

Les fiches perforées “ par caractéristique ” et leur application aux enquêtes par questionnaire

par Roland DEVAUGES et Jacques JENNY

Le propos de cet article est de décrire une technique de traitement de l'information qui n'est pas originale à proprement parler, puisqu'elle est couramment utilisée dans le domaine de la documentation, mais que nous estimons méconnue à tort dans certains domaines de la recherche empirique. Il s'agit du système dit « à entrée de l'information par caractéristique » (ou par terme) (1), qui s'oppose radicalement au système beaucoup plus répandu « à entrée de l'information par individu » (ou par document ou par objet) (2).

Après avoir rappelé les différentes techniques de traitement de l'information qui se rattachent à ces deux grands systèmes, nous présenterons les avantages et les inconvénients propres à chacune d'elles sous leurs principaux aspects. Nous illustrerons ensuite la technique des fiches par caractéristique par l'exemple concret d'une application en cours. A cette occasion, nous exposerons brièvement les arguments qui nous ont conduits à préférer ce système à celui par individu. En conclusion, nous ferons part de nos projets actuels en vue de doter ce procédé de moyens techniques propres à améliorer son rendement.

(1) Système parfois désigné sous le terme générique importé des Etats-Unis de *peck-a-boo*, dont l'application aux travaux documentaires semble remonter dans ce pays à l'année 1915 et en France, à 1923 (Brevet Borgeaud-Liber).

(2) Chaque auteur d'un article ou d'un ouvrage sur ces questions a sa terminologie personnelle, ce qui est le propre de toute spécialité embryonnaire. Nous essayerons de nous faire comprendre ici avec le maximum de clarté et de concision en assortissant les termes employés de brèves définitions, de synonymes ou d'illustrations concrètes. D'autre part, pour éviter de trop longs développements technologiques abstraits, nous indiquerons la marque des principaux modèles fabriqués ou commercialisés en France, qui se rattachent aux différentes méthodes de traitement de l'information. Ces indications n'ont, bien entendu, aucune intention publicitaire ni prétention d'inventaire exhaustif.

Les fiches perforées « par caractéristique »

Toute la gamme des techniques de traitement de l'information peut se classer en deux grandes catégories selon que leur support matériel est en correspondance avec les sources émettrices d'information ou avec le contenu lui-même de l'information (cf. figure 1).

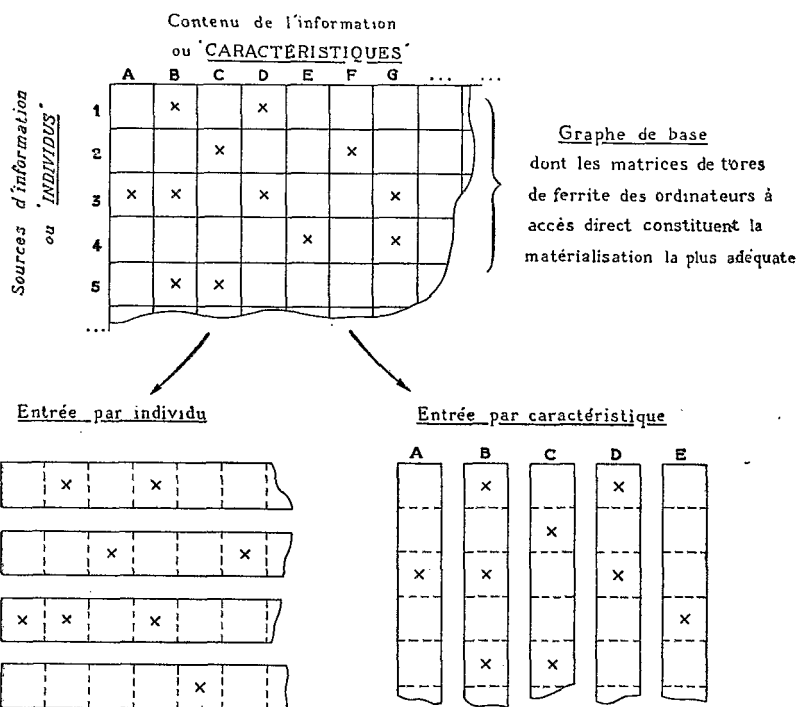


Fig. 1. — Schémas théoriques illustrant les deux systèmes de traitement de l'information.

Dans ces deux schémas, les bandes allongées horizontalement ou verticalement représentent les *supports* matériels, et les croix représentent les *traces* matérielles — supports et traces qui permettent les divers types d'exploitation rappelés dans le tableau ci-dessous.

Il faut noter que le nombre de traces matérielles reste théoriquement identique dans les deux systèmes à celui de la matrice de base. Seul change leur rangement en lignes ou en colonnes.

Les sources émettrices et le contenu codifié de l'information constituent des ensembles couplés de types divers, comme par exemple :

Sources	Contenus
Documents de bibliothèque	Mots-clés d'analyse
Personnes ou organismes de fichier administratif	Caractéristiques codifiées
Objets ou articles en magasin	
Questionnaires d'enquêtes	Réponses aux questions

Quant aux supports qui matérialisent l'un de ces ensembles et aux traces qui matérialisent l'autre, ils sont également de types très variés, comme par exemple :

SUPPORTS	TRACES	TYPES D'EXPLOITATION	EXEMPLES DE MODÈLES
Simplex fiches vierges	Notations manuscrites Signalisation optique sur la marge supérieure	Strictement <i>manuelle</i> et/ou <i>visuelle</i>	Fiche auteur ou matière Intercalaires ou « cavaliers » synoptiques
Fiches sans préperforation	Encoches marginales	<i>manuelle</i> (à l'aide de simples aiguilles ou broches de tris)	Fiches SATO (brevet SABOX) ou ZATOCODING
Fiches à préperforations marginales			RAPIDTRI (C.F.M.) MECANALYSE (KEYSORT), etc.
Fiches à préperforations centrales	Encoches centrales	ou <i>mécanique</i> (à l'aide de trieurs-sélecteurs par gravitation ou par vibration électrique)	DETECTRI, DEQUEKER, SELECTRI, TEXOLLETTE
Pellicules ou films photographiques	Taches ou mots-clés microfilmés (négatif du support)	<i>optique</i> et <i>photo-électrique</i>	FILMOREX, ECCESTRON, TELECESTRON (Procédés Marcel LOCQUIN)
Fiches quadrillées	Perforations centrales	<i>visuelle</i> et/ou <i>photo-électrique</i>	SPHINXO (Borgeaud), SELECTO (Cordonnier), DEQUEKER (3)
Cartes « mécanographiques »	Perforations rectangulaires par « colonnes »	<i>électromécanique</i> (lecture par balais en continu)	Matériel mécanographique BULL, I.B.M., I.C.T., REMINGTON, etc.
Bandes « mécanographiques »	Perforations rondes par « canaux »		
Cartes « mécanographiques »	Perforations rondes indépendantes du dessin de la carte	<i>électromécanique</i> (lecture par plongeurs en discontinu)	SAMAS-I.C.T., TEXOLER
Rubans et boucles magnétiques	Traces de magnétisation	<i>électromagnétique</i> (à accès séquentiel ou semi-direct par déroulement)	« Ordinateurs à mémoire incorporée » BULL (série GAMMA), I.C.T., I. B. M., S. E. A., REMINGTON (UNIVAC 1107), etc.
Tambours et disques magnétiques		<i>électromagnétique</i> (à accès direct avec « adresses » individuelles)	
Matrices de tores de ferrite, de pellicules ferro-magnétiques			

(3) Il s'agit des seuls modèles destinés théoriquement au type de traitement de l'information « par caractéristique ». Leurs équivalents étrangers sont par exemple, en Allemagne, le modèle EKAHA et aux Etats-Unis, les modèles VICREF et TERMATREX-MINIMATREX (brevet Jonker) dont nous parlerons dans la suite de cet article parce qu'actuellement dotés des plus grands perfectionnements techniques (mais non commercialisés en France). Nous verrons que d'autres supports matériels peuvent également se prêter à ce type de traitement de l'information « par caractéristique » (en particulier, les cartes mécanographiques et les mémoires magnétiques à accès direct).

Les fiches perforées « par caractéristique »

Bien que ces divers types de support matériel présentent entre eux de nombreuses analogies, nous nous limiterons à l'examen des techniques utilisant la carte ou la fiche perforée qui est de loin le support matériel de base le plus typique et le plus couramment utilisé (4).

Comme l'illustre la *figure 1*, dans le premier des deux grands systèmes de traitement de l'information distingués plus haut — celui par individu — c'est l'ensemble des cartes ou fichier, qui correspond à l'ensemble des individus, des documents, des objets ou des questionnaires, et c'est l'ensemble des perforations de chaque carte qui correspond à l'ensemble des caractéristiques, des mots-clés, des réponses, des items... Par exemple, chaque carte mécanographique représente une personne et chaque position de perforation sur toute carte représente une seule et même réponse ou caractéristique donnée. Dans le second système — celui par caractéristique — c'est l'inverse : par exemple, chaque fiche représente une réponse ou une caractéristique donnée et chaque position de perforation sur toute fiche représente une seule et même personne.

Dans la pratique actuelle les domaines d'application de ces deux grands systèmes se recouvrent partiellement. Toutefois, à première vue, le procédé par individu semble mieux convenir aux exploitations comptables et statistiques de documents administratifs ou scientifiques (fichiers de personnel ou de clientèle, questionnaires d'enquête extensive, dépouillements de recensements) et le procédé par caractéristique aux opérations de sélection documentaire ne comportant ni dénombrements ni calculs mais seulement sélection et identification de documents pertinents; on trouve en effet organisés sur ce principe les services ou centres de documentation suivants : *Bulletin signalétique* du Centre national de la Recherche scientifique, Laboratoire de Psychologie sociale (5), Centre de Documentation préhistorique (6), pour ne citer que les plus connus en sciences humaines.

En fait, d'une part le premier système connaît aussi des applications remarquables dans le secteur de la documentation, parmi lesquelles on peut citer la trieuse BULL D 3 D et le programme d'analyse automatique du C.A.D.A.N. (Centre d'Analyse documentaire de l'Afrique Noire) sur ordinateur I.B.M. 7090 (7), et d'autre part le deuxième système a déjà été utilisé avec profit dans certains types de recherches scientifiques, généralement des enquêtes d'exploration (telles que certaines recherches médicales ou certaines enquêtes intensives du Groupe d'Ethnologie sociale dirigé par P. H. Chombart de Lauwe, ou l'enquête en cours sur les adolescents scolarisés du Togo dont nous parlerons plus loin). Il existe enfin des techniques alliant ces deux systèmes, tel le procédé « Uniterm », assez répandu dans les services de documentation américains mais peu connu semble-t-il en France et le tout récent matériel américain « Termatrex » et « Minimatrex » avec convertisseur carte/individu-fiche/caractéristique (8).

(4) Rappelons que même les grands calculateurs et ordinateurs électroniques et électromagnétiques utilisent encore actuellement la carte ou la bande perforée comme support matériel d'introduction des données en mémoire ou d'enregistrement du programme.

(5) PAGES, R., BOUILLUT, J., LEMAINE, G., VATAIRE, D. *Une grille d'analyse des documents en psychologie sociale*. 2^e édition. Paris, Laboratoire de Psychologie sociale de l'Université de Paris, 1962.

(6) GARDIN, J. C. « Problèmes de la documentation ». *Diogenes* (11), juillet 1955. GARDIN, J. C. « Documentation sur cartes perforées et travaux sur ordinateurs dans les sciences humaines ». *Revue internationale de Documentation*, 29 (3), 1962.

(7) IZARD, F. *An experiment in automatic data processing for the bibliography concerning Africa, South of the Sahara*. Rapport ronéotypé du C.A.D.A.N. 1961, cf. également la note précédente.

(8) Nous verrons plus loin que les mémoires d'ordinateurs « à accès direct » (mé-

Il n'existe donc pas de liaison systématique et nécessaire entre telle méthode de traitement de l'information et tel genre de travail scientifique, sinon peut-être par routine, mode ou ignorance (9). Un effort d'imagination nous paraît nécessaire pour distinguer les caractéristiques liées à telle ou telle technique de façon intrinsèque ou sous la pression de contingences quasi irréductibles et celles qui n'y sont liées qu'en raison de restrictions technologiques extrinsèques et passagères, donc perfectibles.

C'est pourquoi nous allons maintenant procéder à un rapide bilan des deux grands systèmes décrits plus haut, dans leurs applications les plus fréquentes actuellement mais aussi dans certaines de leurs utilisations plus rares, voire seulement potentielles — c'est-à-dire possibles à brève échéance et à relativement peu de frais.

L'examen portera successivement sur les critères suivantes :

A - *Capacité* de l'information enregistrée et traitée, que nous décomposerons en *capacité en « individus »* (10) d'une part pour rendre compte successivement des problèmes propres aux grands recensements, enquêtes extensives, vastes fichiers administratifs ou documentaires, etc. et *capacité en « caractéristiques »* (10) d'autre part pour rendre compte des problèmes propres aux enquêtes intensives, analyses documentaires approfondies, etc.

Le caractère intensif ou extensif de l'information, indiqué par le rapport N Caractéristiques/N Individus sera examiné sous la rubrique D des « autres considérations pratiques » (§ 8 a, page 427).

B - *Malléabilité*, ou variabilité, c'est-à-dire possibilité d'extension, de réduction ou de permutation des données de l'information. Nous décomposerons également ce critère en variabilité de *l'ensemble des individus* d'une part pour rendre compte des problèmes spécifiques de la tenue à jour de fichiers administratifs ou documentaires, d'enquêtes non instantanées et longitudinales, etc. et variabilité de *l'ensemble des caractéristiques* d'autre part pour rendre compte des problèmes spécifiques des travaux exploratoires comportant une part d'analyse de contenu par « tâtonnement, essais et erreurs », avec modifications et révisions rétrospectives du code, etc.

C - *Souplesse d'utilisation* et commodité pratique aux phases successives d'enregistrement, de lecture, de dénombrement et de traitement mathématique.

D - *Autres considérations pratiques*, telles que coût financier, durée, importance et qualification du personnel d'exécution, etc..., concernant l'établissement, puis l'exploitation du stock d'informations.

moires à tores de ferrite ou à pellicules ferro-magnétiques en particulier) permettent également un enregistrement et un traitement des données dans l'un ou l'autre système indifféremment; et l'on sait que ce type de mémoire est appelé à remplacer l'autre type « à défilement ou à déroulement » (rubans, tambours, etc.) à partir du moment où l'abaissement de son prix de revient permet d'en augmenter la capacité et où sa miniaturisation permet d'en réduire l'encombrement.

(9) Une meilleure connaissance des performances et de la variété des dispositifs techniques peut même libérer le chercheur d'un certain esclavage auquel risque de le contraindre une connaissance trop sommaire ou trop partielle. Les programmes de recherche et d'analyse documentaire ne doivent alors plus être définis en fonction de soi-disant contraintes technologiques mais selon leur logique interne, à laquelle doivent se plier les dispositifs matériels nécessaires à leur réalisation. On pourrait en dire autant de l'organisation collective des centres et des équipes de recherche ou de documentation, qui ne doivent pas sacrifier leurs objectifs propres à un quelconque engouement pour tel ou tel type de matériel.

(10) Rappelons que pour simplifier nous employons ces termes dans leur signification la plus générique : *individus* pour désigner aussi bien des personnes proprement dites, ou des groupes, ou des documents, ou des objets, ou des faits, d'une part; *caractéristiques* pour des réponses, des items, ou des mots-clés, d'autre part.

Les fiches perforées « par caractéristique »

Système à entrée de l'information
par individu

Système à entrée de l'information
par caractéristique

A. — LA CAPACITÉ

1. Capacité en individus

a) *Théoriquement illimitée*, puisque le support matériel correspondant à l'ensemble des individus est extensible à volonté.

(b) Pratiquement limitée, pour des raisons...

- de temps et de commodité dans les systèmes manuels et mécaniques (par ex. fiches à préperforations marginales et centrales, triées par paquets de 150 à 400 à la fois seulement)
- d'encombrement et de manipulation dans les systèmes de tri électromécanique
- de rapidité d'accès dans les mémoires à bandes magnétiques (la capacité et la rapidité d'accès variant en fonction inverse l'une de l'autre).

a) *Limitée* par la capacité même du support matériel utilisé (*).

b) Possibilité de multiplier la capacité par 2, 3...n en juxtaposant 2, 3...n jeux de fiches avec signes distinctifs (dans les limites de la commodité pratique).

« Selecto » prévoit par exemple jusqu'à 20 jeux de fiches à 5.000 positions, soit une capacité de 100.000 « individus ». Les fiches « Minimatrex » sont la copie microfilmée à l'échelle 1/30 de 10 fiches « Termatrex » juxtaposées comportant chacune 10.000 positions de perforation, soit une capacité de 100.000 « individus » sur un seul support.

2. Capacité en caractéristiques

a) *Limitée* par la capacité et le « dessin » du support matériel utilisé, par exemple :

- par le nombre de préperforations marginales ou centrales des fiches à exploitation manuelle et mécanique.
- par le nombre de colonnes des cartes mécanographiques BULL, I.B.M., I.C.T. (qui détermine le nombre de « questions », ou par le nombre de positions de perforation des cartes SAMAS - I.C.T. (qui détermine le nombre de « réponses » dans le système TEXOLER) (*).

b) Avec les trieuses munies de comparateurs (du type BULL D3) et avec les interclasseuses, à plus forte raison avec les ordinateurs, la capacité peut être multipliée pour répondre pratiquement aux besoins scientifiques les plus courants (**).

a) *Théoriquement illimitée*, puisque le support matériel correspondant à l'ensemble des caractéristiques est extensible à volonté.

b) Nécessité d'un classement systématique des fiches selon les caractéristiques afin de faciliter les recherches au sein du fichier (avec éventuellement des cavaliers synoptiques ou des encoches marginales pour pallier les risques de déclassement (**).

(*) C'est-à-dire dans le premier cas : 80 questions à 1 seule réponse parmi 2 à 10 possibles, avec les trieuses les plus

(*) Soit par exemple 1.000 positions de perforation dans les fiches « Sphinxo », 7.000 positions de perforation dans les

Système à entrée de l'information
par individu

simples; et dans le deuxième cas : 960 réponses, quel que soit le nombre de questions, avec l'explorateur « TEXOLER » adapté à une trieuse SAMAS - I.C.T. à 80 colonnes.

(**) Par exemple le calculateur électronique Gamma 30 de la Cie BULL peut traiter jusqu'à 1.000 questionnaires comportant chacun jusqu'à 375 questions (l'équivalent de 8 cartes perforées sur 47 colonnes).

Système à entrée de l'information
par caractéristique

fiches « Dequeker », 5.000, 8.000 ou 14.000 positions de perforation dans les fiches « Selecto », 800 à 960 (10 ou 12 lignes de 80 colonnes) dans les cartes mécanographiques standard à perforation rectangulaire Bull, I.B.M. ou I.C.T. Ces cartes, destinées au système « par individu », sont utilisables dans le système « par caractéristique » comme les fiches « Sphinxo » et « Selecto » auxquelles elles peuvent être préférées en raison de leur faible coût d'achat et de perforation et de leur bonne lisibilité (cf *infra* 5b et 6a). Avec les cartes « Samas-I.C.T. » à doubles perforations rondes, la capacité peut atteindre 1.600 ou 1.920 selon que l'on utilise 20 ou 24 lignes de 80 colonnes.

(**) Fiches « Dequeker » par exemple.

B. — LA MALLÉABILITÉ

3. Malléabilité de l'ensemble « individus »

Parfaite : Il suffit d'ajouter, de retrancher ou de permuter les seules cartes correspondant aux « individus » nouveaux, périmés ou remplacés.

Extension très *commode*, à condition que les « individus » soient numérotés sur les fiches-caractéristiques dans l'ordre de leur enregistrement.

Réduction et permutation moins commodes, mais possibles à l'aide d'une fiche masquant les « individus » périmés.

4. Malléabilité de l'ensemble « caractéristiques »

Incommode dans les systèmes manuels et mécaniques où il faut reboucher manuellement les perforations ou encoches caduques.

Pratiquement nulle dans les systèmes électromécaniques à cartes perforées, où le code est définitivement fixé.

Possible mais coûteuse dans les ordinateurs à mémoire magnétique incorporée.

Parfaite : Il suffit d'ajouter, de retrancher ou de permuter les seules fiches correspondant aux « caractéristiques » nouvelles, périmées ou remplacées.

C'est sans aucun doute la raison majeure qui peut faire préférer ce système pour les travaux d'exploration, qu'il s'agisse d'enquêtes ou d'analyses documentaires, nécessitant une liberté complète à l'égard du code.

Les fiches perforées « par caractéristique »

Système à entrée de l'information
par individu

Système à entrée de l'information
par caractéristique

C. — L'UTILISATION

5. Souplesse d'utilisation et commodité pour l'enregistrement des données (et leur perforation sur cartes)

a) *Enregistrement direct*, puisqu'il y a correspondance bi-univoque entre chaque individu et chaque unité du support matériel : par exemple on peut passer directement du questionnaire ou de la grille d'analyse codée à la fiche à perforation ou à la carte ou bande perforée (elles-mêmes éventuellement introduites automatiquement en mémoire magnétique pour traitement sur ordinateur).

b) Les opérations de *perforation* ont bénéficié du progrès continu qui caractérise le développement des machines mécanographiques mais n'en constituent pas moins encore le principal goulot d'étranglement (*).

c) *Contrôle de perforation* par ré-exécution de la tâche (ovalisation des perforations, etc.), doublant pratiquement la durée et le prix de revient des opérations de perforation.

a) *Enregistrement indirect*, puisque l'information est généralement regroupée par « individu » et qu'il faut l'ordonner selon l'autre dimension, celle des caractéristiques, pour faciliter la perforation. Cela suppose, entre le questionnaire ou la grille d'analyse codée et la fiche perforée, la phase intermédiaire de la feuille de dépouillement (*). On peut cependant concevoir des dispositifs pratiques d'étalement des fiches caractéristiques pour supprimer cette phrase intermédiaire.

b) *Perforation incommode* et longue avec les modèles de fiches et les matériels fabriqués pour ce système (**).

C'est pourquoi on peut préférer actuellement à ces modèles de fiches et de matériels les modèles de cartes et les matériels mécanographiques BULL, I.B.M. - I.C.T., etc. qu'il est possible d'adapter à ce genre de perforation.

c) *Contrôle de perforation* par superpositions relativement rapides et faciles à exécuter des fiches correspondant à chaque item, à condition que soient seulement enregistrées des caractéristiques incompatibles entre elles par item.

(*) A noter, entre autres, les procédés de photolecture, magnétolecture ou graphitolecture, les convertisseurs automatiques bande perforée-carte perforée, les vérificatrices et les reproductrices, etc., qui réduisent considérablement la durée de ces opérations et les risques d'erreurs, mais restent d'un emploi limité dans les enquêtes intensives à questions non précodées.

(*) Appellée « feuille de préperforation Memo 100 » chez « Selecto ».

Notons que cette phase est souvent nécessaire dans les analyses de contenu de questions ouvertes non précodées, sur petits échantillons.

D'autre part, dans le cas d'enregistrement discontinu (par exemple par ordre chronologique d'arrivée de documents bibliographiques) il n'y a pas d'inconvénients à passer directement de l'analyse codée à la perforation document par document.

(**) A l'exception d'un perforateur semi-automatique à mémoire électronique, le « Per Selecto PS 5 », fonctionnant pour 100 « individus » à la fois, caractéristique par caractéristique et d'un projet de lecteur-perforateur automatique pour feuilles de préperforation « Memo 100 ». A noter également, dans la gamme

Système à entrée de l'information
par individu

Système à entrée de l'information
par caractéristique

du matériel « Termatrex », un convertisseur automatique perforant les fiches par caractéristique à partir de données perforées sur cartes mécanographiques par individu.

6. *Souplesse d'utilisation et commodité pour la lecture*
(tris simples et combinés) (1).

a) *Lecture indirecte des tris simples*, nécessitant des appareils adaptés aux différents modèles de fiches ou de cartes, par exemple :

- simple aiguille ou broche de tri ou matrice perforée avec plusieurs aiguilles (trieur-sélecteur avec ou sans vibreur) pour fiches à perforations marginales ou centrales
- trieuse électromécanique pour cartes perforées
- ensemble électronique pour mémoires magnétiques.

Le fichier est donc inutilisable sans l'aide de ces appareils qui, dans les systèmes mécanographiques, exigent des opérateurs, voire des programmeurs très qualifiés.

b) *Lecture indirecte mais commode des opérations ET, OU et SAUF* par tris successifs ou simultanés (*), rendue encore plus commode avec les ordinateurs à programme enregistré et à « langage évolué » : voir par exemple les programmations automatiques pour dépouillements d'enquêtes (analyses factorielles, hiérarchiques, typologiques, structurales, etc.) dans les grands centres de calcul électronique (**).

a) *Lecture directe et instantanée* des tris simples, par extraction manuelle hors de leur fichier des fiches caractéristiques recherchées.

Le fichier est donc utilisable en tous lieux et en toutes circonstances, sans l'aide d'aucune machine si ce n'est, pour la commodité de la lecture, d'une « visionneuse » (*).

b) *Lecture directe et instantanée de l'opération ET* par simple superposition des fiches caractéristiques correspondantes extraites du fichier, pratiquement sans limitation du nombre n de fiches à superposer (**).

Lecture incommode (nécessitant des comptages distincts) *des autres opérations OU et SAUF* qui ont avantage à être traitées comme de nouvelles caractéristiques, c'est-à-dire à être perforées sur de nouvelles fiches. D'où nécessité de prévoir un matériel de reproduction adéquat pour enregistrer et perforer les données résultant de ces opérations (**).

La commodité de lecture des opérations ET, les plus fréquents pour l'établissement de tableaux de contingence ou pour la recherche documentaire par exemple, est un des avantages les plus appréciés des utilisateurs de ce système qui y voient la possibilité

(1) Par tris combinés nous entendons les opérations logiques élémentaires : ET, relation d'« intersection » ou « produit » algébrique (symbole $\cap$) ; OU, relation de « réunion » ou « somme » algébrique (symbole $\cup$) et SAUF, relation de « complémentarité » ou « différence » algébrique (symbole $-$).

Les fiches perforées « par caractéristique »

Système à entrée de l'information par individu

(*) A noter que le dispositif de sélection TEXOLER (adaptation d'un « explorateur » à 42 compteurs sur trieuse SAMAS-I.C.T. à plongeurs pour cartes standard à 65 ou à 80 colonnes) est plus rapide et d'un emploi beaucoup plus commode que les autres trieuses mécanographiques à balais de lecture (du type I.B.M. 101 ou BULL D 3 par exemple). Son emploi paraît particulièrement recommandé lorsque le nombre de caractéristiques ne dépasse pas 960 (cf. § 2a).

(**) Centre national de calcul électronique chez Bull, Institut européen de calcul scientifique chez I.B.M., Département scientifique de la SOGREAH (à Grenoble), Centre de calcul du C.F.R.O., Institut Blaise-Pascal du C.N.R.S., etc.

Système à entrée de l'information par caractéristique

de rester en contact étroit avec les données de l'information.

D'autre part, si l'on a pris soin d'enregistrer les questionnaires ou documents selon un principe de numérotation tenant compte des principales variables de stratification de l'échantillon, la lecture des fiches-caractéristiques sous une feuille translucide (sur laquelle des traits matérialisent le dessin de ces stratifications) fournit instantanément un tableau impressionniste de la ventilation des caractéristiques correspondantes selon ces différentes variables. Un tel procédé est particulièrement commode pour les plans d'échantillonnage comparatif stratifié, dits « plans factoriels orthogonaux ».

(*) Simple glace dépolie posée devant une source lumineuse appelée « lecteur » chez Sphinxo, « Sta Selecto » chez Selecto, « visionneuse » chez Dequeker. Pour l'identification des perforations (ou repérage des numéros perforés) la lisibilité est fonction inverse de la densité des perforations : elle est la meilleure avec les fiches Sphinxo et la moins bonne avec les fiches Selecto ; elle est suffisante avec les fiches Dequeker, ainsi qu'avec les cartes mécanographiques de format standard Bull, I.B.M., Samas-I.C.T., etc. dans lesquelles le numéro de colonne peut indiquer la dizaine et le numéro de ligne l'unité (cf. § 1a).

(**) A noter que Selecto fabrique une visionneuse progressive dite « Sta lumineux à proximité », qui permet de lire également les combinaisons ET à 1 ou 2 caractéristiques près ($n-1$ et $n-2$), ce qui peut être utile pour les analyses typologiques de *clusters* et les recherches documentaires par approximation, entre autres.

(***) Par exemple la reproductrice Selecto 5.000, ou la reproductrice de bureau pour cartes 80/160 colonnes modèle Samas-I.C.T. type 19/0. A noter cependant que la lecture de l'opération SAUF (ou « sélection-négative ») est possible avec les modèles de fiches à grosses perforations (cf. § 6 a) grâce à une feuille translucide colorée intercalée entre les fiches sur la visionneuse et qu'elle est prévue dans la gamme du matériel Termatex, grâce à un dispositif de reproduction photographique de fiches négatives.

Système à entrée de l'information
par individu

Système à entrée de l'information
par caractéristique

7. *Souplesse d'utilisation et commodité pour le dénombrement
et les traitements mathématiques (et l'affichage des résultats)*

a) *Dénombrement* incommode (long et fastidieux) dans les systèmes manuels et mécaniques (comptage manuel des fiches pertinentes) ou approximatif par mesure de poids ou d'épaisseur des paquets de fiches. Dénombrement automatique et très rapide dans les systèmes mécanographiques qui comportent le plus souvent des compteurs indépendants des opérations de tri.

b) *Traitements mathématiques et statistiques* : seuls, parmi les procédés de ce système, les ensembles électroniques remplissent simultanément les fonctions de trieuses et de calculateurs. Pourcentages, pondérations, indices de valeur centrale et de dispersion, analyses factorielles, de variance, typologiques, épreuves de signification, etc. peuvent apparaître en même temps que les chiffres bruts, si les programmes correspondants existent déjà dans les « bibliothèques à programmes ». C'est ce qui en rend l'usage aussi apprécié, surtout lorsque les résultats doivent être connus dans les délais les plus brefs.

c) *Affichage des résultats*, soit à l'aide de compteurs (à relevé manuel ou photographique) sur les trieuses mécanographiques ; soit sous forme de tableaux de chiffres imprimés, à l'aide de tabulatrices ; soit même directement sous forme de courbes et de graphiques, à l'aide des oscilloscopes cathodiques à balayage numérique (*).

(*) Tel que celui qui est connecté à l'ordinateur I.B.M. 704 par exemple.

a) *Dénombrement* incommode (long et fastidieux) en l'absence de matériel approprié. Jusqu'à ce jour, on procédait par dénombrement visuel des perforations, placées devant une source lumineuse, ou bien par pression manuelle à l'aide d'un stylet télescopique à contact électrique (« pointeur-compteur » TEL). On dispose depuis peu d'un appareil de comptage automatique, sûr et rapide, portable et relativement peu coûteux (« photo-compteur électronique » CETEL) qui permet de compter par exemple les perforations d'une carte SAMAS-ICT à 1.920 positions en moins de 3 secondes, avec affichage électronique (*). Le système à fiches caractéristiques, jusque là peu recommandé dans les travaux comportant des dénombrements importants, est désormais efficient et à la portée de toute équipe de recherche de dimension moyenne.

b) *Traitements mathématiques et statistiques* sont ici, comme pour les systèmes manuels, mécaniques et électromécaniques à entrée de l'information par individu, dissociés des opérations de dénombrement. Ils doivent donc être faits séparément, soit sur machines à calculer de bureau, soit sur calculateurs électroniques, soit plus simplement à l'aide de tables de statistiques ou d'abaques dont l'usage nous semble pouvoir s'étendre considérablement.

c) *Affichage des résultats* associé au comptage. Tabulation imprimée des numéros matricules d'« individus » ou de « documents » pertinents à l'étude chez Selecto et réalisable chez Termatex, bien que la photographie des fiches-caractéristiques exposées devant une source lumineuse puisse en tenir lieu.

(*) cf. également l'« Automatic Card Scanner » J 500, compteur et tabulateur automatique pour fiches Termatex.

Les fiches perforées « par caractéristique »

Système à entrée de l'information
par individu

Système à entrée de l'information
par caractéristique

C. — AUTRES CONSIDÉRATIONS

8. Autres considérations pratiques concernant :

a) *L'établissement du stock d'information.* — Coût élevé des fiches à pré-perforations ou à encoches et de leur perforation, dans les systèmes manuels et mécaniques mais coût réduit des cartes mécanographiques et de leur perforation, en raison de la standardisation poussée de ce système. De plus, il existe toute une gamme de métiers spécialisés et d'entreprises commerciales pratiquant soit le travail à façon soit la location de machines.

b) *L'exploitation du stock d'information.* — L'exploitation d'un fichier de fiches à préperforations exige très peu de crédits et très peu de matériel mais beaucoup de temps et comporte des risques d'erreur. En revanche, l'exploitation d'un fichier de cartes mécanographiques exige des crédits importants soit pour achat ou location de machines, soit pour prestation de services (*). Et dans ce dernier cas, le chercheur ou le documentaliste est lié à une programmation rigide qui exclut pratiquement toute possibilité d'intervention permanente et progressive au fur et à mesure de la fourniture des résultats. Il faut noter également le caractère « éphémère » de l'information introduite en mémoire magnétique, que des raisons financières obligent souvent à traiter dans de brefs délais pour ne pas immobiliser trop longtemps un support matériel coûteux, quitte à réintroduire plusieurs fois en mémoire l'information enregistrée initialement sur cartes perforées.

a) *L'établissement du stock d'information.* — Coût élevé des fiches caractéristiques vendues comme telles (SPHINXO, DEQUEKER et SE-LECTO) et de leur perforation, mais ce coût peut être sensiblement réduit si l'on utilise les cartes mécanographiques et poinçonneuses standard (cf. § 1a et 5b).

A noter que, si le nombre de perforations est théoriquement le même dans les deux systèmes, par individu ou par caractéristique, dans certains cas (informations ou documentations extensives) le nombre de fiches-caractéristiques peut néanmoins être très inférieur au nombre de cartes-individus, d'où encombrement et manipulation moindres (*).

b) *L'exploitation du stock d'information.* — Un des principaux avantages du système à fiches-caractéristiques est de fonctionner avec le minimum de crédits et de matériel et avec le maximum de souplesse. Mais il faut reconnaître qu'il exige beaucoup de temps à l'exploitation; des progrès technologiques sont encore nécessaires pour passer à un stade d'automatisme qui rende ce système compétitif avec les systèmes mécanographiques et électroniques (par cartes-individus) dans une gamme plus étendue de problèmes de recherche et de documentation. Nous en parlerons plus en détail à la fin de cet article.

(*) A titre d'exemple : la fourniture de tableaux de contingence à double entrée avec indication de tous les pourcen-

(*) Nous entendons par informations extensives celles dans lesquelles le nombre de « caractéristiques » est inférieur

Système à entrée de l'information
par individu

tages, des fréquences théoriques, et du Khi carré global, par le calculateur GAMMA 60 du C.N.C.E. de la C¹^e BULL est facturée à plus de 5 centimes par tableau et par individu (soit par exemple 5.000 francs pour 1.300 tableaux de contingence à deux dimensions d'un échantillon de 70 sujets ou pour 70 tableaux d'un échantillon de 1.300 sujets — en 1963).

Le coût d'utilisation des moyens ensembles et grands ensembles électromagnétiques à traiter l'information (I.B.M. 1.401, I.C.T. 1.301, C.A.B. 300 et 3.900 de la S.E.A., Gamma 30 et 60 de BULL, I.B.M. 704 et 7.090, etc.) varie autour de 50.000 F par mois.

Système à entrée de l'information
par caractéristique

au nombre d'« individus ». Par opposition, les informations intensives sont celles dans lesquelles le nombre de « caractéristiques » est supérieur au nombre d'« individus ». Dans les cas extrêmes, le rapport Nombre de caractéristiques/Nombre d'individus peut donc avoir une importance pratique pour le choix de l'un ou de l'autre système, en liaison avec les facteurs de capacité ou de maléabilité étudiés aux paragraphes 1 à 4.

L'adoption d'un système et d'une technique de traitement de l'information dépend donc des caractéristiques particulières du ou des problèmes à traiter. Il ne saurait donc être question de recommander telle ou telle technique comme panacée, car aucun procédé ne réunit à la fois tous les avantages : le choix résulte le plus souvent d'un compromis dans lequel les différents critères, examinés ci-dessus, doivent être appréciés à leur juste poids.

Les solutions les plus satisfaisantes nous semblent néanmoins devoir être recherchées dans la combinaison des deux grands systèmes afin que soient préservés au maximum leurs avantages respectifs. A cet égard les procédés et matériels Jonker, qui réalisent cette combinaison du *conventional data processing* et de l'*inverted data processing* en un *total data processing*, grâce à un convertisseur automatique de cartes par individu en fiches par caractéristique, ouvrent des possibilités techniques intéressantes (11). Les ordinateurs les plus récents, associés à des organes périphériques appropriés, devraient être en mesure de procéder à la même conversion en utilisant à l'entrée et à la sortie les mêmes modèles de cartes mécanographiques standard à 80 colonnes, comme en témoignent les projets de programme à l'étude ou en cours chez I.B.M. ou à la SOGREAH par exemple.

Comme il nous a semblé que la méthode des fiches par caractéristique était soit méconnue des chercheurs, soit dévalorisée par l'utilisation artisanale qui en est faite habituellement en documentation, nous allons maintenant en présenter une application concrète et récente qui fera mieux comprendre les modalités pratiques du système en même temps qu'elle en illustrera les possibilités et les perfectionnements possibles. Nous décrirons surtout les techniques de manipulation directe qui sont les seules à nécessiter des explications et qui nous paraissent revêtir un intérêt tout particulier pour les chercheurs isolés ou lorsque les données à enregistrer imposent certaines conditions. Cela n'exclut pas pour autant la possibilité de traitement de ces cartes par ordinateurs, dont nous parlerons ensuite.

L'étude dont il sera question ici est une enquête psycho-sociologique de

(11) JONKER, F. « The new 'Termatex' and 'Minimatex' line of information retrieval system (also available for data retrieval correlation purposes) » *American Documentation*/14 (4), October 1963.

Les fiches perforées « par caractéristique »

grande étendue sur la jeunesse scolarisée au Togo (12). La taille de la population à sonder, à l'échelle d'une nation, le désir de préciser le rôle de certaines variables « de situation » telles que l'ethnie, la profession du père, etc., commandaient de prendre un échantillon important.

Le nombre de variables à étudier dans une population très hétérogène, la complexité des problèmes à aborder dans un milieu « ethnographique » très varié et dont on ignorait de nombreuses données de base, nécessitaient une investigation suffisamment approfondie. Au cours d'une enquête d'approche de près d'un an un questionnaire « en profondeur » comportant 125 questions principales — dont un grand nombre de questions « ouvertes » — a été établi. Son remplissage par écrit durait en moyenne trois heures et quart. Un système de passation collective dans les établissements scolaires, animé par un enquêteur spécialement formé, a permis de faire remplir 3.309 questionnaires en un peu moins de cinq semaines. L'information obtenue — au total plus de 46.000 feuilles de questionnaires — était suffisamment étendue et nuancée pour justifier d'un système d'enregistrement d'une capacité assez grande pour contenir l'échantillon, mais aussi et surtout, offrant la plus grande malléabilité au regard des caractéristiques qu'il s'agissait de définir.

A la première analyse de contenu, 2.446 caractéristiques ont été déterminées sans dépasser le niveau des données manifestes, ce qui excédait déjà, avec le procédé d'enregistrement par individu, la capacité (960 positions au maximum) des cartes mécanographiques de modèle classique, employées par l'Institut de Statistiques local qui devait effectuer la perforation. Avec ce système, en outre, l'étalement des données sur plusieurs cartes se heurtait à une impossibilité au niveau de l'exploitation du fait de l'obligation de recourir à un ordinateur ou à une interclasseuse qui n'existaient pas au Togo.

En outre, l'analyse psycho-sociologique d'un matériel, volontairement peu structuré au niveau de la collecte, demandait des élaborations successives entraînant de nombreuses retouches et additions de caractéristiques. Elle rendait par là particulièrement inadéquate l'utilisation d'un procédé d'enregistrement mécanographique astreint à un dessin fixe de carte.

Toutes les limitations du procédé par cartes individuelles, énoncées ci-dessus de façon théorique, apparaissaient donc ici d'une manière particulièrement frappante. Aussi, après étude de la question avec l'Institut de Statistiques du Togo, et en dépit du fait que celui-ci possédait les machines permettant l'exploitation du procédé de perforation classique, on a finalement décidé d'adopter le procédé décrit ici des cartes par caractéristique.

La première opération du dépouillement a été le numérotage des questionnaires au moyen d'un tampon humide. Chaque questionnaire comportant plusieurs pages, toutes les pages ont dû être numérotées, chacune portant évidemment le même numéro. Ce n'est qu'à ce moment qu'on a pu les séparer pour la commodité du codage. Le numéro, à trois chiffres, était en deux parties : les deux premiers chiffres correspondaient au numéro de colonne de la carte (à partir de 01), le troisième au numéro de la ligne, ce qui aboutissait pratiquement à commencer le numérotage à 010 au lieu 001. Afin de conserver un ordre décimal, on a retenu seulement dix lignes, soit 800 positions correspondant à 800 individus enregistrables, sur l'ensemble de la carte (13).

(12) Cf. DEVAUGES, R. « La jeunesse instruite au Togo. Etude psycho-sociologique » (en préparation).

(13) Avec le procédé SAMAS-I.C.T. permettant de perforer par décalage sur 20 lignes au lieu de 10 on aurait pu doubler la capacité et la porter à 1.600 individus, mais on sortait ainsi des possibilités du matériel disponible au Togo.

L'établissement du premier plan de codification, en raison des nombreuses questions ouvertes, a duré environ six mois. Mais, aussitôt que les premiers codes ont été prêts, c'est-à-dire très rapidement, le chiffrement et la perforation des questions correspondantes ont été entrepris. Et ainsi de suite, jusqu'à la fin, à mesure que l'analyse d'une question était terminée. On notera ici le gain de temps réalisé sur l'ensemble de ces opérations par comparaison avec le procédé par individu qui aurait nécessité d'attendre que toute l'analyse soit achevée pour arrêter le dessin de carte et passer à la perforation.

Le codage a été effectué de la façon habituelle, dans les blocs de chiffrement des questionnaires. Pour les réponses multiples ou lorsqu'on analysait une réponse selon plusieurs « facettes » — ce qui était généralement le cas pour les questions ouvertes — on a conservé comme règle de n'enregistrer qu'une réponse par item et de prendre alors autant d'items que de réponses retenues (le plus souvent deux ou trois). Celui qui n'avait fourni qu'une réponse figurait alors dans les items suivants sous la rubrique « pas de réponse ». Cette méthode, bien qu'un peu plus longue à la perforation, présente plusieurs avantages : elle fournit une ordination des types de réponses selon leur fréquence en première, deuxième ou troisième série. Elle permet en outre d'enregistrer une « redondance » dans le cas où le sujet donne deux ou plusieurs réponses de la même catégorie. En outre, elle préserve la possibilité de contrôler la perforation des fiches par superposition, ce qui n'est plus possible si l'on enregistre plusieurs réponses dans un même item (cf. *infra*).

Le codage terminé, au lieu de donner directement les questionnaires à l'atelier mécanographique, il est apparu beaucoup plus commode de transférer les données chiffrées sur des fiches intermédiaires de chiffrement (14).

En même temps que les fiches intermédiaires de chiffrement, on préparait les cartes à perforer. Cela consistait à faire figurer sur la marge supérieure les indications permettant de les sélectionner pour l'exploitation : numéro de la question, numéro de l'item. On aurait pu aussi — ce qui eût été d'une grande commodité pour l'exploitation — ajouter, au moins de façon abrégée, le contenu de l'item. La taille de l'échantillon dépassant la capacité d'une carte, il a fallu la répartir sur plusieurs ; on a donc fait figurer sur chaque carte le numéro de la série à laquelle elle appartenait, celle-ci étant signalée en outre par une couleur différente de la carte. On a d'ailleurs profité de cette obligation d'étaler l'échantillon sur plusieurs cartes pour faire, au moyen des séries, une stratification par niveau d'instruction et par zone géographique.

La perforation a été confiée à l'Institut de Statistiques du Togo (15).

Une technique complète de contrôle de la perforation applicable lorsque toutes les cartes relatives à une question étaient perforées, avait été mise au point. Elle permettait, par une série de manipulations assez rapides et faciles à exécuter, de déceler la quasi-totalité des erreurs dues à la perforation (16). Ce contrôle se référait à deux données de base : le nombre de réponses à

(14) On utilisait une fiche par question : cette fiche était divisée en colonnes ; chaque colonne était consacrée à un item et on y reportait, dans l'ordre numérique, le numéro des sujets ayant répondu conformément à cet item.

(15) La perforation de plusieurs positions par colonne, voire de la colonne tout entière, présentait des difficultés avec les machines dont disposait l'Institut de Statistiques local ; il eût fallu enlever le système de sécurité assurant l'avancement automatique de la carte après chaque perforation, ce à quoi les techniciens se sont refusés. On s'est alors résolu à effectuer toute la perforation sur une machine à rectifier, du modèle BULL P 80. Le travail a été fait caractéristique par caractéristique, c'est-à-dire carte par carte.

(16) En effet, pour échapper à ce contrôle, il faut une permutation des deux mêmes numéros se reproduisant d'une carte sur une autre, ce qui est hautement improbable.

Les fiches perforées « par caractéristique »

chaque item, calculé sur la fiche de transfert, et la « plage » de perforations, correspondant au total de l'échantillon. Une seule réponse étant par principe retenue dans chaque item, chaque questionnaire ne devait être « cité » qu'une fois et la superposition par paires de toutes les cartes ne devait laisser apparaître aucun « trou ». Le contrôle consistait donc à mirer deux par deux, selon toutes les combinaisons possibles, toutes les cartes-réponses de chaque question. Si un trou apparaissait, il était le signe d'une double perforation. Au lieu de refaire toute la carte mal renseignée, il suffisait alors de boucher la perforation fautive.

Le second contrôle consistait à compter le nombre de perforations de chaque carte. En comparant ce nombre obtenu à celui figurant sur la fiche intermédiaire pour l'item correspondant, on décelait alors les perforations oubliées (les perforations « en trop » ayant été éliminées lors du contrôle précédent).

La dernière vérification, enfin, consistait à rapporter chaque carte à la « plage » de perforations correspondant à l'échantillon pour s'assurer qu'aucune perforation n'avait été faite en dehors de cette « plage ». A la différence des deux précédents, ce type d'erreur s'est révélé extrêmement rare.

Au total, dans des conditions assez précaires et avec un personnel formé sur place, l'ensemble des opérations de dépouillement, depuis la fin de l'enquête sur le terrain jusqu'à la perforation inclusivement, s'est étendu sur une période de onze mois. Deux cent quatre séries de réponses (en incluant les réponses mutiples), réparties le plus souvent sur dix séries (sur deux seulement pour une vingtaine de questions réservées à l'enseignement secondaire), ont été enregistrées sur une dizaine de milliers de cartes. Cette première étape achevée et les premières conclusions dégagées, au fur et à mesure du progrès de la recherche, de nouvelles caractéristiques sont élaborées, d'anciennes remaniées, des types et des indices définis, donnant lieu à la perforation de nouvelles cartes pouvant à leur tour être mises en corrélation avec l'ensemble du matériel déjà perforé. La souplesse du procédé permet ainsi de suivre pas à pas toutes les étapes de l'analyse, puis de l'élaboration synthétique des données.

*
**

Dans l'état des moyens techniques disponibles jusqu'à une époque toute récente, l'exploitation des cartes par caractéristique pour la confection des tableaux ne pouvait se faire qu'« à la main » en utilisant une visionneuse. Celle-ci est une simple boîte à lumière, avec un verre dépoli éclairé par en dessous. Une équerre permet tout en les superposant de centrer exactement les cartes sur la source lumineuse d'une façon rigoureuse. Pour établir un tableau de contingence entre deux questions *a* et *b*, il suffit de poser la première carte de la question *a* sur la visionneuse, puis de faire défiler successivement dessus toutes les cartes de la question *b* en comptant à chaque superposition le nombre de perforations communes et en l'inscrivant dans la case correspondante du tableau à établir (17). Les cartes de la question *b* épuisées, on

(17) L'opération de comptage des perforations est souvent délicate et provoque de nombreuses erreurs. Pour la rendre plus facile, on a trouvé pratique d'intercaler entre le verre dépoli et la première carte une feuille de papier calque coupée à la dimension des cartes; l'employé travaillant à la visionneuse marque un trait au crayon sur la feuille à travers chaque perforation en même temps qu'il la compte; toutes les cases du tableau remplies, on effectue alors les totalisations partielles en bout de ligne et en bout de colonne. Les chiffres trouvés doivent correspondre à ceux du nombre de réponses aux différents items pour chacune des deux questions.

prend la seconde carte de la question *a* comme carte de fond et on recommence à faire défiler dessus toutes les cartes de la question *b*. On opère ainsi jusqu'à épuisement de toutes les cartes correspondant aux deux questions *a* et *b*.

En dépit des avantages indéniables qui nous sont apparus au cours de l'application qui vient d'être décrite, le procédé des cartes par caractéristique présente dans son état actuel un handicap certain du fait qu'il ne permet le plus souvent que l'exploitation manuelle (18) alors que le procédé par cartes par individu dispose de tout un arsenal de machines électro-mécaniques, électro-magnétiques et électroniques spécialement établies à son intention. Or, le comptage « manuel » lorsqu'il porte sur un nombre considérable de perforations réparties sur une grande quantité de cartes est rapidement fastidieux et sujet à erreurs. Le nombre de cartes à manipuler s'accroît rapidement quand on veut comparer plus de deux questions du fait qu'à chaque carte d'une question il faut comparer la totalité des cartes de l'autre. Pour réduire ce goulot d'étranglement de la méthode, on a donc cherché à mettre au point des procédés « armés » susceptibles d'accélérer le travail.

L'une des directions dans lesquelles ont été orientées les recherches est celle des ordinateurs. Là encore, presque tous les appareils existants ne « lisent » que les cartes par individu. Il existe toutefois depuis quelques années un modèle capable de traiter directement les fiches par caractéristique, c'est l'ordinateur portant la dénomination commerciale « I.B.M. 1401 » à lecture binaire. A partir de cet appareil travaillant en association avec un ordinateur 7094 à grande puissance, chargé d'effectuer les calculs, il a été possible d'établir, avec des crédits de l'O.R.S.T.O.M., un programme permettant de croiser simultanément les données de plusieurs cartes — correspondant ici à plusieurs questions — et d'obtenir ainsi en un seul passage des séries de tableaux tenant compte de deux, trois ou quatre critères. Dans l'étude déjà citée sur le Togo, on a utilisé cette possibilité pour effectuer les croisements de questions en donnant séparément les résultats pour chacun des quatre niveaux d'enseignement (CM2, 5^e, 3^e, classes du Baccalauréat) et, dans chacune de ces strates, les résultats pour les garçons, pour les filles et pour l'ensemble (en « redressant » pour ce dernier résultat les taux de sondage différents adoptés pour chaque sexe). Pour chacune de ces données l'ordinateur fournit alors les fréquences observées, les pourcentages en ligne et en colonne, les fréquences « attendues » et, en bas de tableau, la valeur de « Khi-deux » pour chacune des trois séries (avec introduction automatique de la correction de Yates pour toutes les cases fournissant une valeur attendue inférieure à 10). Ceci n'est évidemment qu'une première approche des immenses possibilités de ce type d'appareil. Des programmes d'analyse hiérarchique, des programmes utilisant divers coefficients de liaison (du type du « r » de Bravais-Pearson ou au contraire de type non-paramétrique pour les données simplement ordinales) existent déjà ou seront mis au point prochainement dans les divers centres de calcul électronique cités plus haut (cf. 6 b et 7 b).

Nous ne nous étendrons pas davantage sur les possibilités offertes par les ordinateurs, puisque ce genre de traitement mécanographique de l'information

(18) Tout au moins dans l'état du matériel actuellement commercialisé en France, car nous avons signalé plus haut (§ 5 a et b, 6 b, 7 a et c) l'existence de projets de machines automatiques chez SELECTO (pour perforation, reproduction, lecture et tabulation), l'existence aux Etats-Unis d'une gamme complète de machines automatiques pouvant travailler dans les deux systèmes (*Total data processing* du matériel TERMATREX de la C¹⁰ JONKER) et en France d'un prototype de photo-comptense électronique, CETEL, livré au Groupe d'Ethnologie sociale, dont nous reparlerons plus loin.

Les fiches perforées « par caractéristique »

n'est pas une propriété spécifique du système « par caractéristique » qui constitue l'objet de cet article, et qu'il est déjà suffisamment connu des utilisateurs du système « par individu ». Il suffit de retenir que le mode d'exploitation des données par caractéristique n'exclut pas l'utilisation des ensembles de calcul électronique les plus perfectionnés, soit que les données aient été directement enregistrées selon ce système et lues telles quelles par certaines machines (comme nous venons de le décrire), soit qu'elles aient été préalablement enregistrées selon le système « par individu » (ce qui représente des avantages au stade de l'enregistrement et de la perfection — § 5 a cf. p. 423, et ce qui permet d'utiliser une gamme plus étendue de matériels), avant d'être converties dans l'autre système « par caractéristique » aux fins d'analyses non programmées, effectuées progressivement et directement par le chercheur lui-même.

C'est pour améliorer le rendement de ces analyses sur cartes par caractéristique qu'un constructeur de matériel électronique vient de mettre au point, sur crédits du Centre national de la Recherche scientifique accordés au Groupe d'Ethnologie sociale, une « photo-compteuse électronique » permettant le comptage et l'affichage quasi instantanés des perforations d'une ou plusieurs cartes superposées de format standard (187 mm × 83 mm), quel qu'en soit le type de perforations (rondes ou rectangulaires) et quelle qu'en soit la capacité (65, 80, 130, 160 colonnes). Cette machine, qui représente un progrès considérable sur le comptage direct, ne constitue toutefois qu'une première étape dans la semi-automatisation de la manipulation directe. L'expérience de celle-ci sur de grands ensembles de données permet d'ores et déjà de dresser une sorte de « cahier des charges » des possibilités que l'on attendrait de l'équipement idéal. Celui-ci devrait effectuer deux types d'opérations importantes de façon automatique tout en réservant la possibilité d'une intervention humaine à chaque stade.

1° Comptage et affichage automatique des perforations

C'est précisément ce que fait cette photo-compteuse électronique. Un dispositif supplémentaire devrait toutefois permettre d'agir sur les modalités de ce comptage et de disposer à volonté des possibilités suivantes :

a) Totalisation séparée carte par carte et/ou cumulative par série de cartes (pour les cas où le nombre d'individus excède la capacité d'une carte et où chaque carte matérialise une variable importante).

b) Totalisation séparée et/ou cumulative par ligne ou par série de lignes et/ou par colonne ou par série de colonnes (pour les cas où le dessin de la carte en lignes ou en colonnes exprime une variable importante).

2° Reproduction automatique des cartes

Comme on l'a déjà souligné, l'intérêt de l'emploi des fiches par caractéristique est de permettre au niveau de la codification une multiplication commode des items adaptés à la finesse désirée du niveau de l'analyse. En revanche, au moment de dresser les tableaux ou pour les besoins d'une analyse hiérarchique, il est généralement nécessaire de regrouper deux ou plusieurs catégories (« somme » algébrique), voire même de dichotomiser les réponses à une question. Pour simplifier les manipulations, il faut pouvoir « recopier » automatiquement sur une carte vierge les perforations d'une ou plusieurs cartes déjà renseignées. La plupart des machines à perforer clas-

siques qui effectuent ce travail avec les cartes par individu ne peuvent le faire avec les cartes par caractéristique en raison du nombre trop grand de perforations par colonne. Il faudrait donc prévoir la mise au point, soit sous forme d'une machine spéciale, soit par l'adaptation d'un équipement supplémentaire au lecteur automatique décrit ici, soit encore par l'adaptation à cette fin d'une reproductrice classique, d'un dispositif de perforation permettant de reproduire automatiquement les cartes des caractéristiques à regrouper (cf. la reproductrice de bureau SAMAS-I.C.T., n° 19/0, décrite p. 425).

L'ensemble de ces dispositifs, machine à perforer, machine à reproduire et compteur automatique de perforations, représente alors un équipement mécanographique complet d'un prix relativement peu élevé (14.300 F) et surtout facile à manier et de ce fait utilisable directement par les chercheurs. Doublé de la possibilité de recourir à un ordinateur pour les programmes suffisamment élaborés, il permettrait alors, dans les meilleures conditions d'économie et d'efficacité, d'utiliser au mieux les possibilités des cartes par caractéristique et de réaliser cette association directe et constante de la mécanographie à toutes les étapes de l'analyse sociologique qui nous a paru être l'avantage déterminant de ce procédé d'enregistrement.

R. DEVAUGES

Office de la Recherche scientifique
et technique Outre-Mer.

J. JENNY *

Centre national
de la Recherche scientifique.

BIBLIOGRAPHIE

- FAVIER, J. THOMELIN, R. *La mécanographie*. Edition de Montligeon, 1960, 208 p.
- PEPE, P. *Procédés et matériels de dépouillements statistiques*. Paris, Dunod, 1962, 149 p.
- CASEY, R. S., PERRY, J. W., KENT, A., BERRY, M. *Punched cards, their applications to science and industry*. 2nd edition. New York, Reinhold, Publishing Corporation, 1958 (en particulier le chapitre VI : « The Peek-a-boo system : optical coincidence subject cards »).
- GAUCHET, F., LAMBERT, R., VIOLET, J. *Initiation au traitement automatique de l'information dans les sciences humaines*. Rapport ronéotypé édité par l'A.F.A.P. 1963 (à paraître aux Presses Universitaires de France). Avec bibliographie annotée.
- VICKERY, B. C. *Techniques modernes de documentation - Analyse des systèmes de recherche de documents*. Paris, Dunod, 1962, 178 p. (traduit de de l'anglais « On retrieval system theory » par la S.E.M.A.).
- C.I.M.A.B. « Les systèmes de sélection utilisables dans les organismes de documentation ». Supplément au bulletin n° 8 de septembre 1962, 56 p.
- C.I.M.A.B. « Appareils et fiches perforées pour sélection manuelle ». Notice technique 1953, 32 p.
- Documentation concernant les matériels : BULL, I.B.M. I.C.T., TEXOLER, TEXO-LETTE SELECTO, SELECTRI, DETECTRI, RAPIDITRI, SATO, DEQUEKER, SPHINNO, TERMATREX, MINIMATREX, etc., fournie par les fabricants.
- Catalogues du Salon International de l'Équipement de Bureau des années 1961 à 1963.
- Divers articles dans des revues professionnelles : *Electro-Calcul*, *Revue de la Mécanographie*, *Bulletin des Bibliothèques de France*, *Bulletin de l'U.N.E.S.C.O* à l'intention des bibliothèques, *C.N.O.F.* (*Revue mensuelle de l'Organisation*), etc.

(*) Les auteurs expriment leur reconnaissance à ceux de leurs collègues qui leur ont fait part de critiques ou d'informations utiles, en particulier aux chercheurs et techniciens du service de Documentation du Laboratoire de Psychologie sociale et à Odile Benoit-Guilbot, du Centre d'Études Sociologiques, à qui revient le mérite d'avoir introduit dès 1952 l'utilisation de cartes mécanographiques standard SAMAS-I.C.T. à perforations rondes comme fiches par caractéristique pour l'exploitation d'enquêtes intensives du Groupe d'Éthnologie sociale.

4

PL 089

revue française de sociologie

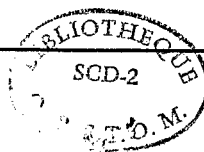
pas un
de Hergon

Les fiches perforées
“ par caractéristique ”
et leur application aux enquêtes
par questionnaire

Par Roland DEVAUGES et Jacques JENNY

O.R.S.T.O.M. Fonds Documentaire

14 SEPT. 1994



N° : 40182

Cote : B

M

P 52

Revue française de sociologie
Vol. V. N° 4. Octobre-Décembre 1964

SOMMAIRE

- François CHAZEL Réflexions sur la conception parsonienne
du pouvoir et de l'influence 387
- René DUCHAC Suicide au Japon, suicide à la japonaise 402
- Roland DEVAUGES
Jacques JENNY Les fiches perforées « par caractéristique »
et leur application aux enquêtes par ques-
tionnaire 416



- Claude GRUSON Planification économique et recherches socio-
logiques 435

ACTUALITÉ DE LA RECHERCHE

- J. L. COURCHET
P. H. MAUCORPS
J. P. PETARD Revendication et empathie 447