



Jacques JENNY - Sociologue retraité (ancien de l'Iresco-CNRS)

**Communication pour le 1^{er} Congrès de l'A.F.S.
(Association Française de Sociologie)
Villetaneuse - 25 février 2004**

**Groupe RTF-20 (Méthodes)
Session n° 4 (Comparaisons de méthodes et de modèles)**

«Quanti / Quali» = distinction artificielle, fallacieuse et stérile !

RÉSUMÉ

La thèse que je voudrais ici exposer et argumenter, c'est que non seulement les deux grands «genres méthodologiques» dénommés habituellement «Qualitatif» et «Quantitatif» sont nécessairement complémentaires – banalité largement partagée – mais surtout que leur distinction même est foncièrement artificielle, fallacieuse et par conséquent stérile et contre-productive.

*Cette division classique du travail méthodologique procède d'une **illusion d'optique**, provoquée par une **perception tronquée** de ce que sont réellement, concrètement, d'une part les pratiques et méthodes dites «quantitatives» et d'autre part les pratiques et méthodes dites «qualitatives».*

*Déjà, **la frontière** entre les deux n'est pas toujours placée aux mêmes endroits, selon les points de vue et selon les auteurs.*

*Et que dire des **espèces «hybrides»** ? où classer par exemple les «**statistiques textuelles**» et autres lexicométries, qui font entrer le langage dans le camp des mathématiques ?*

*Ce qui mériterait **éventuellement** les qualificatifs distincts de Quantitatif et Qualitatif, ce ne sont pas des Méthodes mais ce qu'on pourrait appeler les «**plate-formes techniques**» de ces méthodes. Car il est exact qu'on se trouve confronté, dans un cas, à des répartitions numériques et, dans l'autre, à des énoncés langagiers – et que ces deux **matériaux** ont des structures et des contraintes spécifiques telles que leur analyse, leur interprétation, exige la discipline de spécialités pertinentes et performantes : disons pour simplifier, respectivement les mathématiques, les statistiques et les sciences du langage, la sociolinguistique.*

Mais les méthodes, elles, ne peuvent pas se réduire à ces disciplines spécifiques :

- les traitements mathé-statistiques doivent impérativement prendre en compte, intégrer, les significations précises et circonstanciées des catégories de classement des objets dénombrés, ordonnés ou mesurés et, au-delà, les discours dans lesquels ces catégories prennent sens, sous peine de n'être que des «exercices de calcul»,

- et les analyses discursives de corpus textuels ne peuvent négliger ni les «opérateurs de quantification» que contient tout énoncé ni les formes de répartition spatio-temporelle de leurs éléments constitutifs, qui contribuent à leurs significations, sous peine de n'être que des «exercices de littérature».

la double culture méthodologique = handicap ou métissage fécond ?

Nous sommes peu nombreux, parmi les sociologues, à participer simultanément des deux «communautés et cultures méthodologiques» hermétiquement séparées que l'on reconnaît habituellement sous les étiquettes commodes du «Quantitatif» OU du «Qualitatif» [1]. Je suis de ceux-là, bien que mes travaux et centres d'intérêt ont penché d'abord et plus longtemps du côté Quanti. avant de pencher du côté Quali. mais sans jamais hiérarchiser l'un par rapport à l'autre - malgré la pression exercée par les pratiques dominantes du traitement mathématique et statistique, souvent érigé en modèle idéal de «scientificité». Je revendique cette double culture méthodologique, sous le contrôle d'une épistémologie exigeante, pour m'autoriser à interroger les fondements de cette division du travail sociologique, que je trouve artificielle, fallacieuse et par conséquent contre-productive.

qualitatif = signe de qualité ou mode mineur (non-quantitatif, passage obligé) ?

Dans le contexte de la méthodologie sociologique les usages de l'adjectif «Qualitatif» ne renvoient pas tant aux valeurs positives des «qualités» (opposées par exemple à «défauts») qu'à l'«absence de quantité» : l'adjectif négatif «non-numerical» en est d'ailleurs parfois, en anglais, un synonyme explicite. Autrement dit, dans le couple Quali/Quanti, c'est bien l'univers des nombres (le Quanti...) qui constitue la référence implicite, le pôle dominant par rapport auquel l'autre terme est «marqué», défini «par défaut», comme un mode d'analyse mineur - avec une connotation sinon négative du moins un brin péjorative, dévalorisante.

En outre, dans les pratiques courantes, le qualitatif est souvent considéré comme une étape préalable, «exploratoire» (plutôt inductive) et provisoire, du recueil de matériaux bruts, non chiffrés, destinés à se transformer en produits finis, d'ordre quantitatif, considérés comme supérieurs d'un certain point de vue «scientifique», notamment lorsqu'il s'agit de l'étape dite «confirmatoire» (plutôt déductive) de l'administration de la preuve.

Et il existe aussi une hiérarchie manifeste à l'intérieur de cet «ordre quantitatif», comme si la plus-value de la quantification connaissait elle-même des degrés.

Heureusement la complémentarité de ces deux démarches est de mieux en mieux reconnue, non seulement sans préférence marquée a priori en faveur de l'une ou de l'autre

mais en insistant sur la "plus-value" apportée par la conjonction, le couplage, des deux méthodes - considérées de plus en plus comme également utiles, voire nécessaires [2]. Par rapport à ce nouveau consensus «œcuménique», je voudrais ici attirer l'attention sur le fait que, non seulement ces deux types de méthodes sont complémentaires mais encore qu'il ne devrait même pas y avoir lieu de les distinguer car **en fait**, si l'on y regarde de plus près, on s'aperçoit qu'aucune des deux ne peut fonctionner sans intégrer des éléments substantiels de l'autre - sauf à ne considérer que des caricatures tronquées de leur fonctionnement **réel**, ou à extraire de ces méthodologies ce que j'appellerais leurs «plate-formes techniques» respectives.

Il ne s'agit donc pas pour moi de dire que «cette distinction *devrait être abolie*» (jugement de valeur) mais, plus fondamentalement, de dire que «en réalité, ça ne fonctionne pas ainsi» : la distinction n'est pas tant arbitraire que fallacieuse, elle ne procède que d'un effet d'optique, d'une méconnaissance des conditions réelles (jugement de réalité) dans lesquelles les objets de recherche sont construits, traités, analysés, interprétés.

au sein du quanti lui-même, il existe plusieurs modalités de description et construction mathématiques ...

Certains ouvrages classiques de la méthodologie statistique ont théorisé cette hiérarchie en termes de «niveaux de mesure» [3] au sein d'une graduation dont le premier degré se distingue nettement des deux autres, qui pourraient (de notre point de vue de sociologues) être regroupés.

- le premier degré correspond aux procédures de dénombrement, par agrégation [4] d'objets en *classes d'équivalence*, en principe et *a priori non ordonnées*,

- le deuxième aux procédures de classement hiérarchique, par attribution de *rangs* ordonnés, sans mesure d'intervalles,
- et le troisième aux procédures de *mesure proprement dite*, impliquant une métrique forte capable de comparer les intervalles entre les différents échelons des échelles, avec une «efficacité» (calculée en termes de gain ou de perte) supposée croissante du premier au troisième degré.

Pour bien signifier l'homogénéité du modèle logique sous-jacent, unifié sous le *paradigme de la variation-dispersion*, certains auteurs parlent d'«échelles nominales» [5] pour désigner le premier degré (plus fréquent en sociologie, par ex. les fameuses C.S.P.), de quasi-échelles ou échelles des rangs pour désigner le deuxième (fréquent en psychologie sociale, par ex. les fameuses échelles d'opinion, mais aussi en sociologie avec les fameuses «échelles sociales» en termes de niveaux) et d'échelles métriques stricto sensu pour désigner le troisième degré (surtout pratiqué en économétrie pour analyser les échanges monétaires). Et certains manuels de méthodologie n'hésitent pas à recommander aux chercheurs de construire leurs «données» autant que possible [6] selon les «niveaux» déclarés «supérieurs».

... habituellement hiérarchisées alors qu'elles sont équivalentes, selon le même paradigme fondamental sous-jacent, celui de la variété, de la dispersion ?

Pour ma part, je préfère désigner ces trois «modalités de description–construction mathématique de séries d'objets» par les trois lettres grecques contiguës suivantes, pour manifester la cohérence de l'ensemble sans a priori de hiérarchisation :

- pour les répartitions de simple dénombrement, avec calculs d'«entropie» (ou variété, ou dispersion) sur les fréquences proportionnelles = «statistiques de type *Upsilon*» (avec son prolongement probabiliste des «Rapports de Vraisemblance Lambda», équivalents du Khideux classique),
- pour les répartitions ordonnées, avec calculs sur les rangs relatifs de performances plus ou moins regroupées = «statistiques de type *Tau*» (en référence aux coefficients Tau de Kendall, plus performants que les coefficients rho de Spearman) et,
- pour les échelles métriques stricto sensu, avec calculs sur les valeurs de scores plus ou moins dispersés = «statistiques classiques de type *Sigma*», fondées sur le calcul de l'écart-type Sigma (racine carrée de la variance) et sur les formules de décomposition de la variance, statistiques dites «paramétriques» par excellence [7].

les statistiques par dénombrements et calculs sur les fréquences = du quali pour certains (en sciences de la nature) et du quanti pour d'autres (en sciences sociales)

Ainsi donc la frontière entre les différents *types de construction des objets* de la recherche sociologique (pour éviter les termes impropres de «niveaux» de mesure et de «données») n'est pas aussi simple à tracer qu'il y paraît en première analyse. Par exemple, le «quanti» se limite-t-il aux seules échelles de mesure (ou de rang) stricto sensu comme dans la plupart des sciences de la nature, là où «ça» ne parle pas [8] - ou englobe-t-il aussi (comme le conçoivent la majorité des sociologues) les catégorisations, typologies et classifications, ces «échelles nominales» dont la «métrique» est qualifiée de «dégradée», appauvrie, par la plupart des statisticiens classiques – le «quali» désignant en fait cette spécificité des objets de recherche humains et sociaux que sont paroles, discours, textes, et autres signes et symboles de «représentation cognitive» et de «charge affective» [9] ?

pour nous, sociologues : d'abord, choisir les méthodes et techniques de calcul les mieux adaptées à nos options et présupposés épistémologiques,

....

A ce sujet d'ailleurs, je rappelle qu'il existe en fait des méthodes de calcul méconnues qui réduisent, voire annulent, les soi-disant différences d'efficience entre le «degré inférieur» de mesure et les deux autres. C'est la réalité mathématique simple, non sophistiquée, des transformations logarithmiques qui permet ce rapprochement, notamment grâce à une nouvelle branche de statistiques importée des Etats-Unis (appelée là-bas «uncertainty analysis» et souvent traduite ici par «analyse de l'incertitude») au début des années 60 par les psychologues statisticiens français Favrege et Bacher : réalité particulièrement intéressante à exploiter pour nous les sociologues, lorsque nous travaillons sur des répartitions de fréquences d'individus agrégés en classes permutable, en indicateurs de classifications a priori non ordonnées.

C'est à partir des formules de ce nouvel outil statistique que j'ai développé dès 1965 une méthode originale, baptisée Analyse Structurale des Interférences (A.S.I.), pour donner à la recherche sociologique (à commencer par mes propres travaux empiriques) une méthode d'analyse statistique mieux adaptée à mes/nos conceptions

[10]

épistémologiques d'alors . On pourra consulter l'exposé de cette méthode sur mon site Web personnel.

... sans négliger pour autant les autres «outils de la pensée»

Quant à ces options épistémologiques initiales, qui se sont progressivement enrichies par mes propres expériences de recherche et par la fréquentation de quelques bons auteurs, souvent parmi les plus méconnus ou sous-estimés, il n'est malheureusement pas question non plus de les développer ici mais seulement de rappeler quelques-unes

[11]

des notions-clés qui me paraissent pertinentes par rapport à l'objet de la présente communication .

C'est sous la même expression générale «d'*outils de la pensée*» que j'ai pris l'habitude de désigner aussi bien ces outils matériels et logiciels évoqués plus haut que les conceptions fondamentales, présupposés et paradigmes, concepts et notions, qui structurent souvent à notre insu nos activités et projets intellectuels.

Quelques notions-clés pour critiquer les fondements de la distinction quali/ quanti, à commencer ...

... par la manière dont les chercheurs assument ou non leurs «prénotions subjectives» ou pratiquent l'illusoire «tabula rasa» de J.P. Benzécri, ...

[12]

Ce paragraphe reprend des extraits d'un article que j'ai consacré à une critique du postulat benzécriste, si dominant en France qu'on ne perçoit plus ce qu'il a de relatif, de non-évident, et qu'on en viendrait à oublier les méthodes traditionnelles d'une observation quasi-expérimentale de la réalité sociale, instruite par des paradigmes explicités et des «prénotions» consciemment assumées - dans une perspective résolument constructiviste.

[13]

Rappelons les termes de ce postulat : «il existe **UNE structure DU réel ...**» [que «l'*Analyse des Données*» va pouvoir dévoiler au chercheur qui aura fait *table rase de ses prénotions naïves*]. Cette méthode opère par algorithmes de calcul matriciel qui décomposent automatiquement de grands tableaux à double entrée (du type 'Individus x Caractères descripteurs') en «facteurs additifs» hiérarchisés – ce qui lui vaut d'être rangée dans la catégorie des méthodes «fréquentistes», ou d'analyse «spectrale» (par analogie avec l'effet du prisme sur la lumière blanche).

Dans les débats méthodologiques, surtout franco-français, la distinction fondamentale entre les conceptions des analyses statistiques dites "fréquentiste" d'une part et

[14]

"intuitionniste" d'autre part n'est pratiquement jamais problématisée, tant est puissant l'effet de dominance et de pseudo-évidence dont jouit la conception fréquentiste. Et elle est souvent masquée, parfois même brouillée, par des distinctions secondaires d'ordre technique, portant par exemple sur le choix de telle ou telle métrique au sein d'une même famille de méthodes.

Or cette distinction est capitale parce qu'elle concerne la manière de traiter les relations d'interaction réciproque entre les "prénotions" des chercheurs et les "notions" produites par la recherche. Il s'agit ni plus ni moins des "conceptions gnoséologiques" auxquelles on se réfère (consciemment ou non) dans nos façons de produire collectivement la connaissance du réel : avons-nous bien entendu les mises en garde des phénoménologues contre le *réalisme naïf*, contre le "préjugé du monde objectif" ? Un tel enjeu épistémologique ne mérite-t-il pas une réflexion vigilante ?

Au lieu de postuler l'illusoire "tabula rasa" et de s'en remettre aveuglément aux résultats des calculs (par ailleurs géniaux) de la méthode abusivement appelée **au singulier**, l'«Analyse des données», l'approche intuitionniste oblige le chercheur à expliciter sa représentation du phénomène étudié, voire ses modèles d'organisation de la complexité, en fonction des grilles de lecture et des systèmes d'interprétation du "réel" qu'il ne peut pas ne pas avoir (même s'il se refuse à en prendre conscience). Au lieu de lui fournir **une** représentation graphique préfabriquée, unidimensionnelle (le dendogramme de la C.A.H. par ex.) ou spectrale (les diagrammes plurifactoriels de l'A.F.C.) ou réductionniste (le tableau de Burt, par ex.), l'approche intuitionniste oblige le chercheur à investir ses connaissances nuancées du domaine étudié dans une série de lectures **plurielles** de la réalité, à expérimenter en fonction des multiples points de vue que sa problématique de recherche l'invite à adopter.

... par le postulat des complexités structurelle et structurale ...

C'est principalement à partir de la *notion de complexité* et des procédures de son *traitement* que s'échafaudent la plupart de mes propositions et conceptions méthodologiques dont certaines sont des «trouvailles» à proprement parler alors que d'autres ne sont que des applications de principes généraux consensuels, notamment s'agissant du principe trivial que les phénomènes, tant humains et sociaux que naturels, possèdent tous la propriété d'être structurés ^[15] - postulat impliquant que complexité rime mieux avec complicité qu'avec complication.

Dans plusieurs notes internes adressées aux instances dirigeantes de mes laboratoires et instituts de rattachement, depuis 1985 environ, le *traitement de la complexité* était explicitement inscrit comme objectif susceptible de mobiliser et regrouper les énergies dans des opérations de recherche transversales. Et ces différents projets successifs visaient aussi bien le domaine des «répartitions numériques» que celui des «corpus textuels» (ou vice-versa). Tout en étant alors bien distingués comme deux domaines méthodologiques complémentaires, que je me refusais déjà d'appeler Quantitatif et Qualitatif, il était postulé que ces deux domaines relèvent des mêmes exigences fondamentales, notamment celle de respecter la complexité intrinsèque, constitutive et non seulement construite ou artificielle, des objets sociologiques - quels qu'ils soient.

La complexité se décline, selon moi, en deux grandes modalités complémentaires - que j'appelle *complexité structurelle* (ou combinatoires «intra-objet») et *complexité structurale* (ou correspondances «inter-objets»), entendues respectivement comme,

- pour la structurelle : complexité caractéristique d'objets *relativement autonomes*, composés de multiples «facettes» combinées ^[16], comme par ex. les catégories et niveaux du système de division du travail socio-professionnel, les filières et niveaux de l'institution scolaire, mais aussi comme certaines structures sémantiques fondamentales, qu'on a tendance à oublier ou à méconnaître malgré leur fécondité heuristique, telles que les organisations conceptuelles des systèmes d'objets selon le double axe Syntagmatique *

Paradigmatique de Ferdinand de Saussure et Jakobson, repris par Roland Barthes ^[17] - ou encore l'organisation logique des oppositions conceptuelles, en forme hexagonale selon Robert Blanché ^[18], et ...

- pour la structurale : complexité caractéristique de relations de *type quasi-causal* observables entre les composantes de réalités *relativement hétéronomes* à un «niveau de structuration supérieur», évoluant dans le temps et différenciés dans l'espace des sociétés dites «globales», comme par ex. les imbrications mutuelles des positions sociales des parents et des positions scolaires des enfants et des positions sociales futures de ces mêmes enfants - ce qui peut définir précisément les trois dimensions d'analyse d'une table de mobilité sociale intergénérationnelle.

... et par le principe de la dialectique (chinoise ?), ou «dialogicité» anti-cartésienne, qui exige un «Nouveau Discours de la Méthode» ...

Tout aussi fondamental pour justifier ma critique de la distinction «Quali/Quanti», est le principe selon lequel les éléments constitutifs d'une réalité sociale complexe ou d'un phénomène social complexe ne sont *pas isolables*, même pour les besoins de l'«analyse» - comme le voudrait une théorie des ensembles simpliste. Les meilleures images que je puisse en donner ici, pour éviter de trop longs développements, est l'image connue de l'«entrelacement du Yin et du Yang», qui ne saurait se confondre avec l'image de l'«intersection booléenne de deux sous-ensembles A et B», ou encore le test de la «figure ambiguë» du psychologue suédois Boëring, qui ne saurait se confondre avec le visage à double face de Janus.

[19]

Les linguistes évoqueraient ici probablement le grand Bakhtine, dont la philosophie du langage accorde une place éminente au paradigme du *dialogisme généralisé*, selon lequel toute expression verbale est hétérogène, tout énoncé est «altéré» par la présence, au moins potentielle ou virtuelle, d'autres locuteurs, et une conception «*polyphonique*» du sens, reprise par Ducrot.

En vertu de ces principes de la complexité et de la dialogicité, la Méthode cartésienne du découpage de la réalité en éléments aussi simples que possible est inappropriée, au moins pour nos disciplines de sciences humaines et sociales. C'est pourquoi je propose de lui substituer un «Nouveau Discours de la Méthode», projet méthodologique audacieux et risqué (car confronté à d'insurmontables obstacles intellectuels et pratiques).

Il s'agit en somme de concevoir une méthodologie conforme au principe suivant de "**complexité optimum**", qui appellerait évidemment de longs commentaires épistémologiques et théoriques par rapport au fameux Discours de la Méthode de Descartes :

"construire, enregistrer et traiter les informations sous la forme (dans le 'format') **la plus signifiante pour la recherche**, c'est-à-dire à leur **'niveau de complexité' le plus élevé possible**"

... et une typologie plus élaborée des «formats de questionnements», avec la règle impérative du respect du couple indissociable «Question(s)-Réponse(s)»

S'agissant particulièrement des informations recueillies à l'aide de questionnaires ou d'entretiens d'enquêtes sociologiques, ces principes abstraits peuvent se concrétiser par une diversification et une explicitation des principaux types formels de questions que toute recherche peut être amenée à poser à ses informateurs [20], au lieu et place de la distinction pragmatique usuelle entre questions ouvertes et fermées, réplique de la distinction tranchée établie entre conversations ordinaires et questionnaires.

A côté des banales conversations familières, des confidences, des sondages d'opinion et des interrogatoires judiciaires, quelle place occupent nos pratiques et nos "dispositifs de questionnement standardisé" prétendument rationnels ? Quel "*contrat de communication*" [21] s'institue entre ceux qui parlent et ceux qui font parler ? Peut-on esquisser une théorie générale de ces différents dispositifs sociocognitifs et pratiques sociodiscursives, qui concourent à la production d'une multitude variée d'actes langagiers.

d'énonciations discursives, et qui participent à leur manière à la production et à la circulation des discours sociaux, des débats publics, thème de recherche spécifiquement sociologique s'il en est ?

N'ayant encore que peu publié sur ces sujets, on pourra trouver sur mon site Web, quand il sera mis à jour, un bref aperçu sur ce paradigme de la circulation des discours et sur ma typologie tridimensionnelle des «instances de pratiques discursives». Mais il convient de mentionner ici la place éminente que prennent parfois des résultats d'enquête sociologiques ou de sondages d'opinion lorsqu'ils sont largement médiatisés, et simplifiés de manière caricaturale : il leur arrive de structurer les débats publics, tout comme

[22]

les conversations privées, au-delà même de ce qui serait légitime d'un point de vue déontologique .

De manière générale, qu'il s'agisse de questionnaires (fermés ou ouverts) ou d'entretiens (directifs ou non), la règle d'or impérative serait de ne jamais découpler le «Réponses», quel que soit leur «format», du contexte d'énonciation et des Questions qui les ont suscitées, au moins implicitement (même s'il n'y a pas eu d'énoncés interrogatifs dans le contexte immédiat). Autrement dit, ne jamais analyser les informations en les abstrayant du «dispositif de questionnement» qui les a provoquées sollicitées, voire «formatées».

Ce faisant, non seulement on s'éloigne aux antipodes de la conception classique des «données» (supposées préexister, déjà là) et a fortiori des pratiques usuelles consistant à analyser les informations recueillies en les réduisant à de simples codes chiffrés de nomenclatures, sans rappeler et analyser avec autant de rigueur le contexte de la relation d'entretien, le dispositif concret de questionnement, le système formé par l'ensemble des questions, avec leur chronologie et leur dynamique, les énoncés précis de chacun

[23]

des questions , et bien entendu leurs connotations, les implicites émotionnels, cognitifs et axiologiques qu'elles charrient nécessairement.

... sans oublier d'explicitier la problématique des temporalités et les enjeux afférents, notamment selon le paradigme ontogénétique de Gilbert Simondon

Comment ne pas évoquer ici tout d'abord la théorie ontogénétique de "l'individuation transductive" de Gilbert Simondon, qui ne vise pas moins qu'à renverser les paradigmes classiques du changement - c'est-à-dire à poser les processus de transformation diachronique, d'avènement de nouvelles formes et structures, comme premiers par rapport aux états "individus", accomplis, advenus, des êtres et des choses, à penser *les potentiels de la métastabilité* comme le mode d'être de référence, par rapport auquel les états

[24]

stables et instables ne sont que provisoires et conjoncturels. Je me permets de renvoyer ici à un article récent de vulgarisation sur cette révolution paradigmatique reconnue dans les sciences exactes mais dont on est loin d'avoir épuisé toutes les potentialités en sciences sociales.

Pour ce qui concerne notre critique de la distinction Quali/ Quanti, je ne retiendrai de ce paradigme simondonien que la nécessité du principe de précaution suivant, valable pour toutes les formes de questionnement et de traitement des réponses - avec cependant une vigilance accrue pour les méthodes préformatées, figées, contraignantes, les codifications rigides, voire binarisées, les interprétations simplistes, superficielles, «au premier degré» :

mis à part pour quelques rares informations à caractère objectif incontestable (par ex. les renseignements d'état-civil, dans les sociétés bien «administrées»), ne jamais perdre de vue que les réponses obtenues :

- d'une part, peuvent être des expressions nuancées par le conditionnel (du genre «ça dépend ...»), fluctuantes, voire labiles, voire ambiguës, dont la signification sera d'autant plus contestable qu'elles auront été soumises à des traitements réductionnistes, unidimensionnels, tranchants, à des indexations forcées,

[25]

d'autre part, qu'elles peuvent masquer des pensées indécises, contradictoires, des auto-censures - tous symptômes révélateurs de ce qu'il peut y avoir de métastabilité, de potentiel, de «jeu» (au double sens du mot) dans les pratiques et les pensées humaines.

Transition : entre ma focalisation sur les Répartitions Numériques et mon intérêt croissant pour les Corpus Textuels, où en étais-je ?

Tout au long de ma période de pratiques méthodologiques à dominante statistique, ma curiosité décloisonnée n'a pas cessé de s'intéresser à ce que je ressentais confusément comme une exigence d'équilibre intellectuel, voire de salubrité cognitive, c'est-à-dire au domaine des productions verbales, qu'on les appelle actes langagiers, pratiques ou énonciations discursives, en y faisant des incursions de plus en plus fréquentes.

N'est-il pas banal de dire qu'on ne marche jamais si bien que sur ses deux jambes ?

Etant un des rares chercheurs à fréquenter ces deux univers académiques disjoints, j'étais d'autant plus sensible à tout indice de possibles convergences, voire de possibles fécondations mutuelles. Il y a très longtemps, j'avais fait partie d'un groupe informel de chercheurs sociologues et linguistes, qu'on appelait ironiquement L.S.D. (= Laboratoire de Sociologie de la Dominance) et dont certains membres furent parmi les créateurs de la revue *Langage et Société*. Nos discussions interdisciplinaires, qui étaient passionnantes, ont certainement contribué au moins à initier les sociologues dont j'étais aux charmes discrets de la sociolinguistique naissante et ont probablement contribué à assurer les quelques jeunes linguistes du groupe dans leur projet de fonder la nouvelle discipline socio-linguistique (ou sociolinguistique, sans trait d'union, pour les plus ambitieux et les plus novateurs).

Quelques années plus tard, c'est à la forte personnalité de Pierre Achard, avec le séminaire «Analyse de Discours» que je fréquentais régulièrement et à la jeune revue «Langage et Société» que je lisais de temps en temps, que je dois en particulier de m'être détaché progressivement de quelques postures méthodologiques issues de la (dé)formation sociologique traditionnelle, qui constituaient en fait un obstacle à la pénétration de notions sociolinguistiques importantes, et fécondes pour la sociologie elle-même. La principale de ces postures consistait à ne considérer les textes (écrits ou parlés) que comme supports d'informations, justiciables de ce qu'on appelle (encore) l'«Analyses de Contenu» (A.C.) – notamment avec le classique protocole méthodologique de Berelson - posture assez proche de celle du «paradigme des mots-clés» en recherche documentaire.

La fréquentation des sociolinguistes nous a fait découvrir les richesses et subtilités des différents courants de l'école dite «Analyse de Discours (A.D.) à la française», et tout ce qu'elle recèle de propositions originales susceptibles d'enrichir notre méthodologie de sociologues, avec le risque (évité ?) de tomber dans la marmite sociolinguistique et d'en oublier les spécificités de notre discipline.

L'aventure d'une contribution rédactionnelle inattendue : «quelles sont les pratiques «CAQDAS» des sociologues français» ?

J'en étais là, il y a maintenant huit ans, lorsqu'il m'arriva l'aventure suivante :

Une jeune collègue brésilienne, Wilma Mangabeira, chargée d'organiser un n° spécial de la revue de l'A.I.S. «Current Sociology» consacré aux pratiques de CAQDAS (= Computer Assisted Qualitative Data Analysis Software) dans plusieurs pays européens et américains, et ne trouvant pas de correspondants en France pour rédiger la contribution française de ce numéro spécial, fut par hasard mise en contact avec moi, précisément parce que je pratiquais (un peu) la micro-informatique et que je manifestais un intérêt certain pour les analyses textuelles.

Bien que ne sachant même pas alors ce que signifiait ce sigle CAQDAS, je me suis trouvé embarqué dans cette aventure. Et je ne le regrette pas, car elle m'a permis de découvrir cet univers opaque et passionnant des concepteurs de logiciels : bricoleurs (comme moi) ou pas mais souvent géniaux, parfois méconnus ou méprisés par les autorités mandarino-bureaucratiques (dont j'ai moi-même éprouvé l'inertie) mais persévérants dans l'adversité.

Je peux avouer maintenant que j'ai accepté la proposition de Wilma Mangabeira un peu comme un défi lancé à moi-même. En effet, autant j'avais des biscuits pour valider mes postulats et présupposés fondamentaux (complexités structurelle et structurale, dialogicité, option intuitionniste, paradigme ontogénétique, etc...) dans le domaine bien balisé et arpenté des analyses à dominante statistique, autant la statistique textuelle et particulièrement la lexicométrie à la française me laissaient perplexe, ainsi que beaucoup de sociologues étrangers. Comme je n'avais aucune pratique suivie et des informations très lacunaires dans ce domaine des analyses textuelles, il m'intéressait d'aller y voir de plus près : mes postulats et présupposés, mes paradigmes de référence fondamentaux, si péniblement élaborés, allaient-ils pouvoir s'appliquer aussi dans ce nouveau

domaine ?

Ma dénonciation intuitive de la distinction Quali/ Quanti se précise et se renforce au contact des pratiques d'analyses textuelles informatisées «à la française»

N'ayant a priori aucune affinité ni avec les présupposés technicistes de cette mission rédactionnelle (pourquoi se limiter aux techniques d'analyses assistées par ordinateur, avant même d'analyser les méthodologies sous-jacentes ?) ni avec la distinction Quali/ Quanti explicitement inscrite dans ce projet, j'ai dû argumenter pour qu'il n'y ait pas

d'ambiguïté sur le sens de ma participation à ce numéro spécial [\[26\]](#).

Je connaissais déjà la spécificité franco-française des logiciels lexicométriques, qui dominent largement les (rares) pratiques de recherche sociologique sur corpus assistées par ordinateur. Il était donc inconcevable de ne pas parler des outils de «statistique textuelle» sous prétexte qu'on ne devait traiter que des «données qualitatives».

D'ailleurs, sur le fond du problème, peut-on se désintéresser des répartitions chiffrées des phénomènes et objets discursifs, faits langagiers ou "actes de parole", de leurs fréquences d'occurrence, concordance ou co-occurrence, ce que l'on a coutume de désigner, en Europe tout au moins, sous l'expression de *statistique textuelle* (Lebart et Salem, 1994), ou d'*analyse statistique des données textuelles* ?. Ce courant est né, vers 1980, de l'application des nouvelles méthodes d'Analyse des Données dont on a déjà parlé plus haut à des corpus textuels qui, au début, étaient surtout des écrits littéraires, politiques ou religieux. Mais la statistique textuelle peut prendre d'autres formes mathématiques, comme en attestent les JADT (Journées d'Analyse des Données Textuelles) bisannuelles.

En janvier 2002 j'ai eu l'occasion de participer à une journée de débats organisée par et pour des linguistes de Paris-3 et Paris-13, sur le thème de «la pertinence du chiffre dans le traitement des données langagières». J'ai pu y retrouver certaines de mes conceptions d'alors, notamment concernant «les deux cultures de la quantité et de la qualité dans les sciences sociales», considérées comme «deux aspects de tout phénomène langagier», le pôle quantité étant celui des «données stabilisées, travaillées en laboratoire» et le pôle qualité celui des «données attestées sur le terrain, situées, en contexte» (Patrick Renaud).

Cette journée m'a conforté dans le projet d'aller plus loin en reconnaissant le caractère fallacieux, illusoire, de cette distinction elle-même - ce qui implique qu'il faut expliciter nos présupposés épistémologiques, en amont des lieux communs de la pensée unique.

Que penser des «espèces hybrides» que sont les statistiques textuelles et le «pseudo-langage» (combinaisons de codes alphanumériques successifs) ?

J'ai déjà évoqué la difficulté de classer les «statistiques textuelles», les méthodes lexicométriques, selon la distinction Quali/ Quanti. Peut-on dire qu'il s'agit d'une intrusion, voire d'une contamination, des mathématiques dans le camp du langage, ou l'inverse ? ou n'est-ce qu'une métaphore ?

De même, lorsqu'on pratique des statistiques complexes, par exemple en construisant des indicateurs empiriques par combinaison de codes alphanumériques (comme je l'ai fait souvent, en reprenant le terme classique de «patrons de réponses» ou de séries de réponses à des «batteries de questions»), on est frappé par l'analogie formelle entre ces successions de lettres (de code) et des mots d'un lexique. S'agissant par exemple d'enquêtes biographiques, la succession des positions de chaque enquêté au cours d'une séquence temporelle peut s'exprimer en juxtapositions d'états codifiés - qu'on peut lire *comme des mots* (ou comme des phrases sous certaines conditions) qui raconteraient, résumeraient, cette séquence biographique, au point que certains chercheurs en parlent comme de «récits mathématiques» qui, à la longue, "parlent" aux

enquêteurs comme de véritables signes articulés [\[27\]](#). Peut-on dire pour autant que ce «pseudo-langage» signifie une intrusion des entités linguistiques dans le camp des codes de chiffrage et des calculs, ou l'inverse ? ou n'est-ce qu'une métaphore ?

Dans toute expression langagière les dimensions dites quantitative et qualitative sont indissociables, intrinsèquement associées l'une à l'autre

A mon avis les espèces qualifiées ci-dessus d'hybrides ne sont pas des métaphores, mais des cas-limites qui illustrent au contraire ce fait fondamental qu'il ne saurait y avoir, dans toute information, de dimension quantitative sans dimension qualitative intrinsèquement associée et réciproquement.

Lorsqu'on sollicite ou observe des réponses textuelles en langage naturel, l'affaire est entendue : qu'elles aient été provoquées par des questions ouvertes dans un questionnaire directif ou par de longs entretiens plus ou moins non-directifs (comme d'ailleurs dans tout corpus textuel), ou prononcées dans des conversations spontanées, on peut constater que la quantité et le codage s'y trouvent toujours présents à leur manière.

La *quantité* s'y manifeste par nécessité grammaticale, surtout par ce qu'on appelle les "**opérateurs de quantification**" (le singulier de l'unique et le singulier du générique, le pluriel défini de la totalité ou le pluriel indéfini des proportions intermédiaires entre le Tout et le Rien), les **opérateurs de comparaison** (égalité ou différences en + ou en -) et **d'évaluation** (trop ou pas assez ou juste bien comme ça).

Moins nécessairement que la quantité, le **codage est plus ou moins virtuellement présent dans tout énoncé textuel**, de manière implicite - si l'on veut bien considérer que beaucoup de formulations que l'on croit originales, y compris par exemple dans les récits de vie les plus singuliers, peuvent trouver des échos, des résonances, dans des bribes de discours sociaux stéréotypées ou «topoï» (par exemple, expressions proverbiales, aphorismes, locutions toutes faites, slogans, citations...). A moins de ne partager aucune de ces conventions d'usages linguistiques ou de ne rien connaître de ces «arrière-plans doxiques» des textes, qui constituent la culture commune de tout groupe social les chercheurs ne peuvent pas ne pas rattacher plus ou moins nettement ce qu'ils entendent (ou lisent) à ce qu'ils ont déjà entendu (ou lu), à ce qu'ils savent qu'ils pourraient entendre (ou lire). C'est ainsi que la communication est possible, du moins entre ceux qui participent des mêmes cultures de groupe.

A tout bien considérer, au moins *dans une première approximation* lexicographique, il n'y a guère solution de continuité entre d'une part des énoncés en langage naturel composés d'assemblages de mots, ces signifiants inscrits dans les dictionnaires de la langue - et d'autre part des configurations de "patrons de réponses codées" - composées de séquences de signifiants inscrits dans les instructions de codage, ce «pseudo-langage» qu'on évoquait ci-dessus dans la catégorie des espèces hybrides ?

Il est vrai cependant que le sens du langage naturel, et a fortiori celui du pseudo-langage des codes de chiffrement, ne se trouve pas entièrement dans les simples juxtapositions de signes, comme l'a si bien montré Benveniste, commenté dans un article récent de *Langage et Société* ^[28], ... ce qui constitue pour moi une validation supplémentaire du postulat de la complexité structurelle et plaide contre les abus de ce qu'on appelle «l'entrée lexicale», utilisée dans la plupart des logiciels d'analyses textuelles ^[29], là où on pourrait s'attendre à utiliser au moins les syntagmes nominaux.

Contre les abus de l'«entrée lexicale», ne pourrait-on pas investir davantage les acquis de la sociolinguistique dans les méthodes d'analyse de corpus textuels ?

De même, la commodité informatique veut que, pour trier les occurrences lexicales d'un corpus, on se contente le plus souvent des seuls tris «qui ne posent apparemment pas de problème», à savoir selon l'ordre alphabétique et selon l'ordre de fréquence décroissante. Or l'ordre de tri alphabétique est arbitraire (il n'a comme justification que la logique de consultation des dictionnaires) et il est loin d'être anodin : son apparente évidence peut faire obstacle à d'autres structurations plus pertinentes, notamment par catégorisations en types grammaticaux, en domaines ou registres thématiques, en statuts conceptuels dans telle ou telle grille d'analyse théorique (l'équivalent de ce qu'on nomme thesaurus en documentation), etc ... Mais ces structurations posent tellement de problèmes que seules des équipes pluridisciplinaires auraient des chances de pouvoir en venir à bout, et encore ... ?

Certes les types grammaticaux sont distingués dans la plupart des logiciels, mais certains rejettent dans le néant du non-calcul (quitte à les réintégrer en «variables

supplémentaires» après-calcul) tous les mots qui ne sont ni substantifs ni adjectifs ni verbes et adverbes, et qu'on appelle pour cela des «mots-outils» ou mots «vides de sens». Or, quand on connaît la richesse des informations qu'on peut extraire des usages discursifs des déterminants, pronoms personnels, prépositions et conjonctions adversatifs, déictiques, tournures stylistiques et rhétoriques (informations de type sociologique, sur les rapports sociaux qui se construisent et s'actualisent dans les pratiques discursives et langagières), cette pseudo-neutralité méthodologique ne peut s'expliquer que par la crainte d'être contaminé par la subjectivité des chercheurs et/ou par la crainte de ne pouvoir s'acquitter d'opérations aussi complexes et subtiles – car elles le sont, effectivement.

On retrouve là les mêmes choix (ou non-choix) épistémologiques que ceux repérés plus haut s'agissant des répartitions numériques : fréquentisme par parti pris objectivist

[30]

illusoire (le «réalisme naïf») versus intuitionnisme avec l'audace des subjectivités assumées et le goût du risque (bien calculé, évidemment) .

Comme pour les analyses sur répartitions statistiques de populations qui préfèrent faire table rase des prénotions des chercheurs, certaines analyses sur statistiques textuelles préfèrent faire table rase des connaissances acquises en sociolinguistique, sémantique, pragmatique discursive, etc ... Mais ce n'est qu'un choix (inconscient) parmi d'autres. Ce qui me fait penser que ce type de contraste méthodologique, bien réel celui-là, est plus important et lourd de conséquences que la pseudo-distinction entre méthodes Quali et méthodes Quanti. Autrement dit, les mêmes grandes options épistémologiques fondamentales traversent ces deux grandes méthodologies qui occupent le devant de la scène ; mais elles n'ont pas la même visibilité parce que fondées sur des présupposés qui, comme tout présupposé, sont par définition implicites et constituent pour chacun (individu ou groupe) un horizon de pensée et de pratique a priori indépassable.

alors : un faux problème, en langue de bois, qui stérilise la solution à d'autres problèmes de méthode, bien réels ceux-là, ...

L'opposition entre les modalités d'enquêtes «quantitatives» et «qualitatives» a la vie dure, dans l'institution académique comme dans les agences d'«info-com» et de marketing.

On pourrait dire que cette division classique du travail méthodologique procède d'une **illusion d'optique**, provoquée par une **perception tronquée** de ce que sont réellement, concrètement, les pratiques et méthodes d'enquêtes dans ces deux modes opératoires.

Telle qu'elle est généralement énoncée, cette opposition est selon moi un **exemple typique de faux-problème**, survivance d'une langue de bois qui pollue notre réflexion, entretient des divisions artificielles au sein de notre profession et permet aux tenants de chaque camp retranché d'esquiver les nécessaires mises à jour des connaissances théoriques et des techniques de recherche. Cette affirmation exigerait évidemment de longs développements pour ne pas en rester au stade de la provocation gratuite. Faute de temps, je ne peux maintenant qu'énoncer quelques propositions à discuter, en style quasiment télégraphique :

- **bannir le terme galvaudé de "données"** puisque les informations sont toujours co-produites, co-énoncées, dans des procédures discursives plus ou moins complexes, - et traiter ces informations comme des Réponses-à-des-Questions (voire comme «de la» réponse à «de la» question, pour signifier que ces couples Q-R peuvent être de type vague, indéterminé), y compris lorsqu'elles sont énoncées en contexte d'apparent monologue car il y a toujours au moins des questions virtuelles en amont. Ce qui est significatif, porteur de sens et d'information et digne d'interprétation, ce sont les **enchaînements discursifs de couples Questions-Réponses et parfois Réponses-Questions, avec leurs éventuels renvois intertextuels et références pérertextuelles.**

- ne pas réduire les informations sollicitées, dans quelque dispositif Questions-Réponses que ce soit, au "format" des seules réponses détachées de leur contexte, la notion de format étant ici empruntée à l'informatique où l'on distingue les formats numériques, codés alphanumériques, et textuels "en langage naturel". Toute recherche en sciences sociales comporte nécessairement **une part de "matériaux textuels" à analyser** - au point que la distinction entre le "qualitatif" et le "quantitatif" ne saurait être au mieux qu'une distinction de phases, de moments dans la recherche, et au pire qu'une mystification. Mystification destinée peut-être à masquer les méconnaissances respectives de la réalité discursive chez les "quantitativistes" d'une part, et de la réalité numérique chez les "qualitativistes" d'autre part.

Par conséquent, dans le cas général de tout post-codage et a fortiori dans le cas particulier des questions ouvertes numériques, et des questions précodées, les réponses "chiffrées" (au double sens de numériques et de codées par chiffrage) **n'ont de sens qu'en référence aux catégories conceptuelles ainsi "mesurées", voire à leurs connotations idéologiques, ainsi qu'aux énoncés interrogatifs et à tout le "dispositif de questionnement"** qui ont sollicité ces réponses. D'où l'importance d'analyser scrupuleusement au préalable et la forme et le contenu de tout ce dispositif, avec les ressources disponibles de la théorie des catégorisations, de la critique épistémologique, de la sémantique, de la linguistique et de la sociolinguistique, surtout si on est soi-même l'auteur(e) de ces questions.

... qui entretient des préjugés, faux mais tenaces, qu'il faut revisiter, ...

D'autre part, la fermeture des questions d'un questionnaire n'est pas toujours à considérer comme un handicap quant au contenu thématique, opposée à la liberté prétendument laissée aux interviewés. Un chercheur peut à bon droit solliciter aussi bien des réactions aux paradigmes de sa discipline, avec leurs problématisations et leurs catégorisations (ou aux stéréotypes des discours sociaux, par ex. sous le format de question suivant : "On entend dire parfois que [...]; qu'en pensez-vous ?"), que des expressions soi-disant spontanées de ses interlocuteurs : c'est une question de posture générale de recherche, et pas seulement une distinction entre une étape dite "exploratoire" et une étape dite "confirmatoire". Il n'est pas question ici de réexaminer toutes les facettes d'un débat aussi vaste et fondamental. Car un tel réexamen exigerait une relecture critique de nombreux travaux classiques sur les théories et les techniques de l'entretien, comme stratégies de collecte d'informations, modulables selon les différents types d'informations recherchées (notamment leur statut conceptuel dans la problématique de recherche), ce qui devrait plaider pour le mixage de plusieurs modalités discursives au sein d'un même dispositif d'enquête.

Le qualitatif et le quantitatif s'interpénètrent donc si intimement dans tous les types formels de couples Questions-Réponses, bien que selon des modalités propres à leur "format", que cette distinction paraît de plus en plus anachronique. D'ailleurs, si l'on en croit un compte-rendu synthétique de plusieurs études réalisées au Royaume Uni (Halfpenny, 1997) pour analyser "la relation entre recherches sociales quantitatives et qualitatives", il n'y a pas de différence fondamentale entre les résultats des unes et des autres, dans la mesure où chaque méthode s'applique correctement à "reconstituer la richesse du vécu". Contrairement aux préjugés, il semble même n'y avoir **aucun lien entre les paradigmes théoriques dont se réclament les auteurs et les méthodes qu'ils pratiquent pour collecter les informations**. La conclusion de cet article contredit donc les hypothèses selon lesquelles il aurait pu y avoir correspondance entre des présupposés épistémologiques positivistes et le recours à des méthodes "quantitatives" orientées vers de l'*explication*, associées aux techniques de sondage et de traitement statistique, d'un côté, et entre des présupposés phénoménologiques et des méthodes "qualitatives" orientées vers de la *compréhension*, associées à des études de cas monographiques, de l'autre côté. Et l'auteur suggère que les choix méthodologiques dépendent moins de ces présupposés que de certaines circonstances pratiques conjoncturelles comme les thèmes de recherche, les compétences professionnelles acquises et les prédispositions personnelles, les traditions locales, le montant des crédits alloués,... et l'on constate finalement un certain **pluralisme pragmatique, avec recours fréquent au mélange de ces deux types de méthode**.

En guise de conclusion : abandonner cette distinction de méthodes, mais conserver une distinction, plus exigeante, ...

C'est bien une illusion qui peut faire voir et croire les informations recueillies en enquêtes comme «naturelles», déjà là, par conséquent justiciables d'autonomisation, de dissociation par rapport au contexte et à l'intertexte, et finalement susceptibles d'être réduites par le traitement à de simples étiquettes (ces fameux «*labels*» étriqués des têtes de lignes et colonnes de tableaux statistiques, des graphiques et diagrammes en tous genres).

Je propose donc d'abandonner définitivement cette distinction Quali/ Quanti pour une autre, estimée plus pertinente, consistant à évaluer les couples Questions-Réponses de tout dispositif d'enquête, quel qu'il soit, ...

...selon le mélange spécifique de leurs caractéristiques en termes de formes langagières et de contenus thématiques, ...

de ce point de vue, la plupart des questions de questionnaires et de sondages de type classique, y compris les questions ouvertes, comportent des contraintes de forme standardisée qui ne laissent guère de liberté linguistique aux répondants. La *spécificité* des paroles prononcées au cours des entretiens serait donc à chercher **plutôt du côté des "formes" que du "contenu"**. Dans ces paroles, on peut **décrypter "la manière de le dire" autant que "ce qui est dit"**, alors que dans les questionnaires la marge d'initiative stylistique ou rhétorique des enquêtés est beaucoup plus limitée - ce qui plaide pour l'emploi de techniques d'analyse sociolinguistique, notamment **en référence au paradigme de l'énonciation**, d'autant plus élaborées que l'on a affaire à des corpus plus spontanés. Est-il besoin de préciser que l'analyse socio-sémantique classique, qui porte sur le contenu lexical, le plus souvent après élimination des mots-outils soi-disant "vides", reste un complément indispensable de cette approche sociolinguistique ?

Certaines méthodes d'analyse textuelle proposent, par exemple, des procédures d'identification de traits discursifs formels, comme la "rhétorique spontanée" (D'Unrug, 1974), ou comme le "style", révélateurs des rapports d'interlocution mis en scène. De telles perspectives devraient permettre aux chercheurs de s'intéresser dorénavant de plus en plus à ces *"manières de dire"* des locuteurs sollicités, en complément de leur intérêt pour *«ce qu'ils disent»*.

... en termes de «plate-formes Techniques» distinctes (statistiques, sociolinguistiques) «disciplinées», correspondant à des «matériaux» spécifiques différents ...

Ce qui mériterait éventuellement les qualificatifs distincts de Quantitatif et Qualitatif, ce ne sont pas des Méthodes (au sens plein de démarches et de pratiques d'investigation, qui comportent nécessairement des facettes indissociables de l'un et de l'autre) mais ce qu'on pourrait appeler les **«plate-formes techniques»** de ces méthodes. Car il est exact qu'on se trouve confronté, dans un cas, à des répartitions numériques et, dans l'autre, à des énoncés langagiers - et que ces deux matériaux ont des structures et des contraintes spécifiques telles que leur analyse, leur interprétation, exige la *discipline de spécialités pertinentes et performantes* : disons pour simplifier, respectivement les mathématiques et les statistiques d'un côté et les sciences du langage et la sociolinguistique de l'autre.

Pour ne plus parler en termes de qualitatif et de quantitatif, parlons par exemple de «matériaux», langagiers et numériques, d'outils, sociolinguistiques et statistiques, des traitements successifs aux différents moments de la recherche (avec tout ce qu'implique cette petite «révolution paradigmatique») et efforçons-nous, dans un même mouvement de pensée dialectique, de concevoir cette distinction non comme les deux faces d'une même pièce de monnaie, mais comme les composants indissociables de l'alliage où ils sont fondus.

... mais fondus dans un alliage indissoluble, qui mérite le nom de Méthode, avec référence à Canguilhem (entre autres)

Mais les méthodes, elles, ne se réduisent pas à ces disciplines spécifiques :

- les traitements mathé-statistiques doivent impérativement prendre en compte, intégrer, les significations précises et circonstanciées des catégories de classement des objets dénombrés, ordonnés ou mesurés et, au-delà, les discours dans lesquels ces catégories prennent sens (rappeler au moins les formulations exactes des questions, des «relances», etc..., signaler les commentaires libres des enquêtés, leurs réponses «autres» ou «à côté», ou «éludées», etc...), sous peine de n'être que des *«exercices de calcul»*,

- et les analyses de textes ne peuvent négliger ni les «opérateurs de quantification» que contient tout énoncé ni les formes de répartition spatio-temporelle de leurs éléments constitutifs, qui contribuent à leurs significations, sous peine de n'être que des *«exercices de littérature»*.

Et il ne faut pas oublier, ou sous-estimer, que c'est toujours le discours qui a «le dernier mot» : pour lire et décrire, coordonner et synthétiser, interpréter et commenter, réinterroger et extrapoler, tous résultats numériques ou graphiques (et on pourrait en dire autant des images).

Cela revient à reconnaître, notamment avec Canguilhem, que toute élaboration de connaissances implique de «**travailler les concepts**», simultanément et dialectiquement,
- «**en in-tension**» (ou en «noumènes») : constructions théoriques de systèmes conceptuels et de paradigmes, avec tous les outils de la pensée à notre disposition ou à inventer)
- **et «en ex-tension**» (ou en «phénomènes») : observations empiriques des occurrences de faits réels, y compris bien entendu de faits langagiers, de leurs manifestations spatio-temporelles).

[1] Par commodité je n'emploierai pas toujours les guillemets qui s'imposent pour désigner ces termes dont je récuse la pertinence, et j'emploierai souvent les termes tronqués usuels Quali et Quanti.

[2] On pourra consulter à ce propos un n° spécial de la revue *Utinam* (revue de sociologie et d'anthropologie), Paris, Editions L'Harmattan, 2000, qui, sous le titre "L'analyse des trajectoires : *Ressources qualitatives et quantitatives ?*", rend compte d'une Journée de synthèse (22 juin 1998) organisée à l'Université de Versailles / St-Quentin-en-Yvelines, à l'issue d'un atelier de toute une année, sur le thème : «*Questionnaires et entretiens face aux trajectoires : quels appariements ?*».

[3] cf., notamment, la théorie des trois «niveaux de mesure» de Stevens et les «non-parametric statistics» de Siegel – des grands classiques de la période Lazarsfeld – Boudon.

[4] sur ce problème de l'agrégation, il y aurait beaucoup à dire, tant sur le plan pratique des procédures de catégorisation que sur un plan logique plus fondamental, tel qu'abordé dans toute une tradition, notamment depuis Condorcet et Arrow jusqu'à Guilbaud et Arkhipoff ...

[5] expression qui, à mon avis, mériterait d'être qualifiée d'«oxymoron» (= ingénieuse alliance de mots contradictoires)

[6] et même parfois bien au-delà du raisonnable : songeons par exemple à ces caricatures d'«échelles de prestige socio-professionnel» qui étaient en vogue dans la littérature sociologique anglosaxonne des années 60-70, échelles mesurant au millième de point près les «distances sociales» entre tous les emplois professionnels, de véritables *échelles sociales*, réduisant les antagonismes statutaires fondamentaux de nos sociétés capitalistes à l'unique critère flou d'un «prestige social» que même la sociologie simplificatrice d'un Bourdieu en termes de «capitaux symboliques» ne saurait valider.

[7] en enseignant la méthodologie et l'épistémologie sociologiques à l'EPRASS en 1966-67 puis à l'Université de Montréal en 1967-69, j'ai appris à mettre de l'ordre dans la panoplie désordonnée des différents coefficients et tests statistiques usuels. Dans ma classification multidimensionnelle des outils statistiques, la dimension des trois «niveaux de mesure» de Stevens se combine avec deux autres dimensions principales qui ne sont pas toujours combinées avec autant de visibilité dans les manuels :
- le *mode d'analyse*, descriptif (l'observation) ou probabiliste (l'inférence),
- et l'*ordre de complexité*, selon le nombre de critères de répartition pris simultanément en considération (de 0 à n, où n n'est limité que par nos capacités d'appréhender la complexité booléenne et où l'ordre zéro représenterait une population d'individus indifférenciés).

[8] ... encore que les objets de recherche «matériels» et «naturels» ne peuvent être appréhendés sans leurs «vêtements langagiers et discursifs» : terminologies naïve et savante, formules usuelles d'ordre mathématique, physique, chimique, biologique, paradigmes, concepts et théories «déjà là», postulats philosophiques et logiques sous-jacents, etc Mais il est vrai qu'ils ne sont pas totalement *immergés dans les discours* comme le sont les «objets sociaux», qui en sont littéralement imprégnés (aphorismes, mythologies, proverbes, formules langagières et discursives - cf. J.P. Faye et A. Krieg, voire iconiques, etc...) lorsqu'ils ne parlent pas eux-mêmes *en direct*.

[9] cette distinction entre objets «muets» et objets «parlants» me paraît plus féconde que la distinction classique entre «objectivité» et «subjectivité», voire entre «explicatif» et «compréhensif». Mais cela nous entraînerait trop loin de développer ici cette proposition.

[10] Ces conceptions ont été exposées publiquement dès 1966, bien que publiées après Mai 68 dans une revue animée par P Naville, aujourd'hui disparue : Proposition pour

l'élaboration d'une conception épistémologique cohérente et pour son application à la recherche en sciences sociales, in *Epistémologie Sociologique*, N° 6, Paris, 2ème semestre 1968, pp. 65-78.

[11] Comme je n'ai jamais voulu postuler le grade de directeur de recherche au CNRS, je n'avais pas eu l'occasion de pratiquer ce bilan autobiographique rituel fort utile qu'on nomme «exposé des titres et travaux». C'est pour pallier ce manque que j'ai rédigé mon dernier Rapport d'activité officiel au CNRS (en 1995) sous forme de «Testament scientifique», dans lequel j'exprime mon amertume de n'avoir pas été vraiment entendu ni exploité par mon employeur tout au long de ma «carrière», mais aussi dans lequel j'expose les principales «trouvailles» qui me semblent mériter encore maintenant d'être prises en considération. On pourra consulter ce document et quelques autres, ainsi que ma bibliographie, sur mon site Web personnel en cours d'élaboration, à l'adresse suivante : <http://pageperso.aol.fr/jacquesjenny/mapage/testamentscientifique.html>

[12] Types homogènes calculés et/ou classes significatives construites. Réflexions épistémologiques sur les modes de traitement de la complexité, in *Enquêtes statistiques et indicateurs des pratiques familiales*, IRESCO, mars 1989, pp. 67-75.
Comme il s'agit d'un document interne de l'Iresco, sans grande diffusion, on pourra le consulter sur mon site Web personnel.

[13] expression extraite du début de l'ouvrage fondateur de «l'Analyse des Données»

[14] J'emprunte (en extrapolant ?) certains termes de cette problématique aux statisticiens Halphen, Morlat et Arkhipoff, pour distinguer ces deux grands types de "rapports de connaissance" au réel (mêlés à des rapports de force) que sont :
- la conception fréquentiste/idéaliste, qui implique une croyance substantialiste, objectiviste, en LA *Réalité "en soi"* (dotée de propriétés "déjà là" - qu'il suffit de dé-couvrir), dont les idéalités mathématiques (intemporelles et universelles) constitueraient le moyen d'accès privilégié,
- et la conception intuitionniste/constructiviste, qui postule au contraire la *relativité socio-historique des connaissances*, "assumées" (et non dévoilées) dans et par la pratique sociale et les pratiques sociocognitives, avec leurs conflits et leurs contrats négociés ou décisions imposées.

[15] Point n'est besoin d'endosser les habits contraignants du structuralisme ou du systémisme pour se rallier à ce paradigme et pour en tirer toutes les conséquences.

[16] Qui ne se rappelle ici la leçon magistrale de Guttman, invité par Jean Stoetzel au Centre d'Etudes Sociologiques au début des années 60 pour y présenter sa méthode des «*classifications à facettes*», devenue depuis lors un grand classique mais assez peu pratiquée, du moins en France ?
Qui n'a pas profité également des enseignements méthodologiques d'un Barton, qui nous conseillait de construire ce même type d'indicateurs multidimensionnels «*par substraction logique des espaces d'attributs*» ?

[17] cf. Barthes Roland. *Eléments de sémiologie*, in *Le degré zéro de l'écriture*. Paris, Gonthier, 2ème éd., 1964, pages 79-177.

[18] Blanché Robert. *Structures intellectuelles*. Paris, Vrin, 1966, 147 p.

[19] je me réfère ici à quelques petits livres de vulgarisation, comme celui de P. Achard (*La sociologie du langage*, P.U.F., Que sais-je ? n° 2720, 1993) et celui de G.E. Sarfati (*Eléments d'analyse du discours*, Nathan, coll. 128, n° 156, 1997).

[20] on pourra consulter cette classification en $5 * 5 = 25$ «formats de questions» sur mon site Web.

[21] selon l'expression développée par Alain Blanchet dans : *Dire et faire dire. L'entretien*. Paris, A. Colin, 1991.

[22] J'ai eu l'occasion de repérer de tels effets dans une recherche de type documentaire sur :
Les discours sociaux sur "la Jeunesse" dans les années 60 : production, circulation, évolution et articulation avec les pratiques sociales et représentations collectives, in *la Jeunesse en questions, Orientations de recherche et Sources documentaires, La Documentation française*, sous la direction de Jean-Charles Lagrée et Paula Lew-Fai, Paris, 1983, pages 19-44.

[23] cf. par exemple les analyses comparées de réponses fournies selon la formulation exacte des questions posées, dans : *Le discours des sondages d'opinion*. Richard-Zapella Jeannine et Tournier Maurice (eds.), *MOTS*, n° 23, juin 1990, ... où l'on peut remarquer, par exemple, que la proportion de «croyants

en Dieu» passe de 66% à 81% selon qu'on a préalablement demandé : «est-ce que vous croyez en D.?» ou bien «croyez-vous en D.?».
cf. aussi l'inventaire synthétique et la «théorie du questionnaire», élaborés par J.P. Grémy en 1987.

[24] Jenny Jacques. Un nouveau paradigme pour penser le changement ? le processus d'individuation transductive, selon Gilbert Simondon. *Les Cahiers de l'Implication*, n° 1. Université de Paris-8, Département des Sciences de l'Education, Groupe de Sociologie Institutionnelle, Printemps 1998, pp. 31-44.

[25] On ne conseillera jamais assez de lire ou relire la thèse pionnière de Colette Guillaumin, qui contient ce qu'on peut appeler une méthode intelligente d'analyse de corpus textuels : L'idéologie raciste, genèse et langage actuel. Paris-La Haye, Mouton, 1972, 248 p.

[26] On trouvera l'exposé de ces préambules épistémologiques, ainsi que l'ensemble de l'article, soit dans le n° 44-3 de *Current Sociology*, difficile d'accès, soit dans une version «expansée» publiée dans le n° 54 du B.M.S., sous les titres suivants :

- Analyses de contenu et de discours dans la recherche sociologique française. Pratiques micro-informatiques actuelles et potentielles. in Mangabeira Wilma editor, *Qualitative Sociology and Computer Programs: Advent and Diffusion of Computer-Assisted Qualitative Data Analysis Software (CAQDAS)*, *Current Sociology* (44-3) hiver 1996 : pages 187-321.

- Méthodes et pratiques formalisées d'analyse de contenu et de discours dans la recherche sociologique française contemporaine. Etat des lieux et essai de classification. *Bulletin de Méthodologie Sociologique (B.M.S.)*, n° 54, mars 1997, pp. 64-112.

La version du B.M.S. est également disponible, grâce à l'autorisation de Karl Van Meter, sur le site Web du B.M.S. ou sur mon propre site personnel.

[27] On trouve la même métaphore employée à propos d'itinéraires professionnels de couples, que l'indexation réduit à des "phrases" constituées de séquences d'état assimilées à des "suites de mots", dans : Guérin-Pace France. La statistique textuelle. Un outil exploratoire en sciences sociales, *Population*, 52ème année, n° 4, Juillet-Août 1997, pp. 865-887.

[28] Derycke Marc. Le clivage du signe selon Emile Benveniste. *Langage et Société*, n° 70, Décembre 1994, pp. 35-60.

[29] Jenny Jacques. Pour engager un débat avec Max Reinert, à propos des fondements théoriques et des présupposés des logiciels d'analyse textuelle. *Langage et Société*, n° 90, décembre 1999, pp. 73-85.

[30] à quoi on devrait ajouter cet autre non-choix de recherches qui ne prendraient en compte ni la dynamique diachronique, qui produit des transformations d'usages lexicaux à moyen et long terme, ni les variations langagières et discursives selon les «milieux sociolinguistiques» - avec leurs éventuels effets socio-politiques, en termes de rapports de force ou de dominance..