

Conclusion

Jamais un questionnaire d'enquête n'aura été si simple dans la conception de ses questions. Jamais un questionnaire n'aura été si neutre, si primaire pourrait-on dire, pour ce qui est des réactions aux stimuli proposés : agréables ou désagréables. C'est donc une simple liste de mots qui, bien qu'elle puisse paraître quelque peu arbitraire, a permis de mettre en évidence des traits structuraux sous forme de dimensions stables.

Il aura fallu procéder à des complétions de dizaines de milliers de questionnaires pour valider ces traits structuraux dans le temps et dans l'espace et pour établir un repère socio-culturel permettant de positionner des individus et des groupes d'individus. Malgré l'épreuve périlleuse de la traduction, les mêmes traits apparaissent dans plusieurs pays différents, confirmant ainsi que les résultats ne dépendent pas de la composition exacte de la liste, et que la structure obtenue n'est ni locale, ni nationale....

Par sa transparence, sa distance vis-à-vis de l'actualité, ce repère est propice à des comparaisons intergénérationnelles, interrégionales et internationales. Mais on peut aussi positionner dans ce repère, en procédant à des enquêtes « ad hoc », des attitudes, des opinions, des produits, des supports médiatiques...

Cet outil étonnant a lui-même été objet de recherches, exposées en détail au cours des chapitres qui précèdent, et il continuera à l'être.

La seule hypothèque resterait le choix de la liste de mots. Plusieurs des travaux présentés dans cet ouvrage montrent précisément que ce choix ne constitue pas une hypothèque dirimante, et que le caractère opératoire de cette liste laisse augurer d'autres résultats et d'autres développements.

Une façon pragmatique de « vivre cette hypothèque », à défaut de la lever, est de dire que l'on observe non pas une structure universelle, mais la *trace* de ce qui pourrait être une structure universelle. Un « questionnaire total », aussi fictif et utopique que le « doctorat total » de Ionesco, permettrait d'observer cette structure. La liste de mots retenue n'en retient qu'une trace qui en est une sorte de projection, conditionnée à la fois par des critères de coût et de faisabilité. S'il est pratiquement impossible de prouver que cette « projection » est optimale vis-à-vis de ces critères, nous avons montré qu'elle n'est ni fragile, ni sans intérêt.

On ne s'étonnera pas de nous voir tirer des enseignements – et tracer des perspectives de travail et de recherche – dans des registres très divers, relevant de la statistique mathématique, de la sociologie, de la linguistique, enfin de la méthodologie d'enquête.

– *Comment passe-t-on du plaisir au sens ?*

Cette première interrogation est suscitée par la simplicité même du questionnement. Comment des questions fondées uniquement sur des sensations agréables ou désagréables peuvent-elles déboucher sur une structure sémantique locale aussi fine... ou sur une architecture globale de valeurs aussi stable ? L'émotion et la satisfaction jouent probablement un rôle important dans l'acquisition du langage comme dans tout apprentissage, et même, certainement, dans la construction progressive d'une langue. Au niveau de cet atome linguistique qu'est le mot, peu d'autres qualifications bipolaires pouvaient s'appliquer de façon exhaustive. Le caractère opératoire des réponses obtenues fut une preuve (agréable) du caractère judicieux de l'échelle choisie.

Des travaux de recherche complémentaires restent cependant à mener pour mieux analyser les mécanismes qui ont effectivement conduit de l'agréable au pertinent. Dans le cas du questionnaire ouvert utilisé au chapitre 4, pour lequel il ne s'agit plus de noter une liste fixe de mots, il est possible que le concept d'« importance » soit plus adéquat que celui d'« agrément », pour opérer une sélection et un filtrage plus sévères et

limiter ce faisant le caractère peut-être trop ludique, dans ce contexte libre, de l'exercice. Ce caractère ludique contribue, on l'a vu, à l'extrême dispersion du vocabulaire dans les réponses individuelles.

Les travaux sur les questions ouvertes devraient donc susciter de nouvelles recherches, comportant des phases d'expérimentation assez lourdes.

– *Enseignements sur les enquêtes tendant à mesurer les évolutions... et aussi sur l'évolution des enquêtes*

Il est généralement admis que les mesures de différences entre réponses à des questions subjectives, différences dans le temps ou entre catégories, sont beaucoup plus justifiées et prudentes que des mesures absolues. Les résultats obtenus à propos de la dérive simultanée de l'instrument d'observation et de la réalité observée sont une (petite) pierre dans le jardin de ceux qui pensent qu'il est toujours licite d'observer les évolutions d'une question dont le libellé est fixe dans le temps. L'extrême homogénéité et la relative intemporalité des « questions » du questionnaire sémiométrique permet d'étudier assez finement ces phénomènes d'attitudes ainsi que les problèmes de participation et de notations. L'évolution dans le temps¹ illustre bien l'absence de neutralité et donc l'implication de l'instrument d'observation à la fois dans les résultats observés et dans la qualité de ces résultats².

– *Le modèle sémiométrique et la « double convergence »*

Parmi les apports fondamentaux des techniques statistiques connues sous le nom d'analyse des données³, nous citerons d'une part le renouveau d'intérêt apporté aux individus dans les enquêtes, ou plus

¹ Section 5.8 du chapitre 5.

² Notons que la dégradation très progressive de la participation aux enquêtes, qui concerne à la fois les refus *a priori* et la qualité des complétions est un phénomène assez largement observé dans les instituts de sondage et dans les instances de contrôles de la profession. Le questionnaire sémiométrique, et aussi, nous l'avons vu, le questionnaire ouvert du chapitre 4, peuvent contribuer à étudier certains aspects de ce phénomène, en particulier l'engagement et le degré de participation, pour les personnes ayant accepté de répondre.

³ Ces techniques sont apparues dans les années 70, notamment en France, sous l'impulsion de Jean-Paul Benzécri (1973) qui préconisait une totale réécriture de la science statistique prenant en compte les nouvelles possibilités de l'informatique.

exactement à la « dimension individu »¹, et, d'autre part la prise en considération de variables pouvant être très nombreuses. Corrélativement, les modèles probabilistes classiques, peu réalistes parce que beaucoup plus contraignants dans le cas *multivarié* que dans le cas d'une ou deux variables, sont moins utilisés, laissant souvent la place à des méthodes de validation empiriques à base de calculs intensifs comme les techniques de *bootstrap* auxquelles il a été largement fait appel dans cet ouvrage. A la fois les individus gagnaient en importance, et les variables se banalisaient, autrement dit, les lignes (en général : les individus) et les colonnes (en général : les variables) du tableau classique de données jouaient des rôles de plus en plus symétriques².

D'où une potentialité de double convergence selon les lignes (individus) et/ou les colonnes (variables) qui généralise en quelque sorte la loi des grands nombres et d'autres lois limites de la statistique, fondées sur la multiplication à l'infini des seuls individus... Ces problèmes sont simplement évoqués ici, mais appartiennent à une direction de recherche importante.

Ce nouveau regard sur les tableaux de données fut, et sera, car il s'en faut de beaucoup que ces innovations aient pénétré toutes les disciplines, très enrichissant pour beaucoup d'applications en sciences humaines. La multiplication des variables change la nature des problèmes.

Les praticiens d'études de marché savent bien que quelques interviews très riches valent parfois mieux que des centaines d'interviews pauvres... ce que les cliniciens ont su de tous temps. A la limite, risquons la boutade suivante : les *Essais de Montaigne* (échantillon de taille : 1) nous apprennent beaucoup plus, sur l'homme en général et l'homme du XVIème siècle en particulier, que ne l'aurait fait une enquête de satisfaction auprès de quelques milliers de ses contemporains. Montaigne en est d'ailleurs bien conscient lorsqu'il écrit : « Chaque homme porte la forme entière de l'humaine condition », ce que Sartre exprimera de façon plus lapidaire : « Tout homme est tout l'homme »³.

¹ Les individus statistiques (ou observations) ne servaient qu'à calculer des moyennes ou des corrélations, et, détail significatif, n'étaient que rarement représentés sur les graphiques avec des identificateurs.

² D'ailleurs l'opération mathématique de compression qui est à la base des méthodes factorielles, la décomposition aux valeurs singulières, fait jouer un rôle symétrique aux lignes et aux colonnes (cf. annexe A1.2).

³ Sartre J.-P. (1980) *Les Mots*, Gallimard, Paris.

Et la sémiométrie dans tout ça ? Elle constitue l'exemple emblématique de ce paradigme, le *leading case* de la double convergence¹. En effet, il existe une population parente des lignes (la population française, par exemple) et une population parente des colonnes (le vocabulaire de la langue française, pour le même exemple). Peu de questionnaires se trouvent dans une situation théorique aussi favorable. Mais prendre simplement des mots au hasard dans un dictionnaire engendrerait, au centuple, les difficultés² rencontrées à propos du questionnement ouvert du chapitre 4. Parallèlement aux travaux théoriques qu'il serait intéressant de mener sur ce modèle statistique complexe, les expérimentations à venir devront aussi porter sur les problèmes de sélection et d'échantillonnage de mots.

– ***Structure globale, fractale ou individuelle ?***

Si l'on effectue des mesures sur des individus, les corrélations traduisent bien une liaison mesurée au niveau individuel. Certains notent bien à la fois les mots *Discipline* et *Patrie*, d'autres les notent moins bien, mais il y a peu d'individus qui notent bien *Discipline* et mal *Patrie*. La corrélation vient précisément de cette cohérence interne à chaque individu, cohérence que la composition du questionnaire permet de vérifier, et que les quelques légères redondances existant dans le questionnaire permettent de conforter. S'il existe une structure de corrélation, significative statistiquement, sur un échantillon de n individus pris au hasard, on doit retrouver la même structure sur un autre échantillon au hasard, ou dans un sous échantillon pris également au hasard.

Ce n'est pas à propos de ce phénomène classique que l'on a pu parler de fractalité, mais à propos du phénomène plus paradoxal suivant : des catégories semblent s'opposer sur certains axes de façon significative. En segmentant selon ces catégories, on prend un sous-échantillon délibérément non aléatoire. Et l'on s'étonne, cette fois, de retrouver la même structure à l'intérieur de ces sous-catégories. Dans le cas de

¹ L'exemple le plus pertinent permettant d'étudier l'évolution (la limite) du phénomène lorsque le nombre de lignes et le nombre de colonnes augmentent indéfiniment.

² Ouvrons le Petit Robert au hasard et pointons, aveuglément, sur des mots : « *démission, hectare, monazite, papaine, prélude, rabattre, toupie...* ». Les limites de faisabilité sont tout de suite perceptibles : il faut se restreindre aux mots connus *a priori* de tout échantillon de répondants (une petite fraction des entrées d'un dictionnaire), puis introduire une série de contraintes qui conduisent finalement à un protocole analogue (mais pas forcément identique) à celui qui a prévalu lors des premiers essais sémiométriques (*cf.* chapitre 1).

l'opposition homme-femme, la surprise vient peut-être d'une identification hâtive de certains archétypes. La liaison très significative entre les mots de l'« Attachement » et l'appartenance au sexe féminin fait penser que la différence entre sexes est seule responsable de la formation de l'axe « Attachement / Détachement ». L'existence de la même opposition à l'intérieur de chaque sexe montre donc que l'opposition entre sexes n'est en aucune façon responsable de la structure¹.

En fait, ce que l'on appelle ici *fractalité* pourrait être simplement la validité « quasi-individuelle » d'une structure qui serait alors à la fois sémantique et *psycho-culturelle*. Les proximités locales observées sur les cartes sémiométriques sont influencées par les liens sémantiques qui sont universels parce que propres au vocabulaire d'une langue commune à tous les répondants. Les grandes oppositions observées sur les axes principaux proviennent de facteurs latents qui pourraient dépendre d'une organisation psychique indissociable d'un cadre culturel général (de références, de valeurs) également commun à tous les répondants, avec parfois des nuances déclinables selon certaines catégories socio-démographiques. Le lecteur aura compris que les conditionnels désignent en fait des thèmes d'approfondissement et de recherche à venir.

– *Pour une statistique structurale*

Ce n'est pas une activité courante de la statistique que de mettre en évidence des structures cachées, bien que cette démarche soit historiquement apparue dès le début du vingtième siècle - dans le cadre de modèles spécifiques - avec les travaux des psychologues factorialistes.

Nous nous sommes intéressés dans cet ouvrage à la notion de *fait statistique*, en considérant les outils utilisés (analyses en composantes principales, classifications, cartes de Kohonen) comme de simples instruments d'observation du *multidimensionnel* au même titre que les microscopes ou des appareils radiographiques² sont des instruments d'observation du *petit* ou de *l'opaque*. En utilisant une importante

¹ Celle-ci est stable même à l'intérieur de la plupart des catégories socio-démographiques usuelles... mais l'ordre des axes peut cependant être modifié : ainsi le second axe, fortement lié à l'âge des répondants, passe au troisième rang si l'on analyse une tranche d'âge étroite comme la tranche d'âge 40-45 ans.

² Les théories et mécanismes qui président à la réalisation d'un microscope n'ont pas ou peu de rapport avec les propriétés morphologiques des animalcules, cellules ou autres objets observés. Ils ne supposent aucune modélisation des objets observés.

panoplie de procédures de validation (méthodes de rééchantillonnage spécifiques, recodages, transformations analytiques diverses, procédures empiriques extensives), nous avons montré que l'on pouvait extraire des *structures valides* sans se contenter de contempler des *patterns suggestifs*. En fait, on commence seulement à pouvoir mettre en œuvre les outils qui permettent de donner un caractère de scientificité aux techniques d'analyse des données conçues il y a plusieurs décennies¹... et c'est cette démarche de recherche de *faits statistiques* ou de *traits structuraux validés* qui a été désignée par l'expression *statistique structurale*. Les travaux présentés dans les chapitres précédents pourraient paver le début d'un chemin menant vers une telle branche disciplinaire.

¹ Si les statisticiens spécialistes ont utilisé ces instruments d'observation avec le discernement et le recul critique indispensable, beaucoup d'usages malheureux furent cependant à déplorer, en particulier à cause de l'absence de cadre conceptuel pouvant accueillir les nouvelles visualisations. L'analogie est assez frappante avec la diffusion des premiers microscopes telle que la décrit François Jacob dans *Logique du vivant* : « Quand Leeuwenhoek contemple pour la première fois une goutte d'eau au microscope, il y trouve un monde inconnu, des formes qui grouillent... Mais la pensée n'a alors que faire de tout ce monde... cette découverte permet seulement d'alimenter les conversations... ».