

Ce véritable goulot d'étranglement entre la multitude des informations et leur interprétation, certains croient le franchir en s'en tenant à l'examen détaillé d'un cas déjà célèbre ou controversé. On traite alors un exemple, nullement un échantillon, et les difficultés resurgissent aussitôt qu'il s'agit de prouver la représentativité des faits retenus. En d'autres termes, on en revient tôt ou tard aux enquêtes de longue haleine et aux risques qu'elles comportent. Si prometteuses qu'elles s'annoncent, méritent-elles des mois, voire des années de travail ?

Les techniques mécanographiques mises au point au L.A.S.L.A. aident à résoudre ce problème non seulement parce qu'elles abrègent le classement des informations, mais surtout, parce qu'en exigeant une plus rigoureuse homogénéité des données, elles facilitent des utilisations ultérieures, insoupçonnées lors de la phase initiale des hypothèses de départ.

Classement des informations d'abord.

L'application la plus évidente est la rédaction de tables et indices onomastiques dont il a été question ci-dessus. Pareils répertoires, indispensables pour l'utilisation commode d'ouvrages imprimés, le sont à fortiori pour l'exploitation des documents restés inédits. Tel est le cas, entre autres, des protocoles des notaires d'Ancien Régime. D'enquêtes entreprises à Gênes, à Paris et déjà — mais fragmentairement — à Liège même, il ressort que ces fonds d'archives constituent une mine pour l'histoire des industries, des milieux sociaux, du droit privé, de la vie quotidienne. A l'initiative du Professeur P. Harsin, une prospection systématique est en cours. Il s'agira de classer les milliers d'actes répertoriés. Des inventaires élaborés selon les normes de l'archivistique traditionnelle, n'y suffiraient jamais. Les cartes perforées permettent de consigner et de retrouver, outre les références et précisions de date et de lieu indispensables aux historiens :

- les noms des parties contractantes (permettant les recoupements individuels);
- la nature des actes : emprunts, contrats de travail, baux, testaments (première étape vers la constitution de séries homogènes).

En démographie rétrospective, la statistique des migrations est celle qui laisse le plus à désirer. A tel point qu'il s'impose souvent de recommencer les dénombrements à partir des documents de base : registres d'état civil, souches de passeports, etc... On risque alors d'être submergé par la multitude des cas d'espèce. Ici encore, la trieuse statistique du Laboratoire recense presque instantanément les fréquences, quelle que soit la complexité des mouvements migratoires. Bien plus, elle permet d'opérer l'indispensable regroupement des provenances d'après les fréquences effectivement observées et non pas d'après de quelconques circonscriptions administratives (arrondissements, provinces) dont la signification démographique est souvent nulle. De telles méthodes seront mises en œuvre pour mesurer l'immigration à Liège, à la veille de la révolution industrielle.

Enfin, c'est dans l'interprétation des séries chronologiques de prix et de salaires que l'appareillage du Laboratoire apparaît le plus indispensable à la besogne des historiens. Ces séries, en effet, sont constituées à partir de relevés multiples : les cours des céréales sont hebdomadaires ou mensuels dans le cas des mercuriales. Il s'impose de calculer d'abord des moyennes : entre prix maxima et minima, entre mois d'une même année civile et d'une même année-récolte, des moyennes annuelles mobiles aussi, de loin préférables à l'arbitraire découpage en tranches décennales. Le calcul des écarts-types, celui des droites de tendance et des corrélations permet de repérer puis de mesurer systématiquement les fluctuations des cours. Ici non plus, il n'y a pas de commune mesure entre des observations continues, envisageant un long terme, et les commentaires, fatalement myopes, de crises épidémiques. De telles recherches portent d'abord sur un grand marché-pilote aux Temps Modernes : Amsterdam. Dès maintenant, nous les étendons aux cas plus particuliers, mais pour nous plus intéressants, des villes liégeoises et luxembourgeoises.

Et. HÉLIN,

Chef de travaux à la Faculté de Philosophie et Lettres.

5. — L'Archéologie et l'Histoire de l'Art, disciplines réputées les moins rigoureuses de toutes les sciences humaines, peuvent fort bien tirer parti des techniques mécanographiques. Le Centre d'Analyse documentaire pour l'Archéologie de Paris le proclame avec insistance, preuves à l'appui. Il m'est donné d'en faire en ce moment au Laboratoire d'Analyse statistique des Langues anciennes la passionnante expérience.

Les recherches que je poursuis sur l'orfèvrerie religieuse liégeoise des Temps modernes m'ont permis de répertorier plus d'un millier de pièces. En faire l'étude, c'est d'abord et avant tout leur imposer autant de classements qu'il y a de critères à prendre en considération. Dénominations, dates, poinçons d'orfèvres et thèmes iconographiques sont, en l'occurrence, les principaux. Ces données, il n'a pas été bien difficile de les porter sur cartes perforées, en même temps qu'un numéro de renvoi; il a fallu très peu, on me permettra d'y insister, de ces conventions spéciales, de ces *encodages*, dont on est en droit de se faire un épouvantail. Par passages successifs des quelques 1400 cartes, les machines vont, avec l'infailibilité qui leur est propre, opérer les rangements requis, puis écrire les résultats obtenus. Bien mieux, elles vont rendre aisées toutes enquêtes sur les relations qui pourraient exister entre les différentes données inscrites dans les cartes. C'est ainsi que, par une simple *combinaison de tris*, il sera possible de déterminer dans quelle mesure il existe une *corrélation* entre les thèmes iconographiques d'une part, les textes, les époques ou les maîtres d'autre part.

Ces opérations sont en cours à l'heure où j'écris. L'essai poussé qui les a précédées et préparées (soit dit en passant, l'apprentissage qu'il m'a imposé ne fut pas sans profit, loin s'en faut) en garantit le succès.

L'exemple et l'à peu près, plaies de l'Archéologie et de l'Histoire de l'Art — fief longtemps incontesté des dilettantes — trouvent dans la mécanographie un ennemi mortel : n'est-elle pas à la fois école de rigueur et moyen de dominer d'amples collections de faits précis ? Cette technique paraît dès lors destinée à faire entrer nos deux disciplines dans une phase nouvelle.

Pierre COLMAN,

Assistant à l'Institut supérieur  
d'Histoire de l'Art et d'Archéologie.

### ANNEXE III

Le vocabulaire d'un texte se distribue en classes de fréquences : mots apparaissant 1 fois, 2 fois, ... $x$  fois. Le nombre de mots appartenant à l'une de ces classes en est l'effectif ( $f_x$ ) : tant de mots apparaissant 1 fois ; tant, 2 fois, etc...

La distribution du vocabulaire d'un texte étant connue, quelle est la distribution que l'on doit s'attendre à trouver dans un échantillon de ce texte, si la répartition des occurrences des mots est aléatoire ?

Le but de cette recherche n'est évidemment pas de permettre une prévision, mais de fournir les éléments d'une comparaison. Si la distribution réelle d'un échantillon ne diffère pas significativement des chiffres théoriques, on peut admettre l'hypothèse des chiffres théoriques, on peut admettre l'hypothèse d'une répartition aléatoire et donc d'une homogénéité du texte au point de vue de la richesse et de la distribution du vocabulaire. Mais si les différences sont significatives, on doit affirmer que l'échantillon présente, par rapport à la totalité, des particularités de vocabulaire qui méritent un examen : il constitue, par exemple, un morceau qui, en raison de sa situation dans l'œuvre, a été spécialement soigné, etc...

Pour résoudre ce problème, j'ai défini une méthode qui repose sur une utilisation — peut-être inattendue — du binôme de Newton. En voici une brève description.

Si le texte compte  $M$  occurrences et que l'échantillon en compte  $m$ , la probabilité qu'une occurrence prise au hasard se trouve dans l'échantillon est  $p = \frac{m}{M}$ , et la probabilité qu'elle ne s'y trouve pas est

$q = 1 - \frac{m}{M}$ . A l'aide des valeurs de  $p$  et de  $q$  ainsi trouvées, on peut résoudre le problème posé.

En effet, un mot qui se trouve  $x$  fois dans l'ensemble du texte peut ne pas se trouver dans l'échantillon, ou bien s'y trouver 1, 2, ...  $x$  fois. La probabilité de chacune de ces hypothèses est mesurée par les termes successifs du développement du binôme  $(p + q)^x$ . L'effectif  $f_x$  de la classe des mots de fréquence  $x$  se répartit proportionnellement à ces probabilités entre les classes de fréquence 0, 1, 2, ...,  $x$  de l'échantillon. Il suffit donc de multiplier les termes du binôme par  $f_x$  pour connaître la contribution théorique de la classe de fréquence  $x$  aux diverses classes de fréquence de l'échantillon. L'effectif total d'une classe de fréquence de l'échantillon est la somme des termes partiels qui lui sont apportés par chacune des classes de la totalité du texte. Prenons un exemple. Si la totalité d'un texte