

Problèmes posés par les nomenclatures et l'agrégation

par Henri Duprat

Introduction.

L'élaboration des nomenclatures statistiques, longtemps traitée comme une tâche ingrate et mineure, pose en réalité des problèmes théoriques et pratiques, intimement liés à ceux rencontrés en économie industrielle. Les nomenclatures, et particulièrement celles de produits et d'activités, relient en effet les objets de la microéconomie à ceux de la macroéconomie et conditionnent l'étude de la différenciation industrielle, composante fondamentale du développement économique,

I — La construction des nomenclatures.

Outil qui sert à «nommer», une nomenclature a pour fonction d'identifier, de décrire et de classer les éléments d'un ensemble. Sa forme est celle d'une liste des parties composant cet ensemble. Le plus souvent, plusieurs partitions emboîtées donnent naissance à autant de nomenclatures articulées en niveaux successifs, selon un schéma arborescent qui apparente les nomenclatures aux plans de classement.

Cet aspect des nomenclatures statistiques est bien connu. Leur mode d'élaboration l'est moins. Nous en traiterons ici en nous référant aux nomenclatures d'activités et de produits. La démarche suivie et les difficultés rencontrées seraient analogues dans d'autres domaines, notamment en matière d'emplois et de professions¹.

La démarche technique de construction d'une nomenclature est classique : on cherche d'abord à identifier des «noyaux», groupes d'objets à la fois assez semblables et assez nombreux, puis on s'efforce de résoudre au mieux les problèmes de frontières, autour de cette configuration initiale formée de caté-

1. P. Lestang, «Le nouveau système français de nomenclatures d'activités et de produits», *Economie et statistique*, n° 51, déc. 1973, Paris.

gories disjointes. Le contour exact des catégories définitives se traduit par les mentions d'inclusion ou d'exclusion, dans chaque cas particulier, avec des renvois de poste à poste¹.

Ce travail ne se fait pas a priori, mais en fonction d'une structure industrielle qui existe, et notamment des associations d'activités qu'elle contient. Il présuppose donc une information technique et économique suffisamment détaillée.

D'autre part, une nomenclature est un langage : ce langage doit être commun aux entreprises et aux administrations. Les dénominations et les catégories utilisées dans les questionnaires statistiques doivent être aussi proches que possible de celles que les industriels ont l'habitude d'utiliser entre eux, dans la pratique technique et commerciale².

La construction d'une nomenclature est ainsi le résultat d'un dialogue et non d'un calcul. Certes les méthodes d'analyse de données peuvent apporter une aide utile, mais leur emploi suppose que ces données soient déjà disponibles, fiables et détaillées. La tâche principale reste la discussion avec les représentants des professions, pour parvenir à un accord, d'abord sur les «agrégats naturels aux yeux de l'homme de l'art», puis sur les limites des catégories.

Cet accord sur un langage commun est difficile et s'obtient finalement autour de compromis. Car des intérêts sont en jeu : le partage des tutelles administratives, l'organisation des syndicats professionnels, le champ d'application des conventions collectives font apparaître des découpages différents, souvent contradictoires, qui viennent interférer avec les structures industrielles et les usages commerciaux. La diversité des points de vue, des situations et des objectifs conduit chaque organisation ou chaque entreprise à élaborer ses nomenclatures propres selon ses besoins d'information et de décision^{3,4}.

Le développement spontané des nomenclatures entraîne ainsi leur incompatibilité mutuelle. Les efforts accomplis pour y remédier ne peuvent que limiter les divergences. De plus, les nomenclatures ne peuvent évoluer que très lentement⁵. Elles restent ainsi imprécises, malgré leur détail, et inadaptées en permanence, sans que les conséquences de cet état de fait apparaissent clairement.

// — *Nomenclatures «ouvertes» et nomenclatures «fermées».*

Or le mot même de «nomenclature» est ambigu et n'a pas le même sens pour un industriel que pour un statisticien.

1. P. Lestang, «La construction des nomenclatures», *Sociologia ruralis*, n° 4, 1978.

2. B. Guibert, J. Iaganier et M. Voile, «Essai sur les nomenclatures industrielles», *Economie et Statistique*, n° 20, INSEE, Paris, révrier 1971.

3. H. Duprat, «Conditions linguistiques du transfert technologique». Revue «Automatisme», tome XVII, n° 11, novembre 1972.

4. P. Lestang, «Harmonizing international commodity classifications: an urgent task», Proceedings of the second international symposium. Belgian Institute for Commodity Science and Technology. State University Gent, Belgium, 1979.

5. R. Rousseau, «Pourquoi change-t-on de nomenclatures ? *Economie et Statistique*, n° 70, septembre 1975.

Il s'agit bien d'une liste dans les deux cas. Mais dans une entreprise cette liste comporte un grand nombre d'identifiants très détaillés, par exemple des lignes d'inventaire d'articles en magasin, tandis qu'une nomenclature désigne en statistique la liste des catégories qui correspondent à une certaine partition d'un ensemble, catégories relativement peu nombreuses. Nous sommes ainsi conduits à distinguer des nomenclatures «ouvertes» et des nomenclatures «fermées».

Une nomenclature «ouverte» est une nomenclature dans laquelle le nombre de postes est a priori illimité : ces postes sont homogènes, et si un objet nouveau apparaît, un poste nouveau doit être créé. Les rubriques sont donc le plus souvent très nombreuses et d'une importance extrêmement inégale : diverses procédures pratiques ont pour but de concilier les impératifs d'économie dans l'allocation des ressources avec le nombre de postes imposé par la nature des choses, par exemple en gestion de stocks. Un cas extrême est celui de la nomenclature d'articles d'approvisionnement de l'O.T.A.N., qui comporte près de cinq millions d'articles¹,

Au contraire, une nomenclature «fermée» a un nombre de rubriques invariable, en général compris entre une dizaine et quelques centaines ; les nomenclatures douanières sont les plus détaillées et atteignent quelques milliers de postes. Lorsque des objets nouveaux apparaissent, ils doivent être classés dans la catégorie dont ils diffèrent le moins, tant qu'une révision de la nomenclature n'a pu intervenir. On s'efforce d'autre part d'assurer un relatif équilibre d'importance entre les diverses rubriques. Ces rubriques sont choisies aussi homogènes que possible, mais une hétérogénéité plus ou moins grande est inévitable pour la plupart d'entre elles.

Ces différences entre les deux types de nomenclatures s'expliquent par des fonctions très différentes,

Les nomenclatures ouvertes sont des outils de gestion et ont pour rôle de fournir des identifiants sur lesquels on puisse raisonner, ou, à plus forte raison, en charger un ordinateur, en lieu et place des objets eux-mêmes, tandis que les nomenclatures fermées, que sont les nomenclatures statistiques, servent à donner une description résumée d'ensembles nombreux^{2,3}.

Les rubriques des nomenclatures statistiques devraient donc, en principe, correspondre à des regroupements plus ou moins vastes des postes des nomenclatures de gestion, et leur perfectionnement consiste à accroître le détail et la précision technique de cette correspondance⁴.

Mais une différence essentielle subsiste en tout état de cause : dans le cas des nomenclatures de gestion, les catégories agrégées sont de simples repères

1. Bureau interarmées de Codification des Matériels, «Nomenclature interarmées ». *Notice à l'usage des services d'achat et des fournisseurs des armées*, lid. mai 1977 (gratuit sur demande) B.I.C.M., Fort de l'Est, B J>. 257, 93205 Saint-Denis Cedex 1.

2. AFNOR - ACANOR, *Guide de la codification*, éd. AFNOR, Paris 1981.

3. National Computing Center, «Commodity classification and Naming», *N.C.C. Ltd*, Manchester 1971.

4. R. Rousseau, «La NOOEP, nomenclature détaillée de produits», i)fo<|Vo<eî, juillet-août 19K1 .édité par l'Observatoire Economique de Paris.

de classement, destinées à faciliter la recherche des identifiants des objets, tandis que les catégories des nomenclatures statistiques deviennent une représentation du réel. Quelle est la validité de cette représentation ?

/// — *Le problème de la pertinence des catégories.*

La question est de savoir si l'on peut raisonner sur des catégories d'objets comme sur des objets, ou s'il existe des différences de type logique, au sens de Russell, entre les objets et les catégories d'objets. La tentation est grande de raisonner sur un poste de nomenclature comme si tous les éléments qu'il rassemble avaient les mêmes propriétés, et de faire dire aux libellés de nomenclature ce qu'ils ne contiennent pas, soit parce que l'information n'a jamais été disponible, soit parce qu'elle a disparu dans une opération d'agrégation antérieure.

L'expérience montre qu'il existe un «niveau de pertinence», au-delà duquel, si l'on poursuit l'agrégation, les catégories cessent de donner une représentation valable du phénomène étudié. C'est ainsi qu'en matière d'environnement on a cherché à se servir des Nomenclatures d'Activités de 1973 pour identifier les industries polluantes et celles qui ne le sont pas.

Or, on trouve, aux divers niveaux :

- la classe 21 «Travail des métaux»
- le groupe 21.03 «Traitement et revêtement des métaux».

U n'existe pas de nomenclature plus détaillée d'activités. Mais la Nomenclature Détaillée de Produits distingue divers façonnages au sein du groupe correspondant et permet d'isoler :

- 21.03.02 «Revêtement métallique des métaux».

Cette rubrique comprend, entre autres, le chromage. Or les risques de pollution sont notablement différents selon que l'on utilise des bains de chrome trivalent ou hexavalent. En tenir compte exige donc une information beaucoup plus détaillée que celle véhiculée par les nomenclatures existantes. Ajoutons d'ailleurs que le Répertoire des Etablissements ne contient que l'indication de l'activité principale exercée («code APE» à 4 chiffres) et non celle des autres activités existant dans l'établissement. L'étude de l'impact sur l'environnement de l'activité industrielle exige donc un complément d'information, dont la collecte sur le terrain implique l'emploi d'une nomenclature «ad hoc», plus détaillée et plus technique, plus proche du niveau de la pratique industrielle elle-même. Une telle nomenclature est d'ailleurs utilisée depuis longtemps par le Service des Etablissements Classés.

Cette notion de niveau de pertinence apparaît dans de nombreux domaines. Elle conduit l'INSEE à pousser plus loin l'analyse des activités en cherchant à identifier les techniques de production mises en œuvre au sein des unités statistiques recensées, c'est-à-dire des entreprises et des établissements. Cette

1.- O. Arkhipovl, «Peut-on mesurer le Bien-être National ?» in *Collections de l'INSEE* (Série Cn° 41), avril 1976.

identification est nécessaire dès lors que les diverses techniques rencontrées au sein d'une même activité peuvent avoir des incidences très variables dans un domaine donné, que ce soit sur l'emploi, les conditions de travail, les consommations d'énergie, l'approvisionnement en matière première, l'intervention d'une technique nouvelle, la protection de l'environnement, etc.. Dans chaque cas, une information fine est nécessaire, dont il faut préciser le mode de collecte : unité à observer et mode de description.

IV — La description du système productif.

La réalité industrielle est faite de processus élémentaires de production, qu'il faut identifier et décrire. Lorsqu'on analyse ce qui se passe dans une usine, un moment vient où le processus de production cesse d'être physiquement décomposable : on ne peut plus envisager de l'interrompre pour constituer deux phases distinctes et successives entre lesquelles pourrait intervenir un marché. Ces processus élémentaires peuvent d'ailleurs changer avec l'évolution technique : par exemple, la commercialisation de béton prêt à l'emploi fait apparaître deux opérations techniques distinctes là où il n'en existait qu'une.

Dans l'ensemble de la pratique industrielle, il existe un très grand nombre de ces opérations techniques élémentaires. Ce sont les «éléments simples» dont les combinaisons constituent le système productif, de même que les éléments du tableau de Mendeleev constituent la base de la nomenclature chimique. Comme en faire l'inventaire et la description ? L'organisation actuelle des systèmes d'information technique ne s'y prête guère. Une étude a été faite par le CNAM, il y a quelques années, sur un projet de banque de données technologiques. Ce propos fut abandonné, mais semble devoir être repris. C'est un outil de travail nécessaire¹.

Il existe ensuite différentes façons de regrouper en catégories ces unités élémentaires. Les notions classiques d'entreprise et d'établissement résultent de la prise en compte de la personnalité juridique et de l'unité de lieu. La notion de filière, elle, apparaît lorsque l'on suit un même flux de matière à travers ses transformations successives. Mais on peut remarquer que le terme de filière avait été utilisé, il y a quelques années, dans l'industrie nucléaire, pour désigner les formes successives apparues dans le développement d'un même concept technologique : on parlait ainsi de «filière française» à propos des réacteurs à uranium naturel, modérés au graphite et refroidis au gaz carbonique. Le terme de «lignée» eût été, semble-t-il, mieux indiqué. Cette notion d'unité technologique doit, en tout cas, être conservée, avec l'ensemble du savoir-faire qu'elle implique.

Nous retrouvons ici autant de modalités différentes de regroupement des activités élémentaires que de dimensions différentes dans la description d'une

1. H. Duprat, «Un projet de base de données technologiques», *Courrier des Statistiques* n° 4, octobre 1977.

unité de production : des hommes et des équipements mettant en œuvre un savoir-faire et transformant des flux de produits. L'analyse économique classique a réduit pour l'essentiel ces dimensions à une projection sur le plan comptable, au prix d'une simplification grossière de la réalité industrielle. Une description du système productif conservant les données qualitatives nécessaires à une bonne compréhension de l'industrie moderne n'est possible que si les caractéristiques physiques des activités et des produits, en quantités et en performances, gardent un sens.

Deux conditions majeures devraient, selon nous, être remplies, l'une pratique, l'autre théorique.

D'une part, la description des techniques n'est pas séparable de l'identification et de la description des produits : les procédés et le savoir-faire sont matérialisés par des équipements et mesurables par les caractéristiques techniques de ceux-ci. Les progrès de l'industrie de la mesure sont la mesure des progrès de l'industrie, et l'espace de description des produits constitue le référentiel nécessaire à la représentation du système productif¹.

D'autre part, ce système productif est lui-même formé d'un très grand nombre de sous-systèmes, liés par de multiples interactions. La description d'un tel ensemble pose encore quelques problèmes théoriques, «Un ensemble», dit Bourbaki, «est formé d'éléments susceptibles de posséder certaines propriétés, et d'avoir entre eux, ou avec les éléments d'autres ensembles certaines relations»... Les molécules d'un gaz, les grains d'un sac de blé, les pièces issues en série d'une machine constituent des ensembles d'éléments possédant certaines propriétés. On peut appeler «populations» de tels ensembles. Leur description est familière aux statisticiens. Les atomes d'une molécule, les organes d'un être vivant, les pièces d'une machine constituent des ensembles d'éléments ayant entre eux certaines relations. On peut appeler «systèmes» de tels ensembles. Leur description, s'ils ne sont pas isolés et peuvent donner lieu à couplage, est beaucoup plus difficile, et l'on confond souvent des relations entre agrégats avec des agrégats de relations,

V - Economie et technologie : vers une science des techniques et des produits ?

Ces deux conditions peuvent-elles être remplies ?

Oui, sans doute, en ce qui concerne la seconde : le nombre, la diversité, la richesse des travaux en cours permettent d'espérer de la synthèse de la théorie des jeux, de la théorie des systèmes, de la théorie des catastrophes, de la théorie de la décision, etc. dans une «dynamique générale» cohérente.

En revanche, la première condition peut sembler utopique : comment imaginer la collecte et le traitement de telles données détaillées ?

1. H. Duprat, « Conditions linguistiques... », op, cit.

2. P. Mehl, «On the Key-significance of the term cCharacteristcs», *Forum Ware*, 8 (1980).

Une possibilité imprévue nous semble pourtant ouverte par le développement actuel de banques de données sur les produits¹. L'information industrielle et commerciale répond en effet à un besoin impératif des entreprises³. Ce besoin suffira à financer le développement des banques de données. Mais celles-ci, à la différence des supports antérieurs, sont disponibles pour d'autres utilisations à un coût très faible.

L'évolution prévisible des systèmes documentaires, en information industrielle, en assurant ainsi la collecte et la saisie informatique de la description des produits, pourrait bien fournir enfin ses données expérimentales à cette «science des produits» fondée par Beckmann au XVIII^e siècle, et injustement négligée en France⁴. En même temps, deviendrait possible une véritable technologie, science qui traiterait des techniques comme la biologie traite des êtres vivants, science nécessaire aux progrès de l'économie industrielle.

Fermer cette fenêtre pour revenir au Sommaire

1. H. Duprat, «Rôle économique et problèmes documentaires du catalogage industriel», *Courrier des Statistiques*, n^o 12, octobre 1979.

2. H. Duprat, «La segmentation des marchés de l'information industrielle», à paraître *Courrier des Statistiques*, n^o 22, avril 1982.

3. O. Gekeler, K.D. Herdt, W. Oberender, «Waren katalogisierung und Kommunikation über die Ware», *Verlag Dokumentation*, Pullach bei München, 1973.

4. J. Holzl, «Significance and content of technology and commodity science in a classification system of subject fields», *Proceeding of the second symposium*, Belgian Institute for Commodity Science and Technology, State University Gent, Belgium 1979.