sont prévues, qui représentent l'absence momentanée de ces renseignements. Illustrant l'exemple des contrats de mariage au XVIIIe, choisi par M. GARDEN, contrats souvent incomplets, M. MASSARD explique que la machine ne rejette pas les fiches incomplètes, mais n'opère le calcul que sur les indications fournies.

Revenant sur la question du sondage au 1/15e, dont M. MASSARD a parlé dans son exposé, M. VITAL-CHOMEL demande si ce pourcentage a été choisi pour des raisons mathématiques. M. MASSARD précise qu'il a été établi par un calcul de l'I.N.S.E.E. basé sur une constatation de nombreux cas identiques, mais réalisé selon la technique d'un sondage aléatoire, cependant tenu pour hautement scientifique par les statisticiens.

Dans le cas de cet échantillonnage au 1/15e, la duplication doit être faite, sous peine d'obtenir des pourcentages faux (réponse à M. GARDEN)

Pour ce qui est de l'intérêt de la méthode, M. MASSARD se déclare convaincu de sa valeur pour la confirmation rapide et certaine des hypothèses de travail. D'autre part, il pense qu'en matière de démographie elle permet d'aboutir à des résultats tout à fait nouveaux.

M. DURAND se déclare déçu par une telle rigidité. Il constate que la machine n'enregistre pas les transformations d'un dénombrement à l'autre. M. MASSARD estime qu'un programme différent pour chaque dénombrement permettrait de définir une évolution. M. GARRIER l'approuve : la machine ayant fourni deux séries complètes de pourcentages, il ne reste plus au chercheur qu'à établir les comparaisons qu'il juge fondamentales.

En conservant les résultats, les bandes ou les fiches, on peut aboutir à des études comparatives. M. MASSARD répond à M. LEQUIN qu'en codant plusieurs centaines de communes, on peut faire ressortir les mouvements privilégiés, qui animent les déplacements migratoires. Les possibilités restent donc immenses

M. VITAL-CHOMEL pense que ces méthodes deviendront bientôt classiques et seront diffusées dans l'enseignement supérieur. M. GARDEN songe à la création d'un C2, qui permettrait aux étudiants s'initiant à la recherche de connaître les possibilités offertes par l'informatique.

M. GARRIER rappelle qu'un chercheur, qui désire écrire lui-même son programme, peut suivre un stage de huit jours organisé par le C.N.R.S. M. MASSARD précise qu'aucune formation mathématique particulière n'est requise pour l'initiation à l'informatique et qu'il suffit d'avoir le niveau d'un élève de lycée d'une section classique.

(Nous pouvons aujourd'hui préciser, que l'expérience tentée en 1970 - 1971 par l'Education Nationale confirme cette assertion).

=====