

Congrès AFSP Toulouse

Table Ronde 1 "Réflexions sur les méthodes en science politique des deux côtés de l'Atlantique"

Session 3

DOGAN Mattei (CNRS et UCLA)

mattei.dogan@wanadoo.fr

<p>La France Découpée en Dix Tranches Sociologiques <i>La Méthode de la Décilisation</i></p>

La notion de contexte écologique

On donne à cette notion des spécifications diverses, allant du simple agrégat statistique au complexe « espace d'attributs ». Dans cette étude, on entend par contexte écologique ces quatre versions ou éclairages :

- a) On est en présence d'un contexte écologique lorsqu'un changement quantitatif entraîne un changement qualitatif, c'est-à-dire lorsque certains seuils des variables indépendantes entraînent des mutations dans la variable dépendante. Exemple : La propension à voter socialiste augmente plus que proportionnellement par rapport à l'accroissement du pourcentage d'ouvriers dans l'électorat.
- b) En termes statistiques, il y a contexte écologique quand la convergence de plusieurs variables produit un effet global plus important que la somme des variables prises isolément, autrement dit, quand il y a un effet cumulatif, amplificateur, un « effet de halo ».
- c) Dans l'optique de l'école structuraliste, on peut parler d'effet de contexte écologique quand la structure exerce une influence à rebours sur les parties qui la composent. Une telle définition présente l'avantage d'englober la variable dépendante dans l'ensemble des variables. Exemple : dans un village très catholique, où la messe de dimanche est la meilleure occasion de rencontre, la minorité des agnostiques ou indifférents assiste néanmoins à la messe.
- d) Il y a contexte écologique quand une seule variable est si importante par rapport aux autres qu'elle devient, pour ainsi dire, une super-variable, conditionnant tous les autres indicateurs. Disons tout de suite, pour indiquer le fil de la démonstration, que le facteur religieux apparaîtra souvent dans la présente analyse comme une super-variable.¹

¹ *Note méthodologique* : L'analyse de régression multiple donne de meilleurs résultats si l'on dispose de beaucoup d'unités et qu'on limite le nombre de variables, alors que l'analyse factorielle exige une confrontation de nombreuses variables pour un nombre limité d'unités. La complexité du déterminisme social est-elle, que la part d'explication apportée en propre par une nouvelle variable diminue en raison de l'interchangeabilité des facteurs. Si l'on classe les variables en fonction de la part d'explication qu'elles apportent, mesurée par la proportion de variance cumulée, la courbe devient vite asymptotique. Même liées par un fort coefficient de corrélation simple à la variable dépendante, deux indicateurs combinés ne donnent pas nécessairement, s'il y a interchangeabilité, un coefficient de corrélation multiple plus élevé.

L'interrogation centrale dans la littérature d'écologie politique est de savoir dans quelles conditions peut-on inférer les comportements individuels à partir de données collectives. La bibliographie sur ce thème est considérable, particulièrement aux Etats-Unis. Diverses solutions ont été avancées dont beaucoup portent sur le raffinement statistique ou mathématique. Dans ce texte, je propose une méthode qui, curieusement, a échappé à l'attention des politologues et des sociologues. Seuls quelques démographes et statisticiens s'y sont intéressés. La méthode consiste dans l'homogénéisation des espaces sociaux autant que le permet la documentation statistique disponible. Prenons un exemple simple. Supposons que dans une région la moitié de la population se compose d'ouvriers de l'industrie, et l'autre moitié, de classes moyennes ou d'agriculteurs. Cette même région vote à moitié pour la gauche, et l'autre moitié pour la droite. On veut savoir quelle est la proportion d'ouvriers de l'industrie qui vote pour la gauche. Dans une telle situation, on ne peut formuler aucune hypothèse vérifiable. Supposons maintenant qu'au lieu de traiter la région dans son ensemble, on la découpe statistiquement en plusieurs tranches à partir de données agrégées par districts, municipalités ou cantons, en fonction de certaines variables qui sont « bloquées », « gelées », afin de « libérer » les autres variables. On peut ainsi créer des « univers » éminemment ouvrier ou agricole ou à forte prédominance de gauche ou de droite. On peut séparer d'un côté, les unités agricoles, et de l'autre, les localités industrialisées. Dans des contextes ainsi homogénéisés, il devient possible d'inférer des comportements individuels à partir de données collectives et de démontrer d'une manière précise qu'on ne l'a fait auparavant, l'extrême diversité du territoire français, façonné par la nature et l'histoire bien plus que le territoire de bien d'autres pays. La décilisation ici proposée répond à cette logique.

Deux objectifs

L'idée que l'homme est un être social est un principe fondamental dans les sciences sociales. Le comportement d'un individu ne peut être compris s'il est « extrait » de son contexte social, parce que c'est précisément ce contexte qui donne une signification à ces attitudes et opinions. Une telle extraction est pourtant accomplie dans la recherche par sondage stratifié, à l'exception des cas rares où l'information est collectée délibérément sur le contexte social de l'individu. Il n'est donc pas surprenant de trouver dans une bibliographie de dizaines d'études électorales publiées en Europe et aux Etats-Unis durant les trente dernières années que très peu de ces analyses expliquent plus d'un quart de la variance. Ces études ne réussissent pas à expliquer une variance plus significative parce qu'elle se concentre exclusivement sur les caractéristiques personnelles des individus, sans référence aux contextes sociaux. Cette situation a été dénoncée comme une « *individualistic fallacy* » (Scheuch, 1966).²

L'analyse de régression multiple basée sur des données agrégatives n'aboutit pas toujours à de meilleurs résultats. Dans ces études agrégatives les différences locales des contextes sociaux s'annulent les uns par les autres. La négligence du contexte local est

Certes, un grand nombre de variables offre plus de chance de prédire la variable dépendante avec précision. En revanche, un nombre limité de variables permet d'expliquer plus facilement la relation entre chaque indicateur et le phénomène étudié.

D'un point de vue strictement statistique, en multipliant les variables, on peut espérer un coefficient de corrélation multiple plus élevée et une prédiction plus précise. Mais, dans une étude écologique, c'est la relation de cause à effet entre les principales variables indépendantes et la variable dépendante qui nous importe.

² Scheuch, Erwin, 1966, « Cross-National Comparisons Using Aggregate Data », dans R.L. Merritt et S. Rokkan, eds., *Comparing Nations*, Yale University Press, p. 131-168. Une des principales critiques adressées à l'ouvrage classique de Gabriel Almond et Sydney Verba (1963), *The Civic Culture* est d'avoir basé leurs interprétations sur les « erreurs écologiques ».

déplorable car presque tous les pays connaissent une grande diversité territoriale : en particulier, France, Italie, Espagne, Portugal et Canada.

L'inférence des caractéristiques individuelles à partir des données collectives comporte des risques d'interprétations fallacieuses. En 1950, un article de sept pages par W.S. Robinson³ sur « *l'ecological fallacy* » a suffi pour provoquer un séisme méthodologique à travers plusieurs sciences sociales. On a alors changé de méthode mais pour tomber, par l'enquête à base d'échantillonnage stratifié, dans « *l'individualistic fallacy* ». Il faut donc que la recherche s'effectue à plusieurs niveaux.⁴

Par exemple, dans des régressions multiples, on a trouvé au niveau national une très faible corrélation entre la proportion d'ouvriers de l'industrie et la proportion de votes du gauche pour 2.450 cantons (.13), et d'autre part une très forte relation entre la pratique religieuse et le vote de gauche (-.63). Dans une deuxième recherche, l'analyse de régression à l'échelon national a été remplacée par 87 analyses de régression dans le cadre de chacun des 87 départements français. La nouvelle analyse montre clairement le caractère fallacieux de l'analyse de régression multiple à l'échelon national : la corrélation entre la proportion d'ouvriers et vote de gauche a augmenté alors de .13 à .43. Une recherche similaire a été effectuée pour l'Italie centrale avec des résultats similaires.

Cette étude se propose de réhabiliter l'analyse des données agrégatives, en montrant que dans certaines conditions l'approche écologique (dans le sens d'espace social) peut s'avérer statistiquement valable et sociologiquement pertinente. On propose donc de construire des catégories statistiques à partir d'unités territoriales. Les tranches statistiques sont débarrassées des contingences géographiques. La condition requise pour la décilisation est de créer des univers relativement homogènes, écologiquement unidimensionnels. L'idée est de « bloquer » une variable pertinente, par exemple la pratique religieuse, afin de « libérer » la variable classe sociale, ou vice-versa.

Il est vrai que certains aspects du comportement politique peuvent mieux être analysés par des enquêtes effectuées par échantillon stratifié que par des données agrégatives. Par exemple, il n'est pas possible d'observer la « conscience de classe » dans des données agrégatives. Le contraire est également vrai : le contexte social est mieux observé dans des données agrégatives que dans des enquêtes d'opinion. La recherche idéale devrait combiner données collectives et données individuelles, mais elle exige des moyens très importants.

En somme, cet essai a deux objectifs. Le premier est méthodologique de montrer qu'il est possible de saisir l'impact du contexte social sur le comportement individuel, tout en évitant les risques de corrélations fallacieuses en analysant des données agrégatives dans des contextes sociaux statistiquement construits. Le deuxième objectif est substantiel : établir les relations entre classe, religion et politique dans une France dont la géographie est délibérément transmutée en un univers statistiquement organisés où l'individu est placé naturellement dans son milieu social. J'emprunte dans cette étude sur la décilisation une documentation qui a été élaborée en collaboration avec Daniel Derivry.

Analyse par déciles en tant qu'expérimentation simulée

Dans les sciences naturelles le chercheur peut procéder à des expériences en changeant les paramètres selon les hypothèses retenues. Par contre, dans les sciences sociales le chercheur ne se trouve pas dans une situation aussi favorable. L'expérimentation directe est impossible au niveau des groupes sociaux larges (bien qu'elle soit concevable dans des petits

³ Robinson, W.S. 1950, « Ecological Correlations and the Behavior of Individuals », *American Sociological Review*, 15, pp. 351-357.

⁴ Pour une discussion récente des relations fallacieuses dans l'analyse écologique voir : G. Firebaugh, « Statistics of Ecological Fallacy » et D.D. Friedman « Statistics Inference » dans *International Encyclopedia of the Social and Behavioral Sciences*, 2002, Vol. 5, respectivement pp. 4023-26 et 2027-30.

groupes en termes de psychologie sociale). Mais il y a une alternative à l'expérimentation directe : l'expérimentation simulée, par la création de catégories statistiques. Une meilleure compréhension de la logique des expériences simulées peut être obtenue par une comparaison. Le chimiste peut analyser une substance à différents degrés de température. Mais le sociologue ne peut pas changer les données agrégées, ni le niveau d'affiliation politique, ni la fréquence de la pratique religieuse ne peut être modifiée. Il ne peut pas provoquer un processus dynamique ; il peut seulement le simuler, en prenant en considération divers niveaux de différenciation et en observant les caractéristiques des sous-groupes de population artificiellement créés. Le sociologue ne pouvant pas procéder à une expérimentation directe peut en revanche utiliser des classifications expérimentales. La division d'une population en déciles est une de ces classifications. Les catégories ainsi créées peuvent être unidimensionnelles, c'est-à-dire fondées sur une seule variable. Elles peuvent aussi être construites en prenant en considération plusieurs variables, mais alors l'analyse devient plus lourde.

Comment peut-on diviser une population ? Tout dépend du nombre d'individus à inclure dans chaque sous-catégorie. Il serait possible de diviser la population en niveaux équidistants: (par exemple, 5% de croissance de pratique religieuse). Mais la distribution inégale des individus entre sous-catégories équidistantes présente quelques désavantages.

Une autre méthode consiste à diviser la population en sous-catégories numériquement égales. Le nombre de catégories dépend de la dimension de la population analysée, car chaque sous-catégorie doit contenir un nombre suffisant de cas. De plus, l'hétérogénéité de chaque unité devrait être minimisée. Ces deux impératifs sont dans un sens contradictoires. Les variations seraient mieux contrôlées en répartissant la population en dix ou vingt catégories, plutôt qu'en quartiles. On a divisé le nombre de cantons, soit 2 450, en déciles, en utilisant des données agrégatives par cantons. En conséquence, chaque décile contient 245 cantons.

Dites-le avec des chiffres et non pas en paroles

J'ai construit des univers statistiques basés sur l'unité administrative la plus petite disponible en France. Les grands villes ont été exclues parce que l'une des variables les plus significatives, la pratique religieuse, n'était pas disponible pour ces villes. On a ainsi réuni pour chacun de ces 2 450 cantons, des données quantitatives concernant la structure sociale, ainsi qu'une série d'indicateurs culturels, et de même les résultats électoraux. Les données concernant la pratique religieuse furent collectées par le chanoine Boulard et ses collègues, avec la collaboration de centaines de curés durant les années 1960. Ces statistiques sur la pratique religieuse n'ont pas été mises à jour, et en conséquence il n'est pas possible de considérer des élections plus récentes, à moins d'admettre que la pratique religieuse évolue très lentement, et qu'en conséquence on pourrait procéder à une analyse asynchrone. La pratique religieuse a diminué d'une manière significative durant les dernières décennies mais on sait par ailleurs qu'il existe bien une « hérédité » socio-politique dans le milieu rural, un transfert des tendances politiques, par la socialisation, du père ou de la mère au fils ou à la fille. L'analyse ici présentée se réfère aux élections législatives de 1968 qui sont considérées par des historiens comme des élections très significatives, des élections critiques.

Les tableaux statistiques ainsi obtenus ont été traduits en graphiques, qui présentent de manière concise et parlante de multiples aspects et qui évitent de longues descriptions en paroles ou en formules mathématiques compliquées. On a analysé trois séries de déciles : par pratique religieuse, par structure sociale, et par le vote gauche-droite. Ensuite on s'est placé dans des déciles extrêmes: le décile le plus à droite, le décile le plus à gauche, le décile le plus agricole, le décile le plus industrialisé, le décile le plus traditionaliste, le décile le plus indifférent à l'égard de la religion. Les contextes sociaux ainsi créés de toutes pièces mettent

en évidence l'extrême diversité de la France, d'une manière que la vieille géographie électorale n'a pas réussi à établir dans toute son étendue. En effet, la pratique religieuse varie de 1% à 97% ; le vote de gauche varie de 3% à 86%.

Ce texte est basé sur un traitement statistique informatique en plusieurs étapes au cours d'une longue période, en particulier une analyse de régression multiple portant sur 24 variables économiques sociales, culturelles et politiques, totalisant 60000 données quantifiées, qui ont pu expliquer la moitié des fluctuations du vote en France (à l'exception des grandes villes). La nécessité de recourir à une division de l'univers statistique en déciles ne m'est apparue qu'après une recherche préalable sur « l'espace de référence », c'est-à-dire de la position de chaque canton dans son département de référence. Les étapes et les méandres de cette recherche apparaissent dans la liste ci-après des publications qu'elle a engendrées pendant plusieurs décennies.⁵

Déciles par pratique religieuse

La transformation des résultats statistiques en graphiques montre que la courbe du vote de gauche décline de décile en décile de pratique religieuse. Du décile le plus religieux au décile le moins religieux, le vote de gauche perd les deux tiers de sa force. On observe un

⁵ *Etapes d'une recherche écologique (ordre chronologique):*

- Dogan, Mattei, 1960, « Le vote ouvrier de l'Europe occidentale », *Revue Française de Sociologie*, 1, 25-44
- 1965, « Le vote ouvrier en France : Analyse écologique », *Revue Française de Sociologie*, VI, 435-471.
- 1966, « Les contextes politiques en France : Analyse écologique », dans *Les Méthodes d'analyse écologique quantitative dans les sciences sociales*, Rapport présenté au colloque organisé par le Conseil International des Sciences Sociales, sous la direction de M. Dogan et S. Rokkan, Evian, Septembre 1966, *Cahiers du Centre d'Etudes Sociologiques (CNRS)*, 86 pages.
- 1967, « Political Cleavages and Social Stratification in France and Italy », S.M. Lipset et S. Rokkan, *Party Systems and Voter Alignments*, New York, The Free Press, pp. 129-195.
- avec Stein Rokkan, 1967, « Quantitative Ecological Analyses », *Social Science Information*, pp. 35-47.
- 1968, « L'analyse de covariance en sociologie électorale », *Revue Française de Sociologie*, Oct-Déc. pp. 537-47.
- 1968, « "Un fenomeno di atassia politica" dans Partiti Politici e Strutture Sociali in Itali », M. Dogan et O. Petracca, eds., Milano, Edizioni di Comunita, pp. 465-80.
- avec Stein Rokkan, 1969, *Quantitative Ecological Analysis in the Social Sciences*, Cambridge, MIT. Press. Republié sous le titre, *Social Ecology*, MIT Press, 1974.
- Clubb, J.M, W. Miller, G. Marks, M. Dogan et S. Rokkan, 1969, *Ecological Data in Comparative Research, Report on the International Data Seminar*, Institute of Social Research, University of Michigan et International Social Science Council, 70 pages.
- Dogan, Mattei et Daniel Derivry, 1971, « Unité d'analyse et espace de référence en écologie politique : le canton et le département français », *Revue Française de Science Politique*, XXI, 3, pp. 517-570.
- Dogan, Mattei, 1981, « La transmission 'héréditaire', des tendances politiques », *Science et Théorie de l'Opinion Publique, Hommage à Jean Stoetzel*, Paris, Retz, pp. 216-28.
- avec Daniel Derivry, 1986, « Religion, classe et politique en France », *Revue Française de Science Politique*, XXXVI, 2, pp. 157-180.
- avec et Daniel Derivry, 1988, « France in Ten Slices : An Analysis of Aggregate Data », *Electoral Studies*, 7, 3, pp. 251-267.
- Dogan, Mattei, 1993, « De Gaulle et la classe ouvrière », *Espoir*, 91, pp. 53-62.
- 1996, « Classe, religion, parti : triple déclin dans les clivages électoraux », *Revue Internationale de Politique Comparée*, III, pp. 519-540.
- 2000, « Triple Decline in Electoral Cleavages in Western Europe », L. Karvonen et S. Kuhnle, eds., *Party Systems and Voter Alignments Revisited*, London, Routledge, pp. 93-114.

changement rapide entre le 8^{ème} et le 9^{ème} déciles. La tendance exponentielle des déciles extrêmes démontre l'impact du contexte social sur le comportement individuel.

Déciles par classes sociales

La proportion d'ouvriers de l'industrie indique la structure sociale de l'unité écologique. Cet indicateur est lié à d'autres variables comme par exemple le degré d'industrialisation ou d'urbanisation. La proportion de salariés dans l'agriculture est aussi un indicateur significatif, mais seulement pour quelques régions en France.

Au niveau national, la classe sociale et le comportement politique sont faiblement corrélés, mais au delà d'un certain niveau de la proportion d'ouvriers de l'industrie la relation entre classe et vote s'accroît. Ceci peut être facilement expliqué. Les effets de la concentration ouvrière ne sont visibles qu'au delà d'un niveau critique de densité ouvrière. C'est surtout dans les deux derniers déciles que l'impact de l'urbanisation et de la densité ouvrière atteignent un niveau élevé. Dans chaque décile, le vote de gauche est toujours plus élevé que la proportion d'ouvriers. Alors que le vote de gauche varie faiblement d'un décile au suivant, la composition sociale de l'électorat de gauche varie considérablement entre le décile le plus industrialisé et le décile le moins ouvrier.

La composition des déciles montre un accroissement de la proportion d'agriculteurs de 7 % à 68 %, et pour les ouvriers de l'industrie une diminution de 50 % à 13%. Il est clair que dans le premier décile le vote de gauche se compose presque exclusivement d'ouvriers de l'industrie et de petite bourgeoisie, alors que dans le 10^{ème} décile les agriculteurs prédominent. Ces données montrent l'importance de la petite bourgeoisie dans le vote de gauche.

Déciles du vote gauche-droite

Là où la pratique religieuse est élevée le vote de gauche est faible, et vice-versa. Toutefois, bien qu'un haut niveau de pratique religieuse constitue une barrière contre le vote de gauche, un bas niveau de pratique religieuse ne conduit pas nécessairement à gauche. Le contraste entre les deux déciles extrêmes est frappant. Dans le premier décile, où le vote de gauche est faible, les deux tiers au moins des ouvriers ne votent pas pour la gauche. Dans le dernier décile où la gauche est la plus forte, au moins la moitié du vote de gauche provient d'autres catégories sociales que la classe ouvrière. Dans ce décile le plus à gauche, la classe ouvrière est de toute évidence une minorité dans l'électorat de gauche. Cette constatation démolie une montagne de rhétorique et de polémique sur « le parti de la classe ouvrière », puisque dans la plupart des cantons la majorité des ouvriers de l'industrie n'a pas voté à gauche. Les enquêtes d'opinion ne révèlent pas une telle distribution contextuelle, mais la décilisation le montre bien.

Les effets d'interaction entre les variables

Une corrélation simple entre deux variables montre les effets de la variable indépendante sur la variable dépendante mais l'analyse multivariée requiert que le résultat ne soit pas modifié par l'intervention d'autres variables. L'interférence d'une troisième variable peut changer la relation causale. Une telle interférence peut contester des relations apparemment bien établies.

La relation entre classe et vote est très sensible au niveau de pratique religieuse. Le coefficient de corrélation entre la proportion d'ouvriers et la proportion du vote de gauche varie d'un décile à l'autre de pratiquants religieux. Il y a là de toute évidence un phénomène d'interaction. Dans des termes moins techniques, la possibilité de prédire le niveau de vote de gauche dans une série de cantons en connaissant la proportion d'ouvriers de l'industrie n'est pas la même dans les cantons à forte pratique religieuse et les cantons à faible pratique. Cette interaction peut être facilement clarifiée. En divisant la population selon la pratique religieuse

on neutralise les effets de cette variable. La corrélation entre ouvriers et votes dans chacun de ces séries de déciles peut présenter trois configurations différentes.

- a. Les coefficients peuvent ne pas varier du tout ou ils peuvent varier très peu. Dans ce cas, il n'y a pas de fait d'interaction.
- b. Les coefficients peuvent être plus élevés ou moins élevés que la moyenne nationale de corrélation. S'il est plus élevé le facteur religieux perturbe les effets de la classe sur le vote. Libérée de cette interférence, la relation entre classe sociale et comportement politique trouve sa vraie valeur. Si elle est faible, la variance attribuée à la classe sociale peut être en partie imputable à la pratique religieuse. Cette situation implique une relation initiale entre deux variables indépendantes. Ceci signifie que l'effet d'altération n'est pas le même dans des cantons déchristianisés et dans les cantons traditionalistes.
- c. Il y a des cas plus complexes où l'interaction peut être observée entre deux variables indépendantes selon la signification de la variable dépendante. Ceci confirme la difficulté, voire l'impossibilité, d'expliquer les corrélations partielles à l'échelon national, et confirme en même temps la pertinence de l'analyse par déciles.

Les tranches extrêmes

Pour chacune des variables il y a un contraste frappant entre le premier et le dernier décile. Les six déciles extrêmes représentent des contextes homogènes chacun pour sa variable respective. Cela ne veut pas forcément dire que ces six tranches extrêmes sont homogènes en ce qui concerne les deux autres variables. L'avantage d'analyser ces types unidimensionnels est justement de « libérer » les autres variables.

La différence entre le premier et le second décile, et la différence entre le neuvième et le dixième décile ne sont pas uniquement quantitatifs. Les déciles extrêmes apparaissent comme des mondes différents. Au delà d'un certain seuil nous nous trouvons dans un environnement ultra-catholique, où même ceux qui ne vont jamais à l'église sont influencés par la culture religieuse. Au-delà d'un certain degré d'anticléricalisme, on observe le phénomène contraire. Dans ces cas extrêmes, nous nous trouvons dans un contexte extrêmement traditionaliste ou dans des zones d'anticléricalisme généralisé à toutes les catégories sociales. Ce qui est vrai pour la variable religieuse l'est également pour les deux autres. C'est pour cette raison que les tranches extrêmes peuvent être considérées comme des types idéaux et même comme des cas quasi-expérimentaux.

Ces six tranches extrêmes représentent toutes ensemble une partie importante de la population française. Comme le graphique l'indique, les deux extrêmes de déciles de votes et les deux déciles extrêmes de pratique religieuse représentent à eux un tiers des cantons du territoire français. Si nous ajoutons les déciles extrêmes d'ouvriers d'industrie et les déciles extrêmes de population agricole, les dix déciles extrêmes couvrent près de la moitié de la population considérée dans cette analyse écologique, exactement 45 %.

La tranche la plus droite de la France

Ce décile comprend des cantons dispersés entre 29 départements, mais les dix départements les plus peuplés regroupent 190 cantons parmi les 245 appartenant à ce décile. Beaucoup de ces cantons les plus à droite appartiennent aussi à la tranche la plus catholique de la France. Ceci pouvait être prévu en raison de la forte corrélation entre pratique religieuse et orientation de droite. Il est toutefois intéressant de noter que certains cantons s'écartent de ce modèle. En effet, il y a des cantons plus à droite qu'ils ne pouvait être anticipés par la seule variable religieuse. Certains cantons sont plus conservateurs que traditionalistes, dans le sens que les premiers sont motivés par des problèmes économiques, alors que les autres le

sont par des valeurs religieuses. Il est important de distinguer clairement conservatisme économique et traditionalisme religieux.

Le vote de gauche dans cette tranche droitère varie de 3 à 16 %, ce qui nous autorise de considérer cette tranche comme relativement homogène du point de vue politique. Toutefois, dans la partie la plus à droite de la France, la pratique religieuse varie énormément de 12 à 96 %. Il y a là un phénomène que la géographie électorale n'a pas été capable de mettre en évidence que partiellement. Encore une fois, on doit souligner le fait qu'un fort niveau de pratique religieuse constitue une barrière contre la gauche, mais qu'un bas niveau de pratique religieuse ouvre un éventail de comportements.

La tranche la plus à gauche

Dans ce décile la distribution du vote de gauche varie de 59 à 88 %, une dispersion de 30 points. Mais son homogénéité interne est plus faible que pour les autres tranches extrêmes. Ces cantons sont dispersés entre la moitié des départements français, et 21 de ces cantons regroupent ensemble 189 des 245 cantons de ce décile.

Logiquement, nous pouvions nous attendre à un très bas niveau de pratique religieuse dans cette partie la plus à gauche de la France et, en effet, seulement 13 de ces cantons exhibent un taux relativement élevé de messalisants. Toutefois ce décile le plus à gauche ne coïncide pas avec le décile le plus éloigné de l'église : la proportion de messalisants varie de 2 à 90 %. Cette partie de la France caractérisée par le plus haut niveau de votes de gauche est donc très hétérogène. La proportion d'ouvriers de l'industrie varie de 7 à 65 %, mais cette étendue ne révèle pas une dispersion réelle. En regardant de plus près, on observe deux sous-catégories de cantons, la première avec plus de 40 % d'ouvriers de l'industrie, l'autre avec moins de 20 %. En isolant cette tranche la plus à gauche en France, nous pouvons voir dans ce monde statistiquement créé, totalement libéré des contraintes géographiques, comment les autres variables se croisent. Nous observons en effet une forte variation selon les autres variables ce qui nous permet de dire que le vote de gauche est polymorphe, dans le sens sociologique du mot. La plupart des cantons du premier décile et du dernier décile se distribuent entre de nombreux départements. Pourtant, il est surprenant de trouver que certains des cantons les plus à gauche coexistent avec certains des cantons les plus à droite dans le même département. C'est le cas par exemple dans les départements de Finistère, Aveyron et Lozère. Faut-il rappeler que la géographie électorale a amalgamé ces cantons dans les cartes géographiques. Considérant la diversité interne de certains de ces départements, on voit que la technique cartographique cache des informations essentielles et qu'elle n'est en conséquence qu'une représentation approximative.

La tranche la plus agricole de la France

Dans ce décile nous nous trouvons vraiment dans un monde agricole. Malgré un haut degré de homogénéité sociale, la proportion du vote de gauche varie de 3 à 83 %. Cette diversité politique énorme ne peut pas être attribuée à la variation de la proportion des ouvriers de l'industrie, car dans 202 de ces 245 cantons, les ouvriers de l'industrie ne représentent que 10 à 15 % de la population active. Cette variable est donc « congelée ». Nous sommes ici parmi de vrais agriculteurs, mais ils sont assez différenciés, petits et grands fermiers, éleveurs de bovines, grands et petits propriétaires dans des terres riches et des terres pauvres et ainsi de suite. Une telle diversité de la population agricole ne peut pas ne pas avoir des conséquences sur le comportement politique. Il vaut mieux ne pas entrer ici dans une analyse détaillée des comportements politiques des diverses strates de la population agricole car nous risquerions de nous perdre dans des sentiers. De toute manière la relation entre type d'agriculture et comportement politique est obscurcie par l'influence prépondérante de la pratique religieuse, qui s'étend de 2 à 91 %. De toutes les expérimentations par décile

discutées ici, cette dispersion est la plus spectaculaire. La proportion de vote de gauche est la proportion de messalisants dans ce décile atteint un coefficient record -.70. Dans ces tranches de la France construites statistiquement, on peut, sans risque de tomber dans « *ecological fallacy* », inférer des données agrégatives aux comportements individuels.

Les 245 cantons composant la tranche la plus agricole se distribuent entre 51 départements. Pour la plupart de ces départements, l'analyse de régression multiple a montré un haut degré de corrélation entre l'attitude religieuse et l'attitude politique : - .76 pour l'Aveyron. Mais dans d'autres départements déchristianisés le facteur religieux perd tout pouvoir discriminatoire : - .10. Ainsi, dans le dixième de la France le plus agricole, il y a des zones où le contexte anti-clérical laisse peu de place pour les fluctuations entre la pratique religieuse et l'orientation politique.

La tranche la plus industrialisée de la France

Dans ce décile, la proportion d'ouvriers de l'industrie est de 54 %, et la proportion du vote de gauche est de 44 %, c'est-à-dire la même proportion que dans le décile le plus agricole. Mais cette moyenne cache d'importantes fluctuations qui vont de 5 % à 75 %. Cette variation peut être expliquée en partie par la proportion de gens appartenant à la petite bourgeoisie. Elle peut être expliquée aussi par la diversité interne de la classe ouvrière, en particulier par l'importance relative des diverses branches de l'industrie. Pourtant, l'explication essentielle de la variation du vote de gauche dans cette partie la plus industrialisée de la France, est, une fois de plus, la pratique religieuse, qui varie d'une manière étonnante de 5 % à 90 %. Ceci est un fait dur comme l'acier. Devant cette constatation irréfutable que reste-t-il des millions de pages écrites sur la « classe ouvrière » et sur le parti de cette « classe ». Ici se trouvent face à face deux mondes contrastés de la France, le plus agricole et le plus industrialisé, et où toutefois l'importance du facteur religieux dans l'explication du vote s'impose d'une manière également évidente.

La tranche la plus dévote de la France

Dans cette tranche, la proportion des messalisants oscille de 67 % à 90 % avec une moyenne de 78 %. La proportion moyenne d'ouvriers d'industrie est de 29 % avec une variation de 1 % à 67 %. Ici, nous avons une grande diversité sociale. Dans ce décile les gens sont allergiques au socialisme agnostique : les 245 cantons offrent une faible attention aux partis considérés comme anti-cléricaux. Nous ne sommes pas surpris de trouver pour ce décile une corrélation entre la proportion d'ouvriers de l'industrie et le vote de gauche presque nulle (-.06). Dans un contexte très catholique, la classe sociale n'influence pas le comportement politique.

La partie de la France la plus détachée de la religion

Les cantons de ce décile, comme tous les autres déciles, ne représentent pas une aire géographique mais une construction statistique où se trouvent probablement l'une des poches les plus déchristianisées de l'Europe, à l'exception de certaines zones de l'Espagne ou de l'Italie centrale. Seulement en Toscane, Emilie et certaines régions de l'Espagne méridionale on trouve une proportion aussi faible de messalisants. La moyenne est de 7 % et elle varie de 1 à 10 %. Nous pouvons donc considérer le comportement religieux de ces cantons comme « congelé » pour notre analyse, mais les deux autres dimensions varient grandement : la proportion d'ouvriers de l'industrie varie de 9 à 64 %, et le vote de gauche varie de 19 à 89 %. Le facteur religieux étant ainsi neutralisé, nous pourrions nous attendre à trouver des circonstances idéales pour observer les relations entre classes sociales et attitudes politiques, car dans ce décile il ne peut y avoir de pression croisée entre classe sociale et attachement religieux, la pratique religieuse étant partout extrêmement faible. La base sociale

du comportement politique devrait pouvoir s'épanouir en toute liberté. Pourtant ce n'est pas le cas, car le coefficient de corrélation entre la proportion d'ouvriers d'industrie et de vote de gauche est curieusement négative : .26. Cette corrélation négative suggère que les cantons les plus à gauche sont à prédominance agricole. Il semble que la probabilité de l'orientation de gauche est plus élevée dans certaines régions agricoles que parmi les ouvriers de l'industrie. L'implication est que l'absence de pratique religieuse est idéologiquement plus significative pour l'agriculteur que pour l'ouvrier de l'industrie. Alors que dans un village l'absence de pratique religieuse peut être remarquée par tout le monde, dans l'anonymat de la ville elle est moins remarquée. Le comportement politique est lié autant aux conditions de la vie quotidienne qu'à la catégorie socioprofessionnelle.

Normalement, un décile correspond très peu à une région géographique spécifique car les cantons d'un décile ne sont pas contigus, ils se trouvent au nord, au sud, à l'ouest et à l'est. Toutefois, pour ce décile, on trouve un certain nombre de cantons qui sont contigus dans quelques régions relativement pauvres du point de vue agricole, comme par exemple la Haute-Vienne ou la Creuse. Dans le premier de ces départements le vote de gauche s'élève à 72 % alors que dans la proportion d'ouvriers d'industrie est seulement de 29 % ; dans le second, le vote de gauche est de 55 % alors que les ouvriers de l'industrie ne représentent que 18 % de la population. Ici, les prolétaires sont des paysans parcellaire ou des ouvriers agricoles.

Dernière tentative pour ressusciter la « classe ouvrière »

Selon les règles de l'analyse statistique, les unités d'une population sont supposées être égales numériquement. C'est ce principe d'équivalence qui nous permet de comparer les unités en fonction de diverses variables. Mais on trouve rarement une telle situation dans la réalité sociale. Il y a très peu d'études écologiques dans la littérature internationale qui prennent en considération la taille démographique des unités d'analyse. Les cantons français comme les *counties* américains ou les *Bezirke* allemands ont une dimension démographique très inégale. Le problème de la dimension démographique des unités pose un problème dans l'analyse de régression multiple. Formulons le problème d'une manière simple : est-ce que la proportion d'ouvriers de l'industrie a la même signification contextuelle quand il s'agit de 1 000 ouvriers ou de 10 000 ouvriers ? Si la signification n'est pas la même, y a-t-il un seuil ? Est-ce que nous ne négligeons pas un phénomène essentiel ? Autrement dit, la question est de savoir si la taille démographique n'est en elle-même une dimension importante.

Trois exemples hypothétiques devraient nous aider à comprendre ce problème méthodologique. Dans le premier cas, le coefficient de corrélation pour la taille démographique est de .73. Dans le second où les unités sont plus peuplées, le coefficient de corrélation tombe à .54. Dans la troisième, les unités démographiques les plus larges sont situées sur la ligne de régression et le coefficient de corrélation monte à .85. Il est clair que la dimension démographique intervient d'une manière importante dans l'analyse.

En bref, la taille qui n'exerce pas une influence directe sur les relations entre variables pour neuf déciles, intervient néanmoins dans le dixième décile avec les cantons les plus peuplés et les plus urbanisés. On a mené aussi une analyse pour les villes de plus de 30 000 habitants qui indique l'importance du vote des classes moyennes pour l'électorat socialiste. Dans une représentation graphique dans les villes en-dessous de la diagonale, une partie importante de l'électorat ouvrier avait voté en 1965 pour de Gaulle. Dans les villes au-dessus de la diagonale, le socialisme était bien diffusé parmi les classes moyennes. A Strasbourg, Colmar, Metz, Nancy, Mulhouse ou Tourcoing, les ouvriers gaullistes apparaissent clairement. A Narbonne, Arles, Béziers, Nîmes, Toulouse, Carcassonne ou Montpellier, on constate un excédent de votes pour le candidat socialiste par rapport au potentiel de l'électorat ouvrier. On y aperçoit des « petits bourgeois » socialistes.

Mea culpa, confession d'une erreur méthodologique que j'ai commise quand je me suis lancé dans une analyse factorielle de la variance, qui a abouti à une simplification excessive, si bien que j'ai dû admettre qu'après un travail acharné, j'ai jeté par la fenêtre la majeure partie de l'information scientifique réunie. Depuis, je suis devenu allergique à l'analyse factorielle et j'ai recommandé aux étudiants de s'en méfier, sans avoir réussi à convaincre certains statisticiens chevronnés. La lecture de mon article sur « La covariance en sociologie électorale » (1968) devrait être considérée comme une vaccination contre l'analyse factorielle. On apprend beaucoup en commentant, parfois inévitablement, des erreurs.

Conclusion : La moyenne nationale est presque toujours fallacieuse

La France est un des pays les plus anciens à avoir achevé son unification nationale et son identité. Mais son territoire reste un des plus diversifiés en Europe. Est-ce que cette diversité territoriale, qui est un héritage historique, continue à créer des clivages électoraux sous l'influence de l'urbanisation, du déclin de la population agricole, de la montée des classes moyennes, de l'influence des médias et surtout en raison du déclin des croyances et des pratiques religieuses ? La France était encore en 1968 une mosaïque. Le déclin de l'influence religieuse a réduit quelque peu la signification de cet indicateur mais le classement de 2450 cantons par déciles était encore à peu près le même en 1981 qu'en 1968. Des recherches plus récentes (1988-2002) montrent qu'aujourd'hui encore la France est un pays électoralement polymorphe, où le comportement politique n'a pas de signification s'il n'est pas observé dans des contextes sociaux, malgré la forte mobilité géographique de la population. La diversité territoriale des relations causales aide à comprendre pourquoi les partis politiques en France ne sont pas bien organisés à l'échelon national. Ils sont polymorphes dans le sens que le même parti se trouve parfois dans un contexte agricole ou industriel, parfois dans un contexte à prédominance religieuse ou encore dans un contexte anti-clérical. Le système électoral uninominal à deux tours correspond à cette symbiose locale. Ce polymorphisme a dans beaucoup de régions de profondes racines historiques, comme l'ont montré André Sigfried et François Goguel.

L'analyse par décile permet mieux que toute autre méthode d'expliquer pourquoi il est si difficile d'engendrer en France un régime bipolaire, bipartite. Pourtant, la bipolarité existe au niveau local, mais elle prend des configurations différentes d'une zone à l'autre : ici, c'est un conservateur qui affronte un socialiste, là c'est un socialiste qui se trouve face à un catholique centriste, et ailleurs le socialiste et l'extrémiste de droite se trouvent face à face. C'est ce polymorphisme du territoire français qui a amené à la classe politique à adopter le scrutin à deux tours. « Au premier, on choisit, au second, on élimine ».

En Angleterre où la très forte urbanisation a homogénéisé d'une manière significative le territoire national, et où la religion intervient relativement peu dans la vie politique, les compétiteurs ont partout à peu près le même profil et, en conséquence le scrutin uninominal à un seul tour est pleinement efficace pour générer un système bipartite.

Le régime politique italien est lui aussi pluripartite et pour la même raison que le régime français : le polymorphisme du territoire. Il est vrai aussi que le système électoral proportionnel adopté après la guerre a facilité le pluralisme paritaire. Le scrutin proportionnel a été d'ailleurs adopté justement à cause de ce polymorphisme, qui dans certaines régions a des racines historiques remontant à plusieurs siècles. Dans une étude de Barberis et Corraziari⁶, qui distingue dix régions écologiques (zone montagnaise, plaine fertile, etc.), mais qui n'est pas une analyse par décile, et qui porte sur 8000 communes d'Italie, le polymorphisme écologique apparaît en toute clarté, par exemple, la relation entre l'altitude

⁶ Barberis, C. et G. Corraziari, 1968, « Strutture economica e dinamica elettorale » dans M. Dogan et O. Petracca, *op. cit.* p. 411-464.

(montagne, collines) et le vote pour le parti d'inspiration chrétienne. On dirait que les habitants de ces zones montagneuses se sentent plus près des divinités célestes. En réalité, il s'agit de petite propriété paysanne, alors que dans les plaines, où les terres sont plus fertiles, les grands propriétaires emploient une main d'œuvre composée d'ouvriers agricoles et de métayers.

A l'époque récente, des nouveaux facteurs sont intervenus, par exemple la densité d'immigrants d'origine non-européenne, car la xénophobie est un phénomène éminemment écologique/contextuel : c'est une maladie qui s'attrape par contact direct, par voisinage. Autrement il serait difficile d'expliquer certains résultats électoraux durant l'époque la plus récente.

En conclusion, cette analyse basée sur des données agrégatives démontre que les moyennes nationales, particulièrement ceux qui concernent les relations entre religion, classe sociale et orientation politique sont au niveau national largement fallacieuses. Les coefficients de corrélation à l'échelon national cachent les relations causales, et ne peuvent pas refléter la diversité des comportements.

Il serait théoriquement possible de reproduire des enquêtes par sondage dans divers contextes sociaux, mais la multiplication des enquêtes serait très coûteuse, tant en énergie qu'en argent. Il y a toutefois un moyen d'éviter à la fois la *individualistic fallacy* et la *ecological fallacy* : c'est l'analyse par décile ici proposée. « Rien n'est plus menteur qu'un moyen » disait Proudhon. Mais si on replace les moyennes dans leurs contextes sociaux, elles deviennent valides.