

L'impact de 2002 sur les instituts de sondage

par Claire Durand,

Professeur titulaire, département de sociologie, Université de Montréal, Québec, Canada

L'impact possible de l'élection présidentielle de 2002 sur le vote des électeurs en 2007 a suscité de nombreux commentaires, entre autres dans les médias. Comme les sondages avaient été considérés en partie responsables de la "catastrophe" de 2002, il convient de se demander si celle-ci n'a pas eu également un impact sur les sondages, en fait sur les instituts de sondage et sur les méthodes qu'ils utilisent. Cet article compare les sondages de la campagne électorale de 2007 à ceux de 2002 sur plusieurs indicateurs de façon à répondre à cette question.

Que fut reproché aux instituts de sondage en 2002? De n'avoir pas prévu la possibilité de la qualification de Jean-Marie Le Pen au second tour ou du moins de ne pas en avoir informé les électeurs et d'avoir laissé croire qu'il n'y avait pas d'autre possibilité qu'un deuxième tour entre Lionel Jospin et Jacques Chirac. Comment en était-on arrivé là? Les estimations publiées étaient normalement sélectionnées parmi un nombre parfois important de redressements fondés sur la reconstitution du vote aux élections précédentes, la certitude du choix et la certitude d'aller voter. Mais encore, on s'autorisait parfois de modifier ensuite les chiffres jugés "improbables" en fonction de "l'expérience professionnelle du sondeur". Cette pratique consistait à lisser certaines évolutions jugées trop rapides ou invraisemblables. Cette situation a amené une unanimité anormale des estimations des intentions de vote présentées aux électeurs : cinq des six derniers sondages publiés donnaient une même intention de vote de 18% pour Lionel Jospin. Les erreurs dans les sondages ont toutefois cet immense avantage d'amener normalement les sondeurs à se questionner et à raffiner leurs méthodes. Cela s'est-il produit?

Les sondages ont-ils été meilleurs en 2007 qu'en 2002?

Pour répondre à cette question, on peut comparer les derniers sondages publiés avant l'élection par les six instituts (Ifop, IPSOS, CSA, BVA, LH2, Sofres) en 2002 et en 2007. Pour le premier tour, si on prend la moyenne de ces sondages, on réalise que les sondages ont fait un moins bon score en 2007 qu'en 2002. Alors que seul Le Pen était mal estimé en 2002 (- 4 points de pourcentage), en 2007, Le Pen est surestimé de 3,8 points (à 14,2) et Sarkozy est sous estimé de 3,3 points (à 27,9). Si l'on prend un autre indicateur, soit la meilleure estimation des derniers sondages réalisés, celle-ci s'écartait de 2,9 points pour Le Pen en 2002. En 2007, l'écart est de 2,6 points pour Le Pen toujours, mais en sens inverse. Pour ce qui est de la pire estimation, elle sous-estimait Le Pen de 6,9 points en 2002; en 2007, Le Pen est surestimé de 6,1 points alors que Sarkozy est sous estimé de 4,7 points, Royal de 3,4 points et Bayrou de 2,6 points.

Bref, le principal succès des instituts en 2007 est d'avoir prévu le bon ordre d'arrivée – du moins pour cinq des six instituts – mais cela est en partie dû au fait que les écarts entre les candidats de tête était plus important qu'en 2002. La mauvaise estimation du vote Le Pen est toujours présente mais en sens inverse, comme si les instituts, traumatisés par 2002, avaient toujours choisi leur estimation la plus élevée.

Pour ce qui est du deuxième tour, notons que les intentions de vote pour le duo Sarkozy-Royal sont demeurées au même niveau à partir du début février. Les analyses statistiques de tous les sondages publiés prédisaient Sarkozy à 53,4%. Il a obtenu 53,1%. Toutefois, parmi les six derniers sondages, on note deux estimations à 55% et une à 54,5% pour Sarkozy. Ces sondages, dont deux ont été publiés sur Internet dans les minutes précédant l'interdiction de publication, laissaient penser qu'il y avait une remontée de Sarkozy suite au débat. Ceci ne s'est pas avéré et ces sondages ont été les pires de l'ensemble des sondages publiés entre les deux tours.

Les instituts ont-ils changé leurs méthodes?

Certains indices permettent d'examiner de façon indirecte si des changements dans les méthodes se sont produits. Une analyse des derniers sondages publiés par chaque firme montre une plus grande variété des intentions de vote publiées en 2007 : pour Sarkozy et Le Pen, il y a eu trois estimations similaires sur six (à 28% pour Sarkozy et à 14% pour Le Pen), pour Royal, deux (24%) et pour Bayrou, aucune. En comparaison, en 2002, il y avait cinq estimations semblables à 18 % pour Jospin, trois à 20% et deux à 19,5% pour Chirac et trois à 14% pour Le Pen. Il y a donc plus de variation en 2007. Un autre indice est fourni par les différences entre les instituts. En 2002, il n'y avait aucune différence significative entre les instituts quant aux estimations publiées pour les trois candidats de tête entre le 1^{er} mars et l'élection. En 2007, de telles différences existaient pour trois des quatre candidats en tête, seul Bayrou étant estimé de façon similaire par les divers instituts. Ceci laisse penser que les instituts utilisent des méthodes différentes – puisqu'ils obtiennent des résultats différents – et qu'ils recourent moins à des redressements *ad hoc* qu'en 2002.

La consultation des notices produites par les instituts à la Commission des sondages de même que des informations disponibles sur les sites Internet des instituts permettent de conclure qu'aucun institut ne recourt plus, du moins formellement, à des redressements "basés sur l'expérience" comme en 2002. Toutefois, les méthodes décrites sont à peu près identiques aux descriptions de 2002, la principale différence entre les instituts se situant dans l'identité et le nombre d'élections précédentes utilisés dans les redressements. Comme en 2002, il est généralement impossible de savoir quel redressement a été utilisé pour un sondage donné et les chiffres publiés peuvent provenir de redressements différents pour divers candidats.

A la décharge des instituts, notons que la tâche semble ardue. Même dans les sondages effectués le soir du premier tour, le 22 avril, le vote pour Le Pen est généralement mal reconstitué, de 4 à 7 pourcent des électeurs ayant déclaré avoir voté pour Le Pen selon les informations disponibles. Ceci amène à penser que le problème de l'estimation des intentions de vote ne résulte pas d'un problème de mémoire du vote antérieur chez les répondants mais bien de difficultés d'échantillonnage, de non réponse et de dissimulation des préférences.

Si certains indices laissent croire que des changements sont survenus dans les méthodes des instituts de sondage français suite à la "catastrophe" de 2002, rien ne permet toutefois de conclure que les méthodes utilisées ont été modifiées de façon significative et sont plus fiables. La moins grande unanimité des estimations présentées peut toutefois inciter les électeurs et les media à relativiser les prédictions. En utilisant des méthodes similaires à 2002, les instituts ont produit cette fois-ci une surestimation de Le Pen, mais non pas une meilleure estimation. Si les

Français avaient été aussi divisés qu'en 2002, il n'est pas certain qu'une autre catastrophe aurait été évitée.

A l'aube des élections législatives, ce sont les électeurs qui risquent encore de sauver ou de perdre les instituts. En effet, si les écarts sont suffisamment importants entre les partis, la prédiction de l'ordre d'arrivée s'en trouve facilitée et l'attention médiatique est détournée de la justesse des estimations chiffrées. Par contre, lorsque les écarts sont faibles, la situation devient très difficile et toute mauvaise estimation peut avoir des conséquences désastreuses, entre autres sur la crédibilité des instituts. Il serait sans doute approprié que les instituts et les chercheurs collaborent ensemble pour trouver des méthodes plus fiables d'estimer les intentions de vote pour l'avenir.