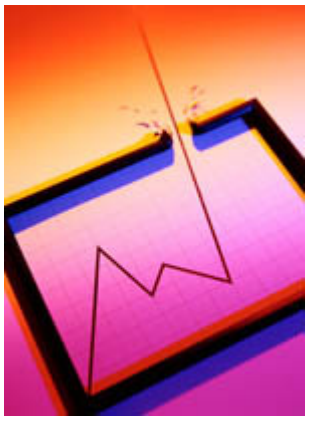
Sondages de printemps ou printemps des sondages ?

Par Dominique Joye



La période électorale que vit actuellement la France est bien sûr un moment de débats. Ceux-ci ne se limitent pas à la seule campagne *stricto sensu*, mais englobent le rôle politique de plusieurs acteurs et institutions. La question des sondages ressurgit ainsi non seulement parce qu'elle renvoie à l'opinion publique[1] et à l'image du votant libre de son choix et, donc *in fine*, à la démocratie ; non seulement parce qu'elle représente un marché économique non nul[2] mais aussi parce qu'elle met à jour des débats touchant aussi bien à la définition des sciences sociales qu'à la manière de les mettre en oeuvre. Ces thématiques sont explicitement abordées dans un article récent, dans lequel Roland Cayrol et Stéphane Rozès prennent la défense des sondages et des sondeurs[3], arguant en particulier de la démarche scientifique propre aux sciences sociales. Sans reprendre ici un vieux débat sur la nature de l'opinion[4] ou les effets des sondages politiques, quelques éléments de réflexion peuvent être avancés sur les traces lancées par Cayrol et Rozès, à commencer par le caractère scientifique invoqué par ces auteurs et par la méthodologie qui y correspondrait.

Les sondages reposent certes sur une base scientifique, le calcul des probabilités et la statistique inférentielle, mais leur usage dans ce contexte

suppose idéalement une série de conditions. La première est un échantillon réellement probabiliste, tiré à partir d'une liste exhaustive, ce qui pose une série de problèmes.

1) Jusqu'où les bases de sondage utilisées couvrent-elles l'ensemble de la population visée ? S'il s'agit du « bottin de téléphone », cela n'est probablement pas le cas : tout le monde n'a pas le téléphone, en tous cas pas nécessairement de lignes fixes, et tout le monde ne figure pas dans ce fameux « bottin » : listes rouges et portables non inscrits sont monnaie courante dans la plupart des pays et même de plus en plus fréquents.

2) L'idée d'appeler des numéros de téléphone au hasard[5] ne résout qu'une partie des problèmes : bon nombre sont des numéros professionnels qu'il est alors bien difficile de distinguer de numéros privés. Et même si l'on ne disposait que de numéros privés, en mélangeant numéros de lignes fixes, correspondant à un ménage, et numéros de portable, correspondant à un individu, on se retrouve un peu emprunté ![6]

Cela montre déjà qu'avoir, et mettre à disposition de la recherche, une liste des habitants est la première condition pour des sondages de qualité. Cela n'est de loin pas réalisé dans les différents pays européens, sous prétexte d'une défense du secret statistique qu'il s'agirait de contrebalancer par un droit à une information de qualité.[7] Ceci dit, les difficultés ne s'arrêtent pas ici pour autant.

La bonne pratique suppose comme on l'a dit un tirage aléatoire plutôt qu'un modèle basé sur des quotas tels que le plus souvent pratiqué en matière de sondages politiques. Là encore, il y a bien des raisons à cela.

1) La logique des quotas suppose qu'il suffit d'avoir une « maquette » de la population respectant quelques caractéristiques pour que la « représentation » soit bonne. Cela ne fonctionne qu'à la condition nécessaire, mais pas suffisante, que les critères pris en compte soient « les bons ». Or dans quelle mesure l'âge, le sexe et la Csp suffisent-ils à expliquer ou simplement à prédire le vote ?

2) Les quotas reposent sur l'illusion que tous les répondants sont « interchangeables » pour autant qu'ils remplissent les critères explicites. Cette hypothèse n'est certainement pas vraie : il est vraisemblable qu'un répondant qui

« aime répondre aux sondages », et qui le fait, ne soit pas identique à quelqu'un qui n'aime pas répondre aux sondage… et qui ne le fait pas, même s'ils appartiennent tous les deux au même « quota ».

La solution est bien sûr d'insister pour que toute personne contactée réponde. Encore une fois, les quotas n'incitent en aucune manière à travailler dans ce sens. Certaines données laissent supposer l'ampleur du problème. Dans les années soixante-dix, les sondages étaient rares et c'était un privilège d'y répondre : le taux de réponse aux enquêtes sociales ou politiques pouvaient avoisiner les quatre-vingts pour-cent. Aujourd'hui, dans bon nombre de pays d'Europe, il est descendu aux alentours de cinquante pour-cent pour des enquêtes longues, mais de qualité. A combien est-il en l'occurrence dans les sondages politiques tels qu'invoqués par Cayrol et Rozès ? Ce genre de chiffre n'est évidemment que rarement publié dans les sondages de faible qualité[8] mais Alexandre Duyck, dans le *Journal du dimanche* du 8 avril 2007, mentionne plus de 10 000 appels pour moins de 1000 répondants, soit un taux de réponse qui serait de l'ordre de dix pour-cent. Ce faible taux de réponse ne serait pas un problème si la non-réponse se répartissait de manière aléatoire. Ce n'est pas le cas ! Deux effets au moins doivent être distingués.[9]

1) Les répondants qui voudraient bien répondre mais ne sont pas chez eux quand le téléphone sonne. En moyenne, ces gens sont plus intéressés par la politique que les autres mais simplement, aussi du fait d'une vie active, plus difficile à atteindre. Un sondage trop vite fait va les ignorer et sera donc biaisé.

2) Les répondants qui tendent à refuser sont d'un tout autre profil, le plus souvent peu enthousiastes à communiquer leurs intérêts ou leurs préférences, souvent peu intéressés par la politique. Un sondage trop vite fait va aussi ignorer ces potentiels répondants et sera donc aussi biaisé !

La réponse proposée par certains sondeurs est alors simple : il suffit de redresser l'échantillon en pondérant par l'un ou l'autre critère. Cela souffre cependant de plusieurs faiblesses.

1) Il y a des difficultés statistiques dans l'usage des pondérations visant ce type de correction, non seulement pour l'estimation d'une variable en particulier mais

aussi pour celle de la relation entre plusieurs variables.[10]

2) La pondération sur des critères démographiques est insuffisante et peut même amener à augmenter les erreurs plutôt qu'à les diminuer comme, par exemple, les travaux de Voogt l'ont montré.[11]

3) Utiliser les informations du vote passé des répondants est probablement insuffisant car on ne sait pas si ces répondants ont un comportement en la matière identique à ceux n'ayant pas répondu à l'enquête, ni comment tenir compte de la participation : dans de très nombreux cas, les répondants aux enquêtes sont ceux qui sont les plus intéressés par la politique… la preuve, ils répondent à des questions politiques !

4) De nombreuses études montrent que l'information à utiliser est celle fournie par les non-répondants, par définition difficile à obtenir ! Suivre une telle voie implique bien sûr la mise en place de techniques spécifiques et d'études dédiées.

Faut-il dès lors jeter les sondages électoraux à la poubelle ? Certainement pas car leurs prévisions sont correctes dans l'hypothèse où les mécanismes à la base du vote ne changent pas et que l'on puisse donc calculer la différence entre les résultats des répondants et le résultat final. Ceci dit, on peut se demander si la multiplication de sondages de qualité douteuse constitue la stratégie la plus pertinente, et s'il ne serait pas plus intéressant de mettre en place des stratégies méthodologiquement plus correctes, même au risque d'être plus coûteuses. Cela permettrait peut être de construire une démarche qui permette d'avancer dans la compréhension des rapports entre citoyens et politiques, bref une démarche scientifique.

Ce dernier point nous permet de revenir à Cayrol et Rozès, qui écrivent, dans l'article mentionné ci-dessus : « Que signifierait donc la publication de données brutes, que nous savons fallacieuses ?… Non : cette demande vise, croyons-nous, à décrédibiliser les sondeurs, en laissant entendre qu'ils "manipulent" leurs données. Or, ce travail de nature scientifique sur les données, nous le revendiquons pleinement. C'est notre métier. Les sciences sociales ne sont pas des sciences exactes ; elles sont fondées sur une démarche scientifique ».

Or qu'est-ce qui constitue jusqu'à nouvel ordre le propre de la démarche scientifique ? Ce n'est pas l'invocation d'une technologie, quelle qu'en soit le succès, mais bien l'ouverture au regard des « pairs », la transparence de la méthode, la capacité à pouvoir reproduire les résultats en suivant la même démarche. Ce n'est pas pour rien que les bonnes revues scientifiques, même de sciences sociales, demandent la mise à disposition des données à l'ensemble de la communauté scientifique pour qu'un article puisse être publié. Ce n'est pas pour rien non plus qu'il existe des archives de données qui permettent le caractère cumulatif de la recherche. Mais cela implique d'accepter de publier ses astuces de fabrication et non de les cacher. Bref il y a encore du travail à faire dans le monde des sondages !

Résumé

La période électorale que vit actuellement la France est bien sûr un moment de débats. Ceux-ci ne se limitent pas à la seule campagne stricto sensu, mais englobent le rôle politique de plusieurs acteurs et institutions. La question des sondages ressurgit ainsi non seulement parce qu'elle renvoie à l'opinion publique1 et à l'image du votant ...

Notes

1 Un sondage Csa-Cisco publié le 9 avril 2007 par *Le Parisien* mentionne en effet que à la question « *Selon vous, qui sera élu président de la République ?* », près de six sondés sur dix (59%) répondent Nicolas Sarkozy (Ump), contre 18% Ségolène Royal (Ps) ce qui est bien sûr sans rapport avec les intentions de vote et tendrait à laisser penser que le sondage en lui-même devient l'évènement, au delà de l'élection elle-même.

2 Le *Journal du dimanche* du 8 avril mentionne plus de deux cents sondages déjà réalisés pour les élections de cette année.

3 Roland Cayrol et Stéphane Rozès , « Plaidoyer pour les sondages », *Le Monde*, 20 mars 2007 .

4 La question est en effet déjà posée dans les travaux de Converse, dans les années soixante.

5 Le célèbre « *Random Digit Dialing* » dans la littérature spécialisée.
6 D'un point de vue formel, la probabilité d'atteindre quelqu'un n'est pas la même si un numéro de téléphone correspond à cinq personnes ou à une seule… Ceci en supposant que la probabilité d'être en mesure de répondre au téléphone soit la même dans un cas ou dans l'autre.
7 Ce n'est pas le lieu non plus de discuter ce point. Notons simplement que les instituts de sondage ont un besoin vital de protéger l'identité de leurs répondants, faute de quoi leur survie serait immédiatement condamnée. L'argument invoqué aboutit plutôt à assurer un monopole de fait sur la connaissance pour les agences officielles de statistique, ce qui pose une question de transparence et d'accès à la connaissance, bref de démocratie.
8 Et encore la manière de le rapporter est soumis à controverse pour donner une image pas trop défavorable, à tel point que l'association américaine pour la recherche en opinion publique a dû mettre au point des normes détaillées et sévères.
9 Encore une fois les enquêtes de qualité enregistrent ce type d'information pour que l'analyse puisse en tenir compte. En pratique chaque tentative de contact fait l'objet d'une documentation minimale.
10 Cf. par exemple John Brehm, *The Phantom Respondents*, The University of Michigan Press, Ann Abor, 1993.
11 J. J Voogt Robert. et Willem E. Saris, « To Participate or not to Participate: The Link Between Survey Participation, Electoral Participation and Political Interest », in *Political Analysis*, 11, 2003, pp. 164-179.