

J.-P. BENZÉCRI

Interprétation logique et analyse multidimensionnelle

Les cahiers de l'analyse des données, tome 19, n° 4 (1994),
p. 475-484

http://www.numdam.org/item?id=CAD_1994__19_4_475_0

© Les cahiers de l'analyse des données, Dunod, 1994, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

INTERPRÉTATION LOGIQUE ET ANALYSE MULTIDIMENSIONNELLE

[INTERPRÉTATION]

J.-P. BENZÉCRI

1 Déduction et interprétation

1.1 Le sens du discours

Selon une conception déductive, le sens d'un discours est déterminé par la combinaison des sens de ses éléments. On peut affirmer, au contraire, le rôle indispensable d'une interprétation.

C'est clair pour le discours ordinaire, généralement ambigu: ambiguïté syntaxique, si les formes des mots se prêtent à plus d'une combinaison hiérarchique compatible avec les règles de la grammaire; ambiguïté sémantique, si tel mot ou locution offre plusieurs sens entre lesquels il faut choisir; ambiguïté pragmatique, si l'intention de celui qui parle, pour louer ou ironiser, n'est pas clairement révélée par son seul discours.

Dans de tels cas, l'ambiguïté ne se résoud que par une interprétation fondée sur le contexte. Pour faire image, nous dirons que, dans sa propre langue, tout lecteur, tout auditeur, est comme dans une langue étrangère qu'il connaît imparfaitement et dont il doit, parfois, deviner le sens.

Ce qui, dans le discours ordinaire, résulte de la maladresse ou d'une élégance qui ne s'embarasse pas de rigueur là où les i se passent de points, est, au contraire, un effet de l'art, ou plutôt la condition nécessaire de l'expression, pour le discours philosophique ou la Révélation céleste.

Platon eût-il mieux pu inciter à penser qu'en proposant des dialogues dont les suggestions, reçues dans des esprits bien disposés, prendraient plus tard forme de doctrines pour lesquelles les mots restaient à définir. Et, certes, il ne convenait pas que la théologie de l'Évangile dont, en dépit de la solennité de la Révélation, des hérétiques ont, de tout temps, proposé une interprétation insignifiante, fût prêchée, d'abord, selon l'ordre des traités dogmatiques qui en défendraient, au cours des siècles, la transcendance.

Mais qui considère attentivement, le compte-rendu de tel débat public, où Saint AUGUSTIN affronte un évêque arien, découvre que, tandis que celui-ci pose sur un seul plan les versets sacrés, le Docteur berbère en restitue la perspective comme pour une ascension vers le mystère (cf. P.L., Tome 42).

L'interprétation, telle que celle mise en œuvre par les Pères de l'Église pour comprendre l'Écriture Sainte, n'est pas un jeu improvisé; mais, non moins que la logique strictement déductive, c'est un système formel; système dont il conviendrait de distinguer les lois, pour servir à l'informatique des corpus de documents.

Certains tentent de mesurer l'intelligence par une note, qui, comme celle dite QI, se calcule d'après les réponses à un ensemble de questions. En principe, la réponse exacte requiert certaines connaissances ou aptitudes opérationnelles; mais nous dirons que parfois il s'agit plutôt d'un piège, le sujet devant prendre la question à la lettre et s'abstenir de toute interprétation.

L'enfant à qui l'on demande "la couleur du cheval blanc d'Henri IV", n'hésite à répondre "blanc" que parce qu'il imagine que, pour que la question ne soit pas absurde, "cheval blanc" doit être compris comme un seul mot, désignant la monture des jours de bataille ou de triomphe...

1.2 Science physique et interprétation des phénomènes observés

La connaissance de la nature visible a depuis des siècles pris la forme rigoureuse d'une physique dont les termes sont liés par des lois mathématiques. Mais entre ces termes et la réalité, le contact n'est pas si intime que la nature elle-même soit liée par les formules. Parfois, l'on découvre que celles-ci impliquent des conséquences contradictoires; d'autant plus strictes que l'énoncé mathématique en est plus rigoureux. Dans le premier quart du XX-ème siècle, il fallut fonder sur de nouvelles bases la science de la matière, de l'espace et du temps.

Quand, dans son mémoire de 1905 "Sur l'électrodynamique des corps en mouvement" (*Zur Elektrodynamik der bewegter Körper*), A. EINSTEIN introduit explicitement l'hypothèse relativiste, entendant remédier aux inconséquences des schémas reçus, il pose comme absolument vrai le principe que "pour tout système de coordonnées où valent les équations de la mécanique valent aussi les mêmes lois électrodynamiques et optiques" (*...für alle Koordinatensysteme, für welche die mechanischen Gleichungen gelten, auch die gleichen elektrodynamischen und optischen Gesetze gelten,...*). Il ajoute que, dans le vide, la lumière se propage avec une vitesse déterminée, toujours la même. Les coordonnées d'espace et de temps sont alors définies relativement à un cadre, ou réseau, solide avec, en chaque nœud, une horloge; un système de coordonnées équivalent pouvant être fourni par un deuxième cadre, en mouvement uniforme relativement au premier, avec ses propres

horloges. Ainsi la mesure explicite des coordonnées d'un événement résulte, pour EINSTEIN d'une "expérience imaginée" (*Gedankenexperiment*).

Or, dans ces constructions, comme l'auteur le note explicitement, tout s'accorde, au premier ordre d'approximation, avec le schéma reçu; le trait de génie consiste à distinguer, dans le tableau qui s'offre du réel, ce qu'on interprétera pour être absolument vrai afin de lui donner la dignité de l'idée.

1.3 Fondations logiques et interprétation en mathématiques

Au sein même des mathématiques, d'une part, l'expression serait impossible sans raccourcis et abus de langage. D'autre part, les démonstrations de non-contradiction qu'apporte la logique formelle, établissent seulement, e.g., que l'adjonction de tel axiome à telle théorie ne rendra pas celle-ci contradictoire, pourvu qu'elle ne le soit pas déjà; ce que rien ne permet, toutefois, d'assurer.

Le paradoxe de l'ensemble de tous les ensembles (ensemble dont une partie X pourrait être définie comme "l'ensemble de tous les ensembles x qui ne vérifient pas la relation $x \in x$ "; partie X pour laquelle on ne peut postuler ni $X \in X$, ni la négation de cette relation), fut résolu en introduisant la distinction entre ensemble usuel et classe.

On peut imaginer que, selon un théorème encore indémontré!, le fait de recéler une contradiction soit inhérent à tout système d'axiomes; la contradiction pouvant toutefois se résoudre en introduisant une semblable distinction; afin que le cercle vicieux se résolve en une hélice vertueuse (; cf. [ÂME], §5, *in fine*, où l'hypothèse de résolution est justifiée d'après l'existence d'un modèle physique).

Quoi qu'il en soit d'une telle hypothèse, l'histoire des mathématiques atteste que le corps n'en a subsisté que par des mises au point décidant de ce qu'il fallait prendre pour vrai, ou rejeter, d'un système imparfait. C'est ce qu'enseigne N. BOURBAKI, en une période fameuse qu'il place, comme un acte de foi, au terme de l'introduction à son traité:

Nous croyons que la mathématique est destinée à survivre, et qu'on ne verra jamais les parties essentielles de ce majestueux édifice s'écrouler du fait d'une contradiction soudain manifestée; mais nous ne prétendons pas que cette opinion repose sur autre chose que sur l'expérience. C'est peu, diront certains. Mais voilà vingt-cinq siècles que les mathématiciens ont l'habitude de corriger leurs erreurs, et d'en voir leur science enrichie, non appauvrie; cela leur donne le droit d'envisager l'avenir avec sérénité.

Procédé de convergence par correction où nous suggérons de voir une sorte d'interprétation.

Le calcul automatique est souvent présenté comme réalisant pratiquement la fonction de la machine de TURING. Pour grande qu'ait été la contribution des logiciens à la conception des ordinateurs, il nous paraît que cette assimilation est abusive. L'automate de TURING, qui a, en lui-même, un

nombre fini d'états possibles, calcule en déplaçant un ruban indéfini sur lequel il écrit ou efface les symboles de son alphabet. En cela, sa mémoire est potentiellement infinie.

L'ordinateur, notamment dans son traitement des nombres réels, nous apparaît, au contraire, plutôt comme un objet fini. Muni d'un programme, il opère sur l'ensemble des tableaux déclarés, qui sont comme les objets d'une catégorie, par des procédures, qui en seraient les morphismes. Dans le calcul sur les entiers, est substitué, à l'anneau infini Z , un anneau fini des entiers *modulo* une puissance de 2 (puissance 16, avec des entiers écrits sur deux octets; puissance 32 avec 4 octets...). Pour les réels, particulièrement si l'on calcule en virgule flottante, la structure mathématique n'est pas si régulière; mais c'est nécessairement une structure finie, dans laquelle les problèmes d'analyse reçoivent une traduction approximative.

Du point de vue des fondements des mathématiques, nous croyons que l'ordinateur offre une nouvelle classe de modèles strictement finis; mais se succédant fonctoriellement; des machines finies, de capacité indéfiniment croissante, tendant vers un modèle infini. (Ainsi l'on irait même au delà des exigences de l'intuitionisme: cf., e.g., [ÂME], §5.)

Pour un mathématicien de formation classique, la surprise est grande de voir réalisé par une machine, dont le mode d'emploi est compact, le programme que RUSSEL ne propose qu'en plusieurs volumes des *Principia*. Peut-être une interprétation de la nouvelle pratique du calcul suggérera-t-elle, pour celui-ci, de nouvelles fondations.

2 Modèles ensemblistes et modèles objectifs dans la pratique du raisonnement

Extension et compréhension sont des notions de logique, qui tout en étant familières aux philosophes, semblent se prêter à un modèle ensembliste, tel que ceux de la logique formelle. Ce parallèle est considéré dans [ÂME], §6, en partant de cette dernière discipline; on le reprend ici en partant du langage usuel. Puis on le retrouve en taxinomie animale et végétale.

M T
Tout homme est mortel

Homme (M) entre tout entier
dans l'extension de Mortel (T).

t M
Or Pierre est homme

Pierre (t) entre dans
l'extension de Homme (M).

t M
Donc Pierre est mortel

Pierre (t) entre dans
l'extension de Mortel (T).

2.1 Modèles du syllogisme

Sur l'exemple le plus commun, (cité d'après R. JOLIVET; *Traité de Philosophie*, T.I, §46, 2^o éd.; Vitte, Lyon-Paris; 1945) on peut voir l'inadéquation (aux applications pratiques) du modèle en extension qu'on donne du syllogisme.

Selon ce modèle, le nom "homme", avec le quantificateur universel "tout" représente l'ensemble de tous les hommes; le prédicat (ou la qualité) "mortel" est également pris pour l'ensemble de tous les individus mortels; enfin, "Pierre" tient lieu de l'ensemble à un élément, noté, en mathématiques, {Pierre}.

Or "homme" n'est pas, à strictement parler, un ensemble; son extension est indéfinie, car il s'étend aux hommes qui ne sont pas encore nés; et à ceux, déjà morts, auxquels la qualité de mortel n'appartient pas au sens usuel; on peut dire, en terme philosophique, qu'il s'agit d'un ensemble potentiel: mais une telle notion n'est pas reçue en mathématiques.

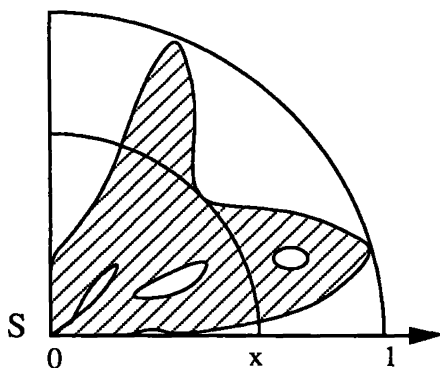
Associer un ensemble au prédicat "mortel" est encore plus incertain: faut-il prendre mortel comme se disant seulement des êtres vivants et décomposer l'ensemble de ceux-ci (encore moins bien délimité que celui des hommes) en "mortel" et "immortel"? À un prédicat désignant un état transitoire (affamé, endormi...), le modèle ensembliste en extension conviendrait encore moins bien.

Afin d'éviter de telles critiques, tout en poursuivant la rigueur mathématique, on a recouru au modèle en compréhension: où les termes, au lieu d'être des ensembles d'individus, sont des ensembles de propriétés (ou de qualités). Dans l'exemple choisi, T, mortel, est un ensemble réduit à une seule propriété; M, homme, est l'ensemble des qualités comprises dans la notion d'homme; enfin, t, Pierre, est l'ensemble des qualités que possède l'individu Pierre. Ainsi les relations d'inclusion sont inversées: la conclusion s'interprète: Mortel est compris dans l'ensemble des qualités de Pierre.

2.2 Théorie et pratique de la définition

En ce qu'il ne considère pas explicitement les sujets individuels, le modèle en compréhension s'accommode bien des ensembles potentiels tels que l'ensemble de tous les hommes. Mais il requiert des définitions parfaites. Or l'ensemble des qualités que possède Pierre est encore moins bien délimité, encore plus fluctuant que l'ensemble de tous les hommes.

Pratiquement, sans logique ni métaphysique, un homme est reconnu pour tel par contiguïté: est homme celui qui est né de parents humains; celui qui converse dans la société des hommes. Mais le progrès des techniques physiologiques pose aujourd'hui, comme un problème moral, le problème de la définition de l'homme: e.g., dans le cas d'un sujet inconscient, au tracé électroencéphalographique plat, et dont les fonctions végétatives sont entretenues mécaniquement. ARISTOTE a dit que, séparée du corps, la main n'est une main que de nom (homonymiquement): n'en est-il pas ainsi du corps entier séparé de la vie, de l'âme.



On a couvert de hachures le domaine des qualités que possède un individu i ; on convient d'attribuer à i la notion limite parce que, des qualités situées à la distance $x < 1$ du centre (sur un quart de cercle) il possède une part supérieure à $(1-x)$.

Nous avons proposé ailleurs (cf. [ÂME], in CAD vol.V, n°2, 1980), sur l'exemple même du vivant, de donner la définition, non comme une partie de l'ensemble de toutes les qualités, mais comme une sorte de filtre de parties sur cet ensemble. En bref, homme serait un point limite, H, adjoint à l'ensemble, Q, des qualités; H étant défini par une suite de parties emboîtées de Q; la possession des qualités qui sont dans les parties les plus étroites étant le plus nécessaire pour qu'un individu soit un homme. Ainsi, on peut appeler homme celui qui est malade; et même estropié; plutôt que celui qui n'a point de pensée. Mais à partir des dispositions matérielles, qui seules tombent sous les sens, la notion même de pensée ne se peut comprendre que comme une limite...

Sans prétendre que l'analogie de ce schéma permette de fonder une logique sur des calculs de limites, nous dirons que, selon les critiques que nous avons rappelées, les modèles reçus ne suffisent pas à rendre raison du raisonnement.

ARISTOTE a lui-même décrit minutieusement les fonctions logiques et ontologiques des termes du discours; il distingue entre substance et accident; entre contingent et nécessaire; entre le sujet, ce qu'on dit du sujet et ce qu'on dit être dans le sujet. Inventeur du syllogisme, il découvre aussi, dans l'être et la pensée, une structure dont le formalisme syllogistique ne nous paraît rendre compte qu'en partie.

Plus parfaite dans son calcul, la logique mathématique d'aujourd'hui, suit d'encore moins près cette structure.

2.3 Définition des espèces animales et végétales

En taxinomie animale et végétale, est généralement acceptée une hiérarchie taxinomique dont les degrés sont : {règne embranchement classe ordre famille genre espèce}; e.g.: {animal vertébré mammifère *rodentia cricetida* *Cricetomys gambianus*}, pour une espèce de gros rongeurs que l'on rencontre en Afrique tropicale; ce qu'on énoncera: l'espèce *gambianus* du genre *Cricetomys*, (qui appartient à une famille de rongeurs, les cricétidés,) est comprise dans la classe des mammifères (; laquelle rentre dans les vertébrés, embranchement du règne animal).

Mais à cette rigueur du langage, la Nature ne se conforme pas strictement.

D'une part, il existe des groupes naturels d'animaux, ou populations, disposés dans l'espace comme en une suite: les individus de groupes immédiatement voisins étant interféconds; tandis que la stérilité est de règle entre groupes éloignés, et particulièrement entre les groupes extrêmes de la suite.

On devrait dire que deux groupes consécutifs sont de même espèce; cependant, en terme mathématique, on doit dire que cette relation de ressemblance n'est pas transitive; ce qui empêche que toute la chaîne soit comprise dans une seule espèce.

Les ensembles actuels d'individus, que sont les populations, ne peuvent donc être universellement répartis suivant des ensembles potentiels, deux à deux d'intersection vide, qui seraient les espèces.

D'autre part, entre les individus d'une même population, la distribution des qualités n'est certainement pas uniforme.

Sans prétendre assimiler qualité et caractère marqué dans les chromosomes, le seul fait qu'en un même locus peuvent se placer l'un ou l'autre des allèles d'un gène, assure que la diversité qualitative des individus est au moins celle de l'ensemble des combinaisons de gènes compatibles avec la vie; l'identité des combinaisons afférentes à deux individus étant quasi impossible, hors le cas des jumeaux.

L'espèce devrait donc être définie par une équivalence entre combinaisons de caractères tous issus d'un même fond: mais, comme avec les populations contiguës, on doit craindre que cette équivalence ne réunisse des individus trop différents entre eux pour être de même espèce.

Ainsi, ni l'extension considérée dans les populations, ni la compréhension schématisée dans les chromosomes, ne suffisent à définir l'espèce.

3 Définition inductive et analyse des correspondances

3.1 Du qualitatif au quantitatif

Il y a un demi siècle, des philosophes épris de rigueur mathématique, tels R. CARNAP et Ch. MORRIS, voulaient refondre en un bloc l'ensemble des sciences suivant l'élaboration logique des données élémentaires de l'observation. Ce projet encyclopédique du "positivisme logique" tourna court.

Les termes de la description ne sont pas les termes du raisonnement. Entre les traits élémentaires des observations premières, il n'y a guère de relation qui doive subsister telle quelle dans le calcul logique ou le calcul différentiel. Afin de comprendre, il faut s'élever dans la hiérarchie des êtres. Un ensemble de qualités n'est pas une définition; un ensemble d'individus n'est pas une espèce. Plutôt que de combinaisons de qualités ou d'individus, il s'agit de termes d'un ordre supérieur que l'observation suggère mais ne donne pas.

Comprendre, c'est découvrir. La considération des découvertes déjà faites peut fortifier l'esprit pour les découvertes à faire; mais il n'y a pas de méthode algorithmique de la découverte.

Et nous prétendons pourtant, par l'élaboration statistique multidimensionnelle, amplifier la force de suggestion propre aux données.

Tous ne comprennent pas de même cette élaboration.

Certains préféreraient retrouver partout le modèle probabiliste d'un ensemble potentiellement infini d'individus décrit chacun par un système de variables aléatoires liées suivant une loi normale. Ils affirment poursuivre ainsi une étude quantitative, en réduisant les variables données à quelques facteurs qui en sont des combinaisons linéaires.

Nous répondrons, d'abord, en rappelant que, suivant ARISTOTE, "le caractère propre de la quantité est qu'on peut lui attribuer l'égal et l'inégal"; tandis que "semblable ou dissemblable se dit uniquement des qualités". De ce point de vue, une description multidimensionnelle est toujours qualitative même si elle ne comporte que des variables numériques précises, parce que la multiplicité des descriptions possibles est telle qu'on n'en rencontrera jamais d'égales, mais seulement de semblables.

Par l'analyse factorielle, on réduit le nombre des dimensions; par la classification automatique, on ramène l'infini à un nombre fini de types. C'est ainsi, qu'on va du qualitatif au quantitatif. Il ne s'agit pas d'accepter des données telles qu'elles sont relevées; mais de les élaborer dans une synthèse profonde qui découvre des entités nouvelles; et, entre celles-ci, des rapports simples.

3.2 Critique de la géométrie appliquée

Il faut ensuite prendre garde que la critique des données doit en guider le choix avant toute analyse.

Quelque méthode qu'on applique, mêler, dans la description d'individus, variables de taille (e.g., pour des pays: population; total des exportations...) et variables de forme (revenu moyen par ménage; espérance de vie à la naissance...) ne peut produire que des résultats contingents, dépendant du poids relatif des deux groupes de variables entre lesquels aucune règle ne permet d'arbitrer.

L'analyse de la correspondance entre deux blocs de variables est en butte à des effets plus fâcheux encore: si, explicitement ou implicitement, les deux blocs ont une intersection non vide (e.g., dans une enquête, du fait que l'âge des sujets est omniprésent), l'analyse est dominée par cette redondance inter-bloc, et ne montre aucunement les corrélations non triviales entre les blocs.

L'idéal est que le tableau considéré offre, d'un domaine homogène, une description, qui, d'un certain point de vue, soit exhaustive. D'ordinaire, le statisticien reçoit des données déjà recueillies qu'il doit analyser: il reste libre de maîtriser les effets de la redondance par un choix approprié des blocs de variables et l'introduction éventuelle de pondérations *a posteriori*.

Selon un modèle logique, telle définition délimite une espèce; telle conjonction de deux qualités en implique une troisième. L'analyse statistique prend des objets complexes dont les termes ne sont pas connus *a priori*: on ne sait quelle conjonction de symptômes a rang de syndrome; quel groupe de sujets doit être accepté comme présentant un type social. Mais en analysant la correspondance entre un ensemble I d'individus et un ensemble J de traits de description, ou entre deux ensembles J_a et J_b de traits, on peut découvrir des axes et des types.

La cohérence des entités ainsi construites n'est pas constante. De même que la superposition de nuées, étagées dans plusieurs plans, offre parfois aux regards de captivantes figures, de même il se peut que des données se projettent pour s'agréger en des termes inconsistants.

Disons plutôt que le réel est fait d'objets qui ne sont pas tous également solides: un groupe se constitue et se disperse, chaque personne subsiste; renonçant à l'art du diagnostic, un médecin a pu dire: il n'y a pas de maladie, il n'y a que des malades.

Autant et plus que les données, les résultats ne peuvent être reçus sans critique.

Pour SPINOSA, *more geometrico*, veut dire "en toute rigueur et certitude". L'analyse multidimensionnelle n'ira pas si loin: mais de ce qui, complexe par essence ne peut être disséqué sans perdre sa nature, elle suggère une interprétation qui se prête au discours et au raisonnement.

Références bibliographiques

A. EINSTEIN : “Zur Elektrodynamik der bewegter Körper”; in *Ann. Phys.*, Vol XVII, pp.891-921; (1905).

R. JOLIVET : *Traité de Philosophie: TI: Logique et Cosmologie*; 2-ème édition; éd. É. Vitte; Lyon-Paris; (1945).

Otto NEURATH, R. CARNAP, Ch. MORRIS & coll.: *International Encyclopedia of Unified Science*; the University of Chicago Press; parurent environ vingt fascicules; (≈ 1939).

N. BOURBAKI : *Éléments de mathématique*; Livre I: Théorie des ensembles, Introduction;

[ÂME] : J.-P. BENZÉCRI : “L’âme au bout d’un rasoir”; in *CAD*, Vol.V, n°2, pp.229-242; (1980).

[HÉTÉROGÈNE] : J. TCHOUANKAM: “Effets de la redondance sur l’analyse d’un questionnaire hétérogène”; in *CAD*, Vol.XIX, n°3, pp.335-342; (1994).

[QUAL. QUANT.] : J.-P. BENZÉCRI : “Qualité et quantité dans la tradition des philosophes et en analyse des données”; in *CAD*, Vol.XIII, n°1, pp.131-152; (1988).