

BENOÎT RIANDEY

Échanges sur les sondages électoraux

Journal de la société statistique de Paris, tome 139, n° 3 (1998),
p. 61-73

http://www.numdam.org/item?id=JSFS_1998__139_3_61_0

© Société de statistique de Paris, 1998, tous droits réservés.

L'accès aux archives de la revue « Journal de la société statistique de Paris » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

ÉCHANGES SUR LES SONDAGES ÉLECTORAUX

Benoît RIANDEY

Institut National d'Etudes Démographiques *

Les sondages électoraux ont conquis l'expression la plus médiatisée de la statistique d'enquête en France. C'est surtout par eux que les instituts de sondage privés accèdent à la notoriété du public. Leurs rares échecs font ressortir les difficultés de cet exercice, mais les débats sont plus intenses dans les quotidiens que dans les milieux scientifiques. Cependant, les questions soulevées sont honorables, qu'il s'agisse de la modélisation des erreurs de mesure ou de celle des erreurs de sondage. Mais elles requièrent de profondes connaissances de sciences politiques à côté de compétences en statistique, et le dialogue n'est pas toujours suffisamment établi entre ces disciplines.

Le thème est ici visité en reprenant le grand débat public provoqué par l'échec des prévisions électorales de mai 1997. Cette discussion est aussi l'occasion d'évoquer les différentes sortes d'enquêtes politiques réalisées en France.

INTENTIONS DE VOTE ET SIMULATION DES SIÈGES : LES ÉLECTIONS LÉGISLATIVES DE 1997

Le 25 mai 1997, la surprise fut grande à l'issue du premier tour des élections législatives et, la semaine suivante, la liste présidentielle a perdu la majorité à la Chambre des députés. Les pronostics convergents fondés sur les sondages publiés avant le premier tour ont ainsi été infirmés, suscitant à nouveau un débat sur la fiabilité des outils statistiques employés.

Sous la plume de Jacques Antoine et de Marie-Thérèse Antoine-Paille, *Le Sondoscope*¹ a récapitulé dès juin 1997 les conclusions des sondages publiés pendant quatre semaines de campagne : l'avance de la Droite en sièges s'est

* 133 Boulevard Davout, 75980 Paris Cédex 20

e-mail : riandey@ined.fr

1. *Le Sondoscope, la revue française des sondages* est une revue mensuelle réalisée par la société CESEM Opinion (27 rue Taitbout, Paris 9ème) avec le concours du Centre d'études de la vie politique française contemporaine, ce laboratoire associant le CNRS et la Fondation Nationale des Sciences Politiques. *Le Sondoscope* recense et analyse l'ensemble des enquêtes par sondage publiées en France. Il retrace l'actualité de la statistique privée comme le fait *Le Courrier des Statistiques* pour la statistique publique.

toujours montrée positive pour les 24 projections établies du 18 avril au 16 mai 1997 (*Le Sondoscope*, juin 1997, p. 12 à 14). Toutefois, cette convergence des pronostics était ternie par de fortes variations aux mêmes dates : les sondages réalisés les 22-23 avril prévoyaient une avance de la Droite s'étageant de 8 à 110 sièges, ceux des 9 et 10 mai, de 52 à 133 sièges. Cette amplitude ne résulte pas des méthodologies spécifiques à chaque institut, car, en quatre semaines, les pronostics de chaque institut ont respectivement connu des sauts de 72, 76, 85, 124 sièges. BVA fait exception avec une amplitude réduite à 18 sièges pour une avance limitée de la Droite comprise entre 34 et 52 sièges. Enfin, aucune projection de Louis Harris n'a été publiée au cours des deux premières semaines.

Ces variations auraient pu n'être que le reflet de l'impact de la campagne que M-Th. Antoine-Paille récapitulait ainsi : après la dispersion des estimations des deux premières semaines, une moindre dispersion en troisième semaine marquée par une détérioration des reports du Front National sur la Droite modérée. Puis des transferts plus favorables à la Droite, enfin une réduction des voix de Droite avec une moindre productivité des reports en sa faveur, ce qu'ont confirmé les sondages non publiés de la semaine précédant le vote.

Mais, *Le Sondoscope* signale que l'avance en sièges pronostiquée est bien peu corrélée au score de la Droite complète au même sondage, constat qui paraît manquer de cohérence. La presse écrite avait souligné cette absence de cohérence (par exemple Nicole Gauthier dans *Libération*, le 5 mai 1997). Elle avait aussi présenté des résultats plus nuancés que la seule estimation des points moyens. Ainsi, le 17 avril, peu avant la dissolution de l'Assemblée, *Paris-Match* publiait des fourchettes de sièges de BVA presque recouvrantes (RPR-UDF-div Dr : 234-306, PS-div G : 231-283, PC : 20-25), plutôt favorables à la Gauche. Autre exemple, l'*Express* du 1er mai indiquait des seuils de basculement de majorité très accessibles, compte tenu des incertitudes d'échantillonnage des sondages et de la proportion d'électeurs indécis.

En définitive, «des projections en sièges très aléatoires» (*Libération*, 5 mai 1997) et un biais commun aux instituts pendant ces quatre semaines de campagne justifiaient un débat sur les méthodes de projection en sièges. De nombreux politologues et spécialistes des instituts y ont activement participé, le 2 décembre au séminaire ASU-INED.

MÉTHODE DE SIMULATION DES SIÈGES

De façon très éclairante, le politologue Jérôme Jaffré a décortiqué les facteurs de variation dont la résultante conduit au constat critique de Marie-Thérèse Antoine-Paille. Il a insisté sur toutes les tâches de politologue n'intégrant pas de données d'enquêtes, qui font de ces projections davantage un exercice de science politique que de statistique de sondage. En premier lieu, pour chaque circonscription, on prend connaissance du vote antérieur à une élection comparable. Il faut tenir compte des modifications des circonscriptions ou du mode de scrutin (comme la proportionnelle en 1986). La stabilité des circonscriptions induit une forte stabilité des candidatures.

ÉCHANGES SUR LES SONDAGES ÉLECTORAUX

On corrige alors ce fichier de référence en affectant des primes aux sortants à la suite de leur première élection (environ 3 %) et des décotes aux battus.

Le système des forces politiques a pu se modifier. Il faut alors intégrer les nouvelles forces absentes de l'élection de référence. On peut alors procéder à la mesure des intentions de vote grâce aux sondages. Les instituts ont abandonné le recours aux intentions sur étiquettes partisans et indiquent aux enquêtés le nom des candidats. De façon pratique, on substitue progressivement aux noms des candidats antérieurs celui des candidats réels qui se déclarent jusqu'à huit jours avant l'élection. La multiplication tardive des petites candidatures induit une sous-estimation du vote émietté jusqu'à la dernière semaine, mais en fait, les politologues sont forcés de réduire le champ des forces politiques (à 6 dans le modèle SOFRES de Jean Chiche et Jérôme Jaffré).

A partir du sondage national, par un calcul d'écart ou d'indice, on actualise le rapport de forces dans chaque circonscription. Ainsi, par exemple, le PS est remonté de 19 % à 28 % des suffrages de 1993 à 1997.

On introduit ensuite les matrices de report des voix sur le second tour. Les duels PS-Droite modérée constituent 70 % des configurations (Tableau 1). L'estimation des reports par sondage national y est alors solide. Les triangulaires PS-Droite modérée-FN se sont élevées en 1997 à 12 % des configurations. Les transferts peuvent également être simulés par une enquête nationale. Par contre, certaines autres situations n'ont aucune signification nationale. Un sondage national ne permet pas à lui seul d'estimer les reports qui s'effectueront dans ces quelques situations locales particulières. Les reports y sont évalués plutôt sur la base de la sociologie électorale et par appréciation du climat politique.

TABLEAU 1
Les configurations politiques du second tour

Duels	Socialistes-div G/Droite modérée	369
	Verts / Droite modérée	14
	PC / Droite modérée	16
Duels	Gauche-Verts/Front National	25
	Droite modérée/Front National	31
Triangulaires :	Socialistes-div G/Droite modérée/FN	68
	Verts-PC/Droite modérée/FN	8
Élus du premier tour et autres configurations		24
Total métropole		555

D'après J. Jaffré, séminaire INED-ASU du 2 12 97.

Le modèle ainsi mis en place, le calcul peut commencer. On détermine d'abord les élus du premier tour. Les erreurs faites à ce stade ont peu d'impact car il s'agit de circonscriptions sans surprise possible au second tour.

La règle des 12,5 % des inscrits sélectionne alors les candidats admis au second tour. Le niveau d'abstention joue là un rôle majeur car un fort taux d'abstention élimine du second tour un bon nombre de candidats. Les accords politiques et les résultats du sondage déterminent la matrice des transferts appliquée à chaque siège. Le programme affecte chaque siège à un candidat gagnant et totalise les sièges de chaque formation.

Le programme repère les sièges sensibles, ceux attribués avec moins de 51 % d'intentions de vote. Leur nombre permet d'établir une fourchette d'incertitude du pronostic.

RÉSULTAT DES SIMULATIONS SELON LE MODÈLE SOFRES

Le modèle de la SOFRES a permis de quantifier l'impact des différents transferts de voix en cours de campagne. Juste avant la dissolution, la SOFRES estimait l'avance de la Droite en métropole à 95 sièges (325 pour la Droite contre 230 pour la Gauche). Au vu du premier tour, la Droite modérée a perdu 3 % des intentions de vote (au profit des écologistes et inclassables, et, faiblement, de l'Extrême-Gauche et du Front National). De ce seul fait, elle aurait ainsi perdu 37 sièges, gagnés par la Gauche (288-267).

Mais l'impact de ce même transfert de voix peut être évalué dans une simulation prenant en compte la nouvelle matrice de reports, celle estimée juste avant le deuxième tour. Au cours de la campagne, le report s'est, en effet, amélioré des Verts vers le PS, et détérioré du Front National vers la Droite modérée (Tableau 2). A elle seule, cette structure plus favorable à la Gauche aurait réduit l'écart initial de 22 sièges (314-241), puis, dans cette structure de reports, les 3 % de transferts de voix auraient fait perdre 50 sièges à la Droite au profit de la Gauche. La conjugaison de ces deux facteurs induit, selon le modèle, une victoire de la Gauche (264-291).

Le modèle calé sur la structure des votes au premier tour et sur les intentions de report à la veille du second conduit à un écart minime à la réalité : un seul siège pour la Gauche et 2 pour la Droite modérée. Le modèle est donc un outil de simulation parfaitement adapté.

Le modèle permet d'étudier la sensibilité du vote à certains facteurs. Ainsi, à la veille du premier tour (*Le Monde*, 18 mai), le modèle de la SOFRES évaluait — parmi d'autres hypothèses — qu'une abstention de 30 % avec un Front National à 15 % permettrait 114 maintiens du Front National au deuxième tour et induirait la perte de 35 sièges pour la Droite (abstention réelle : 31,5 % ; FN à 15,3 et 124 maintiens possibles du FN). En l'absence donc de maintien du Front National, le modèle prévoyait sous ces hypothèses 281 sièges pour la Droite modérée et 275 pour la Gauche : un résultat très serré.

TABLEAU 2

Evolution des pronostics de transfert des voix au deuxième tour
en cas de duels Socialistes-div G/Droite modérée

	Socialistes- div G	Droite modérée	Abstention blanc nul	
1 Transferts estimés avant la dissolution				
Écologistes	54	28	18	100 %
Front National	24	55	21	100 %
2 Transferts estimés juste avant le deuxième tour				
Écologistes	62	14	24	100 %
Front National	21	50	29	100 %

D'après J. Jaffré, exposé du 2 12 97

Ce modèle constitue également un précieux outil d'analyse : Jérôme Jaffré en a déduit les conditions nécessaires à une victoire de la Droite modérée à la suite du premier tour. Elles sont de trois ordres :

- une amélioration de la mobilisation de la Droite modérée (96 % de report des voix acquises au premier tour au lieu de 93 %);
- un meilleur report des voix du FN vers la Droite modérée (60 % au lieu de 50 %);
- un meilleur report des voix écologistes vers la Droite modérée (24 % au lieu de 14 %).

Le modèle évalue à 7, 15 et 16 sièges les gains respectifs pour la Droite associés à chacun de ces facteurs. La conjonction de ces trois facteurs se serait avérée nécessaire pour assurer la victoire de la Droite modérée. Ce ne fut pas le cas.

DÉBAT

Le modèle dégage donc un grand nombre de facteurs affectant le résultat du vote. C'est cette multiplicité qui induit l'apparente incohérence des simulations, relevée par *Le Sondoscope* : le niveau des suffrages de la Droite ne suffit pas à faire l'élection. La comparaison de simulations appelle une analyse beaucoup plus complète.

L'impact des primes ou décotes aux sortants relève d'une appréciation politologique très difficile à individualiser et à quantifier. Jérôme Jaffré illustre cette question par deux exemples : plusieurs instituts ont fait de grosses erreurs dans l'estimation du vote communiste. Le regain du vote socialiste depuis 1993 aurait dû faire basculer les députés communistes nouvellement élus à cette date. Seule, la prime aux sortants explique leur maintien devant le candidat socia-

liste en 1997. Au contraire, cette prime a peu joué au profit des nouveaux élus RPR-UDF et a donc conduit à une surestimation du nombre d'élus de Droite.

Pendant la campagne, les simulations se heurtent de plus à la faible détermination des votes. La proportion d'électeurs sûrs d'aller voter a même décru (de 75 à 71 % selon les sondages BVA). Selon IPSOS, la proportion d'électeurs estimant pouvoir changer d'avis était de 40 % la dernière semaine. S'ajoutent les indécis. Ainsi du fait des filtres successifs, les intentions de vote demandées à 1000 enquêtés reposent en fait sur 500 à 600 d'entre eux, estime Marie-Thérèse Antoine-Paille. Les indicateurs de mobilisation et de sûreté du vote mis en place par BVA ont heureusement permis de définir des fourchettes très illustratives de la fragilité de l'estimation.

L'interdiction de publication pendant la dernière semaine accroît la discordance entre prévisions et vote, mais en 1995, suggère Jérôme Jaffré, après même être allé voter, les électeurs de Droite hésitaient toujours entre Balladur et Chirac. Selon les informations communiquées en octobre au colloque de Communication publique, cette interdiction – aujourd'hui intenable – pourrait être abolie. L'indétermination de fond n'en sera pas pour autant levée.

Mais surtout, la sensibilité des simulations de sièges tient au scrutin à deux tours, cette machine à fabriquer de la majorité. Or les écarts entre les sondages et le vote vont dans le même sens pour tous les instituts. Au-delà de la variance, il y a donc de véritables biais. Ce problème sérieux n'a pas de réponse scientifique : la sous-estimation du vote Front National doit être corrigée par un redressement, mais ce redressement se heurte à la mauvaise reconstitution du vote passé. Cette correction du vote FN au détriment de la Droite a été insuffisante. De même, le manque de fermeté des écologistes dans leur choix est une difficulté par essence irréductible. Ainsi, avec Eric Dupin, certains chercheurs estiment qu'il n'était pas possible de faire des estimations en sièges rigoureuses.

Les sondages et estimations préélectorales sont donc bien fragiles en comparaison des estimations produites par les instituts de sondage au cours du dépouillement des votes (cf. annexe 1). Les Sondages Sorties des Urnes (SSU) présentent également des difficultés. Réalisés à la sortie des bureaux de vote auprès des seuls votants, ils fournissent le soir même du vote des résultats sur les motivations du vote des électeurs, mais leurs biais de non-réponse invalident toute estimation du résultat des élections qui en serait issu. En fait, on cale les SSU nationaux sur des « estimations » dès que celles-ci sont disponibles en soirée (avant même 20 heures). Le public n'a donc accès qu'à une version déjà consolidée des résultats du SSU. Leur fragilité antérieure n'affecte donc – en connaissance de cause – que les états-majors politiques et médiatiques avant le fatidique instant des 20 heures. La déontologie des enquêtes et des médias en est épargnée.

Les instituts ont incontestablement fait un effort de pédagogie, mais en une page d'un hebdomadaire, on ne fait pas le tour de cette mécanique complexe. Les médias audiovisuels ne peuvent assortir les estimations énoncées de précautions d'interprétations suffisantes. Au cours des journaux télévisés, ils présentent généralement directement la simulation de l'hémicycle – le

«camembert» des sièges – construite à partir des seuls points moyens. Cette présentation exclut le doute pour un objet qui n'a pas le statut de fait objectif. Devraient-ils, par déontologie, renoncer à ce type «d'information»? La pression médiatique est forte en faveur des prévisions en sièges. La pression politique également, en particulier du fait de la préparation du second tour car les pronostics donnent corps au vote utile. Les spécialistes sont-ils en situation de résister à de telles pressions? Mais peut-on priver le public des résultats de travaux d'un tel intérêt pour la seule raison que ces recherches fragiles sont pleinement d'actualité?

Les électeurs bénéficient cependant de plusieurs garanties. La Commission des sondages (*cf.* annexe 2) vérifie le sérieux du travail de sondage et la continuité des méthodes de redressement et de projection. La concurrence multiplie les publications : les fluctuations d'échantillonnage en sont «globalement» réduites et les estimations imprudentes soulignées. Enfin, le débat technique — comme la note publiée en 1996 par Pierre Giacometti (IPSOS) dans *Le Point* — donne conscience de la fragilité des méthodes. La proposition de Jérôme Jaffré d'un «dépôt légal des sondages», consultable après la publication, constitue une avancée déontologique majeure puisque sa réalisation permettrait l'auto-contrôle de la profession. Souhaitons son aboutissement, car son cadre potentiel existe déjà, par exemple, avec la banque de données politiques du Centre d'informatisation des données socio-politiques (CIDSP)² à Grenoble.

L'issue du scrutin a donc figuré parmi celles annoncées comme statistiquement possibles, mais, pendant la campagne, elle n'est jamais apparue comme la plus probable, et des pronostics d'avance de 155 sièges (aux 9-10 mai) n'incitent pas le lecteur à un jugement prudent. Les statisticiens traduisent la dialectique du probable parmi les possibles à travers des intervalles de confiance. Cet outil mis à mal par la technique des quotas est en principe accessible à la micro-simulation grâce aux techniques de réplication d'échantillons, telles que le bootstrap. Mais chaque jeu d'hypothèses induit une fourchette et elles sont nombreuses, sans même prendre en compte la dynamique de la campagne électorale. La fourchette empirique établie par IPSOS tient compte des sièges menacés et non de l'aléa du sondage. Elle intègre ainsi les incertitudes liées aux hypothèses contrairement aux intervalles de confiance des statisticiens, toujours conditionnés au modèle, mais on ne saurait lui donner une interprétation probabiliste : il faudrait introduire un genre de modélisation bayésienne des hypothèses et, de plus, les erreurs faites sur les estimations de chaque circonscription sont très loin d'être indépendantes. Cet écueil est sans doute inhérent à tout modèle de micro-simulation qui insère les résultats d'une ou plusieurs enquêtes dans un système d'information plus vaste. Ainsi, la détermination du probable relève certainement davantage du jugement du politologue ou de l'économiste que de la technique du statisticien d'enquête.

2. Ce laboratoire associant le CNRS et l'Institut de Sciences Politiques de Grenoble, archive les données individuelles anonymes de très nombreuses enquêtes électorales françaises. Ce laboratoire est également le siège du DESS PROGIS renommé dans le domaine des enquêtes privées.

ANNEXE 1

Les différents types de sondages électoraux

Il est utile de distinguer les différents types de sondages électoraux évoqués au cours de cette note : les sondages d'intentions de vote, les estimations (des résultats électoraux), les sondages Sorties des urnes (SSU), les sondages post-électoraux, les sondages sur la participation électorale. Précisons ces notions et leurs implications techniques.

Les Estimations : les résultats de votes issus d'échantillons hyper-stratifiés de bureaux de vote – clos et dépouillés en premiers – permettent d'estimer les résultats de l'élection avant l'achèvement de son dépouillement, et même avant la clôture du vote dans les grandes villes où les bureaux de vote ferment tardivement (20 heures). Au fur et à mesure du dépouillement, la largeur des fourchettes se réduit. Il ne s'agit pas de données déclaratives. Elles sont donc exemptes d'«erreurs», si ce n'est de transmission. La précision des fourchettes bénéficie de l'information auxiliaire considérable apportée par l'historique des résultats des élections antérieures dans ces mêmes bureaux de vote. Ces méthodes ont démontré leur fiabilité.

Les sondages d'intentions de vote : il s'agit des enquêtes par questionnaire réalisées tout au long de la campagne électorale. La loi de 1977 soumet la publication de leurs résultats au contrôle de la Commission des sondages (cf. annexe 2), mais interdit leur publication au cours de la semaine précédant le vote. Les difficultés de ces enquêtes ont fait l'objet de nombreux commentaires dans la presse : les échantillons bruts sont très biaisés en raison des refus différentiels selon les appartenances politiques, plutôt plus forts pour les électeurs de Droite, mais surtout pour les partis extrêmes. Si la sous-représentation ancienne des électeurs communistes s'est fortement estompée, celle des électeurs du Front National est actuellement massive. Les questions d'intention de vote posées au cours des *enquêtes omnibus* (enquêtes multi-thèmes et multi-clients) sont moins sujettes à ce biais d'échantillonnage directement lié au sujet d'intérêt. Ces enquêtes sont, en effet, introduites par une demande neutre du type «*Accepteriez-vous de répondre à un sondage sur des questions variées ?*».

Par ailleurs, les déclarations d'intention de vote diffèrent souvent du futur comportement effectif du répondant : son intention même de participer au vote est une information fragile, inégalement d'ailleurs selon les appartenances politiques. Après l'enquête, l'enquêté peut changer son choix politique, comme tout électeur tout au long de la campagne électorale, y compris au dernier moment. Il importe donc de réaliser la collecte d'une enquête politique à des moments précis et jusqu'à la veille de l'élection – ce qui met les acteurs politico-économiques et les citoyens en situation d'information inégale au cours de cette dernière semaine. Cette carence démocratique s'est muée en interdiction vaine du fait de la publication de ces sondages par les médias étrangers et sur Internet.

Cette nécessité d'une collecte «instantanée» est antinomique avec la réalisation d'un échantillon aléatoire qui exige une durée de plusieurs jours afin

d'interroger aussi les personnes sondées les moins disponibles. Cette raison technique – ainsi que des raisons moins légitimes de coût – conduisent les instituts privés à réaliser ces enquêtes sur quotas. La représentativité micro-géographique de l'échantillon (au niveau des quartiers) est très mal assurée lorsque cette enquête sur quotas est effectuée en face à face. Au contraire, le développement des enquêtes politiques par téléphone a largement remédié à ce problème, puisqu'elles reposent enfin sur une base de sondage. Toutefois, les instituts résorbent sans doute rarement le biais dû aux listes rouges (biais de couverture) et celui induit par les enquêtés absents au premier appel (biais de non réponse).

Ces mauvaises conditions d'échantillonnage et de mesure pourraient sembler compensées par la disposition remarquable d'informations auxiliaires directement liées au champ d'intérêt : le résultat des consultations politiques antérieures. Aussi les questionnaires comportent-ils toujours des questions rétrospectives sur les comportements électoraux passés de l'enquêté. Au lieu de devoir mesurer directement des intentions de vote, l'enquête estime donc uniquement les transferts de voix propres à chaque famille politique entre deux consultations successives. On dispose donc d'une opportunité exceptionnelle de réduire la variance grâce à cette post-stratification. Mais celle-ci comporte des limites : les consultations successives sont généralement de natures différentes, ce qui limite la corrélation entre la variable auxiliaire et la variable d'intérêt. (Les seconds tours d'une consultation électorale échappent à cette limitation, ce qui conforte puissamment les pronostics issus de sondages les concernant). Le calage sur cette information auxiliaire corrige non seulement les fluctuations d'échantillonnage, mais les biais de l'échantillon et les erreurs de mesure dues aux dissimulations de l'intention de vote ou aux reconstitutions erronées du vote passé. Il ne semble pas que les implications théoriques de cette correction groupée aient été clairement explicitées. Elles font en tout cas l'objet d'un débat important. Ainsi, Brian Gosschalk récusait absolument le calage sur vote antérieur pour les sondages britanniques, en raison de l'importance des erreurs de reconstitution du vote passé (*cf.* Antoine et Riandey). L'échec des pronostics électoraux de 1992 en Grande-Bretagne a conduit les instituts britanniques à une audit collective volontaire dans le cadre de la Market Research Society suivie d'un débat à la Royal Statistical Society. Ces problèmes théoriques sont loin d'être résolus, et cette difficulté des enquêtes permet au vote de rester un événement, tant il résiste à son anticipation.

Les sondages sorties des urnes : introduits en France en 1983 à l'initiative de Jean-Luc Parodi, les SSU de BVA ou de CSA fournissent le soir même du vote des résultats relatifs aux motivations du choix électoral de l'instant. Ils sont réalisés à la sortie des bureaux de vote auprès des seuls votants dans un échantillon maître de 75 à 90 bureaux de vote choisis en fonction de leur profil politique, de leur région et de leur catégorie de commune. Environ 4000 personnes sont enquêtées par l'institut CSA. Elles sont désignées par l'enquêteur au moyen d'un tirage systématique selon un certain pas (par exemple 1 votant tous les 10). Le questionnaire est auto-administré, ce qui induit 5 à 6 % de déchets. Par contre, les enquêtés s'expriment sur leur vote de l'instant précédent dans le secret de l'auto-administré ne sont pas amenés

à réinventer ou travestir leur acte. Ces sondages sont donc peu sujets aux erreurs de déclaration.

Le taux d'acceptation n'est malheureusement pas mesuré en France et, de fait, le tirage systématique est difficile à opérer rigoureusement par un enquêteur isolé. Ainsi l'accès immédiat à la base de sondage (la population des votants effectifs) ne suffit pas à éviter de très sérieux biais d'échantillonnage - au détriment par exemple des personnes âgées. Les statisticiens anglais estiment réaliser leurs SSU avec un très bon échantillon aléatoire et un taux de réponse élevé (84 %), mais n'en soulignent pas moins la qualité insuffisante (Antoine et Riandey). La technique des quotas ne serait pas d'un bon recours puisqu'elle ne s'appliquerait correctement qu'à l'univers des électeurs inscrits. L'abstention électorale, neutre pour le SSU aléatoire, constituerait une source de biais grave pour tout SSU sur quotas. Ainsi les SSU ne valorisent pas leurs conditions apparemment excellentes (bonne stratification, sondage aléatoire sur une base de sondage immédiatement accessible, les votants eux-mêmes). Les conditions psychologiques de l'enquête constituent peut-être son frein : à quoi bon passer dans un isolement pour déclarer son vote quelques minutes plus tard, même à un enquêteur tenu au secret statistique ?

Réussir un SSU implique donc d'abord éviter de se tromper d'objectif : ces sondages renseignent effectivement dès 20 heures sur la motivation du vote des votants, mais ils ne permettent pas d'estimer valablement les résultats des élections. En fait, on cale les SSU nationaux sur les *estimations* dès que celles-ci sont disponibles avant même 20 heures. Le public n'a donc accès qu'à une version déjà consolidée des résultats.

Les sondages post-électorales sont maintenant réalisés par téléphone auprès de l'ensemble de l'électorat, abstentionniste ou votant, quelques jours après le vote (*cf.* Boy et Mayer éd., pour les présidentielles de 1995) ou à la suite du premier tour pour les législatives de 1997 (*cf.* Perrineau et Ysmal éd.). Ils apportent ainsi également une connaissance essentielle sur les motivations de l'abstention, comportement majeur pour l'issue du vote. Les questionnaires post-électorales peuvent être plus consistants que les questionnaires SSU. Aussi est-ce une source privilégiée des politologues du Centre d'étude de la vie politique française (CEVIPOF).

Les sondages sur la participation électorale : l'INSEE s'interdit d'enquêter les Français sur leurs opinions politiques et donc sur leur choix électoral. Par contre, l'observation de la participation électorale ressort de la seule lecture des bordereaux d'émargement électoral. L'INSEE tire donc un échantillon d'électeurs dont il constate l'éventuelle participation électorale. En fait, il s'agit d'un panel d'électeurs suivis depuis 1995 (*cf.* Héran). Cette information longitudinale exprime donc une dynamique politique que les enquêtes précédentes ne pourraient mesurer que de façon rétrospective et donc imparfaite.

La base de sondage de ce panel est constituée de l'Échantillon Démographique Permanent de l'INSEE. A chacune de ces consultations électorales et avec l'autorisation de la Commission Nationale de l'Informatique et des Libertés (la CNIL), l'INSEE fusionne cet échantillon avec le fichier électoral. Ainsi

ÉCHANGES SUR LES SONDAGES ÉLECTORAUX

peut-il localiser le bureau de vote de chaque enquêté et effectuer un suivi très efficace de l'échantillon. L'Échantillon Démographique Permanent fournit sur l'électeur des caractéristiques socio-démographiques riches, très profitables à l'analyse politologique de l'enquête.

ANNEXE 2

Les contrôles réglementaires de sondages électoraux

Deux textes législatifs encadrent la réalisation et la diffusion des sondages électoraux : la loi sur les sondages électoraux de 1977 et la loi Informatique et libertés de 1978. Tous deux sont à l'origine d'une autorité administrative indépendante qui contrôle l'exécution de la loi.

La Commission des sondages

La Commission des sondages a été instituée par la loi de 1977 sur les sondages électoraux. Cette loi vise à protéger les citoyens contre la publication de sondages préélectorales frauduleux destinés à manipuler le corps électoral. Elle oblige à la publication de certaines caractéristiques (taille et nature de l'échantillon, date de la collecte) et a ainsi amélioré la transparence des sondages préélectorales. Elle interdit la publication des sondages politiques la semaine précédant le vote. Cette mesure est contestée tant sur le fond que pour son irréalisme actuel. Au contraire, le rôle régulateur de la commission est largement apprécié.

La commission est constituée de neuf membres issus du Conseil d'État, de la Cour de cassation et de la Cour des comptes. Elle examine les sondages électoraux avant leur publication et vérifie qu'ils ont bénéficié de tout le sérieux nécessaire, et, plus généralement, veille au respect de la loi de 1977.

La Commission nationale de l'informatique et des libertés

La Commission nationale de l'informatique et des libertés (CNIL), instituée par la loi de 1978, est chargée de veiller au respect de la protection des personnes pour tout traitement automatisé d'informations nominatives. Elle est composée de 17 membres issus du Parlement, du Conseil Économique et Social, du Conseil d'État, de la Cour de cassation et de la Cour des comptes. La Convention du Conseil de l'Europe de 1981 relative à la protection de la vie privée et la Directive européenne du 24 octobre 1995 sur la protection et la circulation des données personnelles ont instauré une telle institution dans chaque État membre.

La CNIL considère que les enquêtes par sondage, y compris celles sur quotas, appartiennent au champ de son autorité dès lors que la collecte est nominative ou indirectement nominative, même si le fichier d'analyse est strictement anonyme. De fait, les instituts exigent de leurs enquêteurs qu'ils reportent sur une feuille le nom et l'adresse des enquêtés en face de l'identifiant du questionnaire. Les clients et les lecteurs des résultats sont ainsi protégés contre la production de questionnaires falsifiés que pourrait induire l'absence de

contrôle. La Commission des sondages insiste pour que cette garantie soit apportée aux enquêtes politiques.

Ce faisant, la CNIL soulève une difficulté majeure : les opinions politiques figurent parmi les informations sensibles que la loi Informatique et Libertés n'autorise à recueillir qu'avec l'accord exprès de l'enquêté (c'est à dire son accord écrit). Cette disposition stricte est incompatible avec la collecte fiable de telles données car elle dénie la garantie apportée par le secret statistique et risque d'induire la défiance de l'enquêté. En définitive, et après un long conflit, la CNIL a accédé à la demande des instituts de sondage que l'accord exprès soit reconnu recueilli quand l'enquêté a lui-même porté son nom sur la feuille de route de l'enquêteur. Cette solution est raisonnable, mais fait ressortir l'inadaptation de la loi informatique et libertés à certains traitements statistiques. La révision de la loi Informatique et Libertés induite par la précédente directive devrait corriger ces contradictions car elle reconnaît la finalité d'intérêt public de la statistique et les garanties apportées par le secret statistique. Par contre, le recueil de l'accord exprès au cours d'une enquête téléphonique s'avère délicat. La seule solution stricte résiderait dans une gestion automatisée des appels sans intervention des enquêteurs, mais elle ne protégerait bien sûr pas des appels déloyaux déguisés en enquêtes politiques par sondage, sans doute les seuls dangereux pour les libertés. L'harmonisation du droit et de la déontologie professionnelle suppose un important sens pratique, puis sa reconnaissance.

L'auteur remercie vivement Jérôme Jaffré, Marie-Thérèse Antoine-Paille et Stéphane Rozès pour leur contribution à ce débat, ainsi que Jacques Antoine, Michel Brulé, Eric Dupin, Pierre Martin, Jean-Luc Parodi, Romain Pache, Jean-Philippe Roy.

BIBLIOGRAPHIE

- ANTOINE J., ANTOINE-PAILLE M-Th. et JADOT A. (1997), "Chronique des sondages pré-électorales français et britannique (de 1997)", *Le Sondoscope*, n° 131, juin 1997.
- ANTOINE J. et RIANDEY B. (1996), "Les sondages d'intention de vote. L'expérience britannique de 1992". *Journal de la Société Statistique de Paris*, n° 4, p. 3-22.
- ANTOINE-PAILLE M-Th. (1997), "Les motivations du vote aux élections législatives de mai-juin 1997", *Le Sondoscope*, n° 130, mai 1997.
- ASSOCIATION COMMUNICATION PUBLIQUE (1997), "La publication des sondages électoraux : quelles règles pour plus de sérénité?", Colloque du jeudi 23 octobre 1997 à l'Assemblée Nationale. Compte rendu dans *Le Sondoscope* n° 136, décembre 1997.
- BLONDIAUX L. (1998), "La fabrique de l'opinion : une histoire sociale des sondages". *Le Seuil*, Paris.
- BON F. (1974), "Les sondages peuvent-ils se tromper?". Calman-Lévy, collection Questions d'actualité, Paris.
- BOY D. et MAYER N. éd. (1997), *L'électeur a ses raisons*, Presses de Sciences Po, collection références.

ÉCHANGES SUR LES SONDAGES ÉLECTORAUX

- HÉRAN F. (1997), "Les intermittences du vote : un bilan de la participation électorale de 1995 à 1997", *INSEE première*, n° 546, sept.
- JAFFRÉ J. (1997), Nombreux articles de presse, en particulier *Le Monde*, le 18-19 mai.
- PERRINEAU P. et YSMAL C. éd. (1998), *Le vote surprise*, Presses de Sciences Po, collection Chroniques électorales.
- MARKET RESEARCH SOCIETY (1994), "The Opinion Polls and the General Election". Market Research Society.
- WORCESTER R. (1996), "Political polling : 95 % expertise and 5 % luck" *Journal of the Royal Statistical Society, A.*, part 1, p. 5-20.