

# Congrès AFSP Paris 2013

## **METD « Le futur des études électorales en France »** Joël Gombin, CURAPP-UPJV, joel.gombin@u-picardie.fr

### *Pour une ANES à la française, contextuelle et multiniveau*

La communauté scientifique des électoralistes a besoin d'une grande enquête de référence, comme beaucoup de nos voisins en possèdent depuis longtemps. Plusieurs contributions présentées dans ce METD dessinent les contours de ce que pourrait être son cahier des charges, et les enquêtes menées par le CEVIPOF ont pour partie joué ce rôle. De ce point de vue, l'époque paraît particulièrement stimulante, tant les innovations méthodologiques sont nombreuses et prometteuses.

Il reste toutefois à les asseoir institutionnellement. En effet, le champ académique français est aujourd'hui fragmenté, donnant lieu à de nombreuses initiatives auxquelles la cumulativité fait parfois défaut. De plus, il n'existe pas de grande institution désireuse de soutenir sur le long terme une série d'enquêtes électorales. Le ministère de l'Intérieur pourrait jouer ce rôle, mais alors même que la diffusion des résultats électoraux s'est sensiblement améliorée ces dernières années, il n'a pas accepté jusqu'à présent de le faire – sans doute pour ne pas courir le risque d'être accusé d'une utilisation politique ou policière de l'enquête. L'INSEE semble se recentrer sur les statistiques économiques et sociales, et la seule enquête politique que l'institut mène – l'enquête « participation » – semble progressivement, sinon laissée en déshérence, du moins désinvestie. Peut-être un consortium inter-universitaire pourrait-il se charger de la gestion, sur le long terme, d'une telle enquête.

### *De l'utilité d'une enquête contextuelle*

La sociologie électorale, et plus largement la science politique française, sont marquées par une controverse ancienne et pérenne concernant la *survey research*.

Les uns soulignent – à raison – que celle-ci est seule à même d’éclairer les mécanismes individuels de la décision électorale, tandis que les autres dénoncent – non sans arguments – l’illusion de la neutralité technologique d’un instrument qui conduirait en réalité à créer son propre objet (l’opinion) et à désencastrer les individus des environnements sociaux et géographiques dans lesquels ils évoluent. Au cours des dernières années, cette controverse – qui semblait naguère s’essouffler, les partisans de la *survey research* s’étant imposés de fait – a connu de nouveaux développements, un certain nombre de chercheurs plaçant pour – et pratiquant – une analyse contextuelle des comportements politiques et notamment électoraux (Jadot et van Egmond, 2003 ; Bussi et Badariotti, 2004 ; Braconnier et Dormagen, 2007 ; Rivière *et al.*, 2009 ; Gombin et Mayance, 2009 ; Braconnier et Lehingue, 2010 ; François et Magni-Berton, 2013). Ces auteurs, sans nécessairement refuser par principe le principe de l’enquête par questionnaire mené sur un échantillon aléatoire ou représentatif, plaident pour une meilleure contextualisation des individus enquêtés – même si les définitions opérationnelles de ces contextes peuvent différer. La recherche française renoue ainsi avec une riche tradition méthodologique et théorique, inaugurée par Siegfried (1913) et dont Goguel (1951) ; Klatzmann (1957) ; Dogan et Rokkan (1969) ; Derivry et Dogan (1971) ; Michelat (1975) ; Mayer et Michelat (1981) ; Derivry et Dogan (1986) ; Mayer (1987) sont les plus illustres représentants.

La contextualisation des données recueillies permet ainsi une meilleure compréhension des comportements électoraux. Le design de l’enquête devra donc s’attacher à recueillir des données contextuelles – et, sur ce point, la participation pluraliste de différentes équipes de recherche est souhaitable pour enrichir les définitions de ce contexte – ou, à tout le moins, à permettre de mettre en relation les données individuelles recueillies avec des données contextuelles, au niveau le plus fin possible (par exemple, pour les données du recensement général de la population mené par l’INSEE, au niveau de l’IRIS, voire même du bloc<sup>1</sup>). L’idéal serait ainsi d’avoir une simple géolocalisation – techniquement devenue très facile –, permettant ensuite de « matcher » l’individu avec des données contextuelles. Celles-ci peuvent être sociologiques, mais aussi politiques, comme les résultats électoraux de son bureau de vote ou des données (le cas échéant longitudinales) relatives à la participation dans ledit bureau. La collecte, lorsque c’est nécessaire, de ces données au moment de la réalisation de l’enquête abaisserait significative-

---

<sup>1</sup>L’enquête Emploi de l’INSEE offre ainsi un exemple intéressant de cas où les coordonnées des individus enquêtés sont connues au niveau du bloc – en outre, cette enquête est menée selon un principe de grappe, chaque bloc étant ainsi enquêté en totalité, sans préoccupation d’échantillonnage aléatoire.

ment les coûts de la recherche et remplirait une véritable mission de service public, en rendant ces données disponibles à l'ensemble de la communauté scientifique. Les contextes explorés peuvent également être familiaux, en renonçant éventuellement à un échantillonnage aléatoire ou, plus couramment, pseudo-aléatoire en faveur d'un échantillonnage en grappe : ainsi, tous les membres d'une même famille pourraient être interrogés. À minima, il semble nécessaire de recueillir des données déclaratives de l'enquêté sur son contexte familial, sur les modèles des « attaches » (ouvrières, paysannes...) autrefois collectées.

On pourrait ainsi imaginer que l'enquête soit livrée avec son jeu de données contextuelles. L'European Social Survey est ainsi livré avec son jeu de données agrégées multiniveau ainsi qu'avec une base d'événements renseignée par les contributeurs de chacun des pays participants. La tâche des chercheurs en serait ainsi grandement facilitée, les coûts liés à la constitution de bases de données en seraient fortement réduits.

### ***De l'utilité d'un échantillonnage permettant une analyse multiniveau***

Au nombre des innovations méthodologiques importantes qu'ont connues les études électorales au cours des dernières années, la modélisation multiniveau (connue également sous le nom de modèles hiérarchiques, ou de modèles mixtes) occupe une place importante. Les travaux de nombreux chercheurs américains (voir par exemple Gelman *et al.* (2007) ; Gelman (2008) ; Ghitza et Gelman (2013)) mobilisent désormais avec profit ce type de modélisation, permettant ainsi de montrer comment les relations entre variable dépendante et variables indépendantes s'articulent de manière différente dans l'espace, voire dans le temps (Ignazi et Wellhofer, 2008).

Ces analyses reposent généralement sur des données individuelles (issues de sondage) et écologiques (contextuelles). Or, cela est permis, dans le cas américain, par le fait que le système fédéral, et la taille du pays, ainsi que le fort intérêt accordé par les médias comme par les universités aux élections présidentielles, ont comme conséquence l'existence de données riches, souvent stratifiées par État. Andrew Gelman a ainsi montré l'intérêt scientifique qu'il y a à analyser les données État par État. Or, dans le cas français, les enquêtes permettent très rarement, si ce n'est jamais, de répliquer ce type de design, de sorte que l'analyse multiniveau trouve surtout à s'appliquer à des données agrégées (Gombin et Mayance, 2010), dans la lignée de Derivry et Dogan (1971) (qui n'utilisent certes pas de modèles multiniveaux, mais dont toute la logique de raisonnement sur le plan statistique les en rapproche).

Il est donc fortement souhaitable que la stratégie d'échantillonnage et de stratification des futures enquêtes permettent une analyse multiniveau dans de bonnes conditions. Cela suppose toutefois que le nombre d'unités de niveau 2 (« espaces de référence », dans le langage de Derivry et Dogan (1971)) soit suffisamment élevé (20 semble un minimum, entre 50 et 100 est encore plus souhaitable), et que le nombre d'unités de niveau 1 (les individus) par unités de niveau 2 soit également suffisamment élevé (là encore, 20 me semble être un minimum). La difficulté vient bien sûr du fait que la taille des unités de niveau 2 étant variable (et d'un facteur pouvant être très élevé selon le type d'unités considéré), le nombre total d'individus à interroger pour en avoir suffisamment dans chaque unité de niveau 2 peut être très élevé, et donc le coût également. À titre d'exemple, l'écart du nombre d'inscrits, en 2012, au sein des 95 départements métropolitains va de 1 à 30, ce qui supposerait que pour avoir 20 enquêtés dans le département le moins peuplé, il en faudrait 600 dans le département le plus peuplé, ce qui représenterait au total un échantillon de presque 14 000 individus – ce qui n'est pas totalement irréaliste mais, soyons lucides, très peu probable.

La difficulté n'est toutefois pas aussi élevée qu'il y paraît, dès lors qu'on sort du schéma de pensée classique du sondage aléatoire. Rien n'oblige en effet, selon le plan de sondage retenu, à chercher à tout prix à sonder de la même manière toutes les unités de niveau 2. Le taux de sondage peut ainsi être plus élevé dans les unités de niveau 2 les moins peuplées, dès lors que la pondération en tient compte au stade de l'analyse. Cette pratique permettrait ainsi d'imaginer un plan de sondage dans lequel, par exemple, les 22 régions ou les 95 départements métropolitains soient représentés, sans engendrer des coûts par trop exorbitants. À titre d'exemple, en décidant de sous-pondérer d'un facteur 2 les enquêtés issus du seul département le moins peuplé (la Lozère), on diviserait par deux le nombre total d'individus à interroger (car le deuxième département le moins peuplé est deux fois plus peuplé). On peut évidemment complexifier ce plan de sondage, pour arriver à un nombre total d'enquêtés acceptable en fonction des contraintes financières – en gardant toutefois présent à l'esprit qu'en diminuant le nombre d'enquêtés, on augmente les erreurs type...

La question la plus sensible est sans doute celle du choix des unités en fonction desquelles cette stratification doit se faire. Derivry et Dogan (1986, 1971) ont démontré l'intérêt du département ; mais c'était sur des données anciennes, et il n'est pas certain que cet espace soit encore le plus adéquat. Aussi de nombreux collègues plaideront-ils sans doute pour le choix d'autres espaces de référence, peut-être moins administratifs et davantage fondés sur une réelle pertinence politique, économique ou sociale. Les aires urbaines ou métropolitaines pourraient

ainsi jouer ce rôle, de même que les bassins d'emploi, ou d'autres découpages encore. Il est clair en tout cas qu'il n'y a pas dans le cas français d'équivalent des États américains ; c'est à la fois un inconvénient et une chance : la communauté scientifique devra donc choisir elle-même le ou les découpage(s) qu'elle jugera le(s) plus pertinent(s).

En poussant la logique de cette proposition jusqu'à son terme, on pourrait imaginer une enquête renonçant à l'ambition de couvrir l'ensemble du territoire français, mais qui procéderait plutôt par sélection itérative, aléatoire ou non, d'unités emboîtées les unes dans les autres (régions, départements, communes, IRIS, blocs d'immeubles, immeubles, foyers...), avec le choix d'un taux de sondage dans chaque unité pouvant varier d'un taux très faible jusqu'à l'exhaustivité (au niveau du foyer par exemple). Après tout, l'intérêt d'une telle enquête n'est pas de connaître la distribution univariée des variables (ce qui est après tout l'objet principal d'un sondage préélectoral), connue par ailleurs, mais bien de mieux connaître leur distribution conjointe, dans des espaces de référence donnés. Il importe donc que le plan d'échantillonnage retenu obéisse à des préoccupations scientifiques clairement explicitées *ex ante*.

### **Références**

Braconnier, Céline et Dormagen, Jean-Yves, *La démocratie de l'abstention : Aux origines de la démobilisation en milieu populaire*, Paris, Editions Gallimard, 2007.

Braconnier, Céline et Lehingue, Patrick, *Une autre sociologie du vote : les électeurs dans leurs contextes : bilan critique et perspectives*, coll. LEJEP, Paris, Lextenso/LEJEP, 2010.

Bussi, Michel et Badariotti, Dominique, *Pour une nouvelle géographie du politique : territoire, démocratie, élections*, coll. Villes-géographie, Paris, Anthropos, 2004.

Derivry, Daniel et Dogan, Mattei, « Unité d'analyse et espace de référence en écologie politique. Le canton et le département français », *Revue française de science politique*, vol. 21, n° 3, 1971, p. 517–570.

Derivry, Daniel et Dogan, Mattei, « Religion, classe et politique en France. Six

- types de relations causales », *Revue française de science politique*, vol. 36, n° 2, 1986, p. 157–181.
- Dogan, Mattei et Rokkan, Stein (dir.), *Quantitative ecological analysis in the social sciences*, Cambridge, M.I.T. Press, 1969.
- François, Abel et Magni-Berton, Raul, « Les dimensions contextuelles de la tolérance à l'égard des immigrés en France : effets de contact et de compétition », *Revue française de sociologie*, vol. Vol. 54, n° 1, 2013, p. 53–82.
- Gelman, Andrew, *Red state, blue state, rich state, poor state : why Americans vote the way they do*, Princeton, Princeton University Press, 2008.
- Gelman, Andrew, Shor, Boris, Bafumi, Joseph et Park, David, « Rich state, poor state, red state, blue state : What's the matter with Connecticut », *Quarterly Journal of Political Science*, vol. 2, n° 4, 2007, p. 345–367.
- Ghitza, Yair et Gelman, Andrew, « Deep Interactions with MRP : Election Turnout and Voting Patterns Among Small Electoral Subgroups », *American Journal of Political Science*, 2013, p. no–no.
- Goguel, François, *Géographie des élections françaises : de 1870 à 1951*, Paris, A. Colin, 1951.
- Gombin, Joël et Mayance, Pierre (dir.), *Droit(es) aux urnes en région PACA ! L'élection présidentielle de 2007 en région Provence-Alpes-Côte d'Azur*, Cahiers politiques, Paris, L'Harmattan, 2009.
- Gombin, Joël et Mayance, Pierre, « Tous conservateurs ? Analyse écologique du vote de la population agricole lors de l'élection présidentielle de 2007 », in Nonna Mayer, Pierre Muller, Bertrand Hervieu, Jacques Rémy et François Purseigle (dir.), « Les mondes agricoles en politique », Paris, Presses de Sciences Po, 2010, p. 193–216.
- Ignazi, Piero et Wellhofer, Spencer, « Class, Property, and Religion : Long-Run Trends on Secularization and Politics in Italy », Barcelone, 2008.
- Jadot, Anne et van Egmond, Marcel, « Réconcilier l'individuel et le contextuel ? L'intérêt de la méthode multiniveaux en recherche électorale », *Revue de la Maison française d'Oxford*, vol. 1, n° 1, 2003.

- Klatzmann, Joseph, « Comportement électoral et classe sociale. Etude du vote communiste et du vote socialiste à Paris et dans la Seine », *in* Maurice Duverger, François Goguel et Jean Touchard (dir.), « Les élections du 2 janvier 1956 », Cahiers de la FNSP, Paris, Armand Colin, 1957, p. 254–285.
- Mayer, Nonna, « De Passy à Barbes : deux visages du vote Le Pen à Paris », *Revue française de science politique*, vol. 37, n° 6, 1987, p. 891–906.
- Mayer, Nonna et Michelat, Guy, « Les choix électoraux des petits commerçants et artisans en 1967. L'importance des variables contextuelles », *Revue française de sociologie*, vol. 22, n° 4, 1981, p. 503–521.
- Michelat, Guy, « Vote des groupes socio-professionnels et variables contextuelles », *Revue française de science politique*, vol. 25, n° 5, 1975, p. 901–918.
- Rivière, Jean, Hérin, Robert et Bussi, Michel, *Le pavillon et l'isoloir. Géographie sociale et électorale des espaces périurbains français (1968-2008). A travers les cas de trois aires urbaines moyennes (Caen, Metz et Perpignan)*, Doctorat, Université de Caen Basse Normandie, Caen, 2009.
- Siegfried, André, *Tableau politique de la France de l'Ouest sous la troisième République*, Paris, Armand Colin, 1913.