

"La statistique éclaire-t-elle
les questions de société" ?

Soirée du 9 mai 2017

Sondages électoraux : peut-on leur faire confiance ?

Synthèse des débats ^[*]

Les sondages électoraux sont de plus en plus décriés, en particulier les enquêtes d'intention de vote. Élection présidentielle américaine, Brexit au Royaume-Uni, primaires en France, pour ne citer que les plus souvent évoqués, sont des exemples récents des difficultés parfois rencontrées pour mesurer des intentions de vote.

Si le contexte politique incertain et la défiance des citoyens constituent une explication de cette nouvelle donne, on peut s'interroger sur l'effet du recours quasi-généralisé aux enquêtes en ligne ces dernières années au détriment du téléphone. Ce nouveau mode de recueil, peu onéreux, présente a priori l'avantage de diminuer les effets de la désirabilité sociale et ainsi de permettre de mieux mesurer des phénomènes socialement peu reconnus, y compris dans le cadre de sondages électoraux.

Il n'est pas pour autant exempt de biais, qu'il importe également de considérer. Quel est l'impact de cette évolution méthodologique sur la qualité des sondages électoraux ? Qui interroge-t-on ? Le redressement des échantillons, souvent présenté comme une boîte noire, apparaît comme un facteur supplémentaire d'incertitude sur les résultats. Qu'en est-il réellement ? Et quel peut être l'effet de la multiplication des sondages, de nature et de valeur scientifique très variables, sur la confiance que leur accorde le public ?

Invités :

Flora Chanvriil

Chargée d'études au Cevipof¹ - Sciences Po

Bruno Jeanbart

Directeur général adjoint d'OpinionWay

^[*] Tant l'exposé liminaire que le contenu des échanges sont structurés en quelques thèmes, sans suivre l'ordre chronologique. Par ailleurs, l'identité des intervenants n'était pas toujours connue et l'on a choisi de ne pas attribuer nominativement les propos. Au reste, ceux-ci ont été reconstitués à partir des notes du secrétariat sans reprendre leur formulation détaillée. Pour retracer le débat, les thèmes sont souvent introduits sous forme d'une question : ce qui vient ensuite n'est pas la seule réponse de l'invitée, mais l'ensemble des contributions des participants.

¹ Centre de recherches politiques de Sciences Po, anciennement dénommé Centre d'étude de la vie politique française. Créé en 1960, il est associé au CNRS depuis 1968.

Exposés introductifs :

Flora Chanvril indique que le Cevipof est spécialisé dans l'étude de la pensée politique et des idées politiques, ainsi que dans celle des attitudes, comportements et forces politiques. Dans la continuité de ses précédentes enquêtes électorales nationales, le Cevipof a mis en œuvre depuis 2015 un dispositif de grande taille et de longue durée dans la perspective des élections régionales, présidentielle et législatives. Le dispositif consiste en un panel en ligne opérationnel entre novembre 2015 et juin 2017 auprès de 25 000 répondants panélisés. Le calendrier des vagues d'interrogation comporte 16 vagues de novembre 2015 à juin 2017.

La Commission des sondages a contribué à la définition d'un sondage telle qu'elle est donnée par la loi du 25 avril 2016² : « *Un sondage est, quelle que soit sa dénomination, une enquête statistique visant à donner une indication quantitative, à une date déterminée, des opinions, souhaits, attitudes ou comportements d'une population par l'interrogation d'un échantillon.* »

La loi évoque aussi la notion de « marge d'erreur » des résultats publiés. Ces résultats doivent être donnés avec plus ou moins quelques pour cent. Cette marge d'erreur dépend à la fois de la taille de l'échantillon et du pourcentage mesuré. Par exemple, pour un échantillon de mille personnes et pour un résultat particulier correspondant à un pourcentage de 50 %, la marge d'erreur est de plus ou moins 3,2 points de %. Il faut comprendre que pour 95 % des échantillons qu'on aurait pu tirer, ce résultat particulier aurait été situé entre 46,8 % et 53,2 %.

Les sondages se multiplient. Leur nombre était de quatorze en 1965 ; on en a décompté environ 500 à l'occasion de la présidentielle de 2017... Cette profusion de chiffres, accompagnée d'une diffusion de plus en plus massive, participe aux questionnements actuels.

Il est important de revenir sur plusieurs points méthodologiques. On distingue les **enquêtes longitudinales** (notamment pour la mesure des intentions de vote) et les **enquêtes en un seul point du temps** (enquêtes « one shot », notamment pour la reconstitution des votes exprimés).

Dans les enquêtes longitudinales, on distingue traditionnellement les *panels* et les *baromètres* de la *technique dite « Rolling Cross Section »*. Chaque formule a ses avantages et ses inconvénients.

Dans les panels, on suit les mêmes individus vague d'enquête après vague d'enquête, ce qui permet d'estimer avec précision les changements d'intention de vote et les raisons de ces évolutions. Le Cevipof suit depuis novembre 2015 le même panel et en est à la 14^e vague d'interrogation à la date de ce Café. Ce panel comportait 24 369 personnes au départ, et notamment des primo-votants. La difficulté principale pour le panel est l'attrition, c'est-à-dire la perte progressive de ses membres pour toutes sortes de causes, d'où l'importance du cadrage initial de l'enquête en termes de durée du panel, de nombre de questionnaires administrés et de taille d'échantillon.

Dans les baromètres, on interroge un nouvel échantillon représentatif à chaque vague, avec pour inconvénient qu'on ne suit pas l'évolution des intentions de vote des mêmes personnes.

Avec le « Rolling Cross Section (RCS) », on cumule en quelque sorte les avantages du panel et du baromètre en renouvelant régulièrement dans le temps une partie de l'échantillon afin qu'il reste représentatif à tout instant et stable en effectif. L'inconvénient est cette fois que le coût associé à cette technique oblige à se limiter à un échantillon de taille généralement assez modérée. Il est à noter que dans une enquête plus classique, les répondants de fin de terrain ne sont pas les mêmes - sociologiquement et politiquement parlant - qu'en début de terrain, d'où l'intérêt du renouvellement régulier de l'échantillon dans un RCS.

² Loi n° 2016-508 du 25 avril 2016 modifiant la loi n° 77-808 du 19 juillet 1977 relative à la publication et à la diffusion de certains sondages d'opinion.

L'enquête « one shot » correspond par exemple à une classique enquête post électorale, ou à un sondage dit « à la sortie des urnes »³ qui part d'un échantillonnage des bureaux de vote afin de produire les estimations publiées à 20 h le jour du vote.

De plus en plus de sondages, alors dits « **en ligne** », se font sur Internet. Cette pratique pose le problème de la fracture numérique. Dans son « Baromètre du numérique » de 2015, le Crédoc⁴ notait que 83 % des personnes interrogées dans leur enquête disposaient d'un accès Internet sur support fixe et 3 % supplémentaires uniquement sur support mobile (téléphone, tablette, etc.), ce qui signifiait que 14 % de la population n'avaient pas d'accès Internet. Cette proportion se réduit au fil du temps mais couvre une population qui semble n'être pas homogène au reste de la population. Il y a des différences politiques et sociologiques entre ceux qui n'ont pas accès à Internet (14 %) et les autres (86 %). On sait par exemple que 20 % des habitants des communes de moins de 2 000 habitants n'ont pas accès à Internet, ou encore que la moitié des personnes de plus de 70 ans n'utilisent pas Internet. Tout cela conduit à des biais de couverture du territoire et de la population. À quoi s'ajoute le problème des « **access panel** »⁵, qui sont constitués de volontaires, donc de personnes plus intéressées que la moyenne pour répondre aux enquêtes. Si l'on reprend l'exemple des personnes de plus de 70 ans, celles qui font partie des « access panel » font à la fois partie des 49 % qui ont accès à Internet puis des volontaires parmi ces 49 % : cela ajoute un biais d'auto-sélection à celui résultant de leur faible utilisation d'Internet.

Les avantages des enquêtes en ligne sont néanmoins nombreux : les coûts sont peu élevés, la désirabilité sociale⁶ se manifeste beaucoup moins sur Internet, les contrôles de la qualité des données sont plus aisés et se sont améliorés ces dernières années, et on peut faire des expérimentations d'outils visuels qui seraient impossibles avec le téléphone, comme par exemple faire réagir les personnes interrogées sur des photos de personnalités politiques⁷.

Les sondages produisent probablement des effets à la fois sur les électeurs, sur le grand public et sur les élites. En 2014, Gérard Grunberg et Nonna Mayer notaient que les effets sont plus importants pour les élites (et les candidats) que pour les électeurs⁸.

Dans les enquêtes sur les intentions de vote, comment mesurer l'incertitude qu'a l'électeur d'aller voter ? Autrefois, la question était simplement dichotomique : on lui demandait s'il allait oui ou non voter. Puis les questionnaires ont évolué et on demande maintenant à l'électeur de coter sur une

³ Cette appellation, relèvera un participant, n'est guère appropriée : ce sont les bulletins qui sortent des urnes et non les électeurs ! Il vaudrait mieux parler d'enquête à la sortie des bureaux de vote.

⁴ Centre de recherche pour l'étude et l'observation des conditions de vie. C'est un organisme d'études et de recherche au service des acteurs de la vie économique et sociale. Pour le Baromètre du numérique, voir par exemple <https://www.data.gouv.fr/fr/datasets/barometre-du-numerique-2007-2016/>

⁵ NDR : Ce sont des viviers de ménages et de personnes qui sont recrutés par les sociétés d'études et qui acceptent de participer régulièrement à des études de marché ou des enquêtes.

⁶ Désirabilité sociale : les personnes interrogées peuvent estimer que certains de leurs comportements ou opinions ne sont guère appréciés socialement. Elles ne les avouent pas dans un entretien en tête-à-tête mais les confieront plus volontiers dans la réponse à un questionnaire en ligne. Ainsi, le vote pour le Front national a longtemps été sous-déclaré. L'intention de s'abstenir à une élection, l'addiction aux drogues ou certaines pratiques sexuelles demeurent autant d'objets de sous-déclaration aux enquêtes. Voir : Flora Chanvriil, Viviane Le Hay (2011), « Le baromètre politique Cevipof sur la confiance : un dispositif d'enquête mixte innovant » dans Tremblay M.-E., Lavallée P., El haj Tirari M. (dir.), *Pratiques et méthodes de sondage*, Dunod.

⁷ Pour les lecteurs qui souhaiteraient aller plus loin sur le bilan avantages/inconvénients, voir :

https://www.cairn.info/load_pdf.php?ID_ARTICLE=POPU_1002_0309

Didier Fripiat, Nicolas Marquis (2010), « Les enquêtes par Internet en sciences sociales : un état des lieux », *Population*, Vol. 65, n°2.

⁸ Gérard Grunberg, Nonna Mayer (2014), « L'effet sondage. Des citoyens ordinaires aux élites politiques » dans Yves Déloye et al., *Institutions, élections, opinion*, Presses de Sciences Po, pages 219 à 236.

échelle de zéro à 10 son degré de certitude d'aller voter. Cette évolution permet de mieux mesurer le degré d'incertitude/certitude et de préciser les intentions de vote en fonction de ce degré de certitude. Ainsi, pour l'Enquête électorale française et la plupart des instituts, les intentions de vote publiées sont filtrées sur les personnes « tout à fait certaines d'aller voter », c'est-à-dire celles ayant choisi la position 10 sur l'échelle d'incertitude/certitude

Les défauts de couverture, la marge d'erreur, le degré d'incertitude d'aller voter, la désirabilité sociale, constituent autant de biais auxquels il faut remédier, avant la publication des résultats de l'enquête, par la technique des redressements. Ces redressements permettent de pallier les déséquilibres sociodémographiques et politiques observés dans les données brutes. Les redressements sociodémographiques correspondent généralement aux critères utilisés pour les quotas (l'âge, le genre ou la profession), déterminés en fonction des résultats du recensement de la population. Le niveau de diplôme est plus difficile à redresser et donc plus rarement utilisé dans les critères de redressement. Les critères politiques utilisés dépendent du scrutin considéré : en général, on se réfère, lors du traitement des réponses, à la fois au scrutin antérieur de même nature, aussi proche que possible dans le passé (par exemple la reconstitution de vote à la présidentielle de 2012 pour le premier tour de l'élection présidentielle de 2017) et au scrutin le plus récent (dans notre exemple les régionales de 2015).

Pour l'élection présidentielle de 2017, les sondages se sont révélés bon pour le premier tour, c'est-à-dire que l'état de l'opinion dont ils ont rendu compte s'est révélé proche des votes effectifs. Les choses ont été plus compliquées pour le second tour, en particulier parce que l'effet du débat entre les deux candidats - débat intervenu le mercredi soir précédant le scrutin - n'a pas pu être pris complètement en compte dans les derniers sondages publiés en raison des dates de terrain. On peut expliquer de cette manière une partie de l'écart entre les meilleures estimations des intentions de vote en faveur du gagnant (63 %) et ce qui s'est passé réellement (66 %). À quoi s'est ajoutée la difficulté d'estimer correctement l'ampleur des abstentions et des votes blancs ou nuls.

Bruno Jeanbart prolonge cette analyse en évoquant plusieurs éléments propres à l'élection présidentielle de 2017. Le sondeur pouvait *a priori* être effrayé mais finalement les choses se sont bien passées. Le deuxième tour était théoriquement plus simple que le premier puisqu'on n'avait que deux candidats en lice au lieu de onze et qu'on pouvait entre les deux tours travailler avec le premier tour alors qu'avant il fallait se référer au premier tour de l'élection présidentielle de 2012. Et cette technique n'est pas sans défauts. Quand on demande à l'électeur comment il a voté en 2012, il peut faire des erreurs de mémoire et c'est un biais majeur. Par ailleurs, il y a eu beaucoup de changements politiques entre 2012 et 2017, et l'arrivée des primo-votants complique aussi la donne puisque leur vote en 2012 est par définition inexistant.

Au deuxième tour, on utilise plutôt ce qui s'est passé au premier tour. La mémoire des personnes interrogées est encore fraîche et on peut être plus précis dans les pondérations. Pourtant, les estimations pour le deuxième tour ont été moins bonnes que celles du premier. Certes, des événements considérables sont cette fois survenus entre les deux tours et il y a en effet le problème des délais puisque le débat, tardif, entraînait que des intentions de vote publiées le vendredi correspondaient au mieux à une enquête le jeudi. Or, le jeudi, une enquête ne pouvait prendre la pleine mesure des effets du débat du mercredi soir puisque ces effets se sont manifestés sur plusieurs jours. Par ailleurs, une sociologie différente des électeurs des deux candidats a entraîné des comportements de réponse obligeant dans certaines zones à des redressements importants des réponses brutes⁹.

Au total, bien que les estimations pour le premier tour aient été bonnes, il reste à propos de cette élection pas mal de questions inhabituelles à résoudre, à la fois sur les techniques de sondage et sur l'expression des intentions de vote. Le contexte était particulier lui aussi. D'une part, l'élection présidentielle aux États-Unis et le Brexit avaient suscité des doutes sérieux sur la validité des

⁹ En données brutes, on a pu avoir des scores 70-30 en faveur de Macron.

sondages. D'autre part, en liaison avec ces doutes, beaucoup d'investisseurs internationaux, hantés qu'ils étaient par le risque de ne pas anticiper suffisamment tôt un éventuel succès de Mme Le Pen, se sont manifestés très tôt auprès des instituts de sondage.

Plus généralement, on peut relever les difficultés propres aux sondages électoraux : en premier lieu vient la difficulté de mesurer des intentions de vote dans une société dans laquelle de plus en plus de gens sont difficiles à joindre. Opinion Way a été en 2006 le premier institut à faire des enquêtes en ligne et, trois ans plus tard, donc aux élections européennes de 2009, à utiliser l'Internet pour ses enquêtes sur les intentions de vote. En 2012, sur huit instituts de sondage, cinq opéraient par téléphone et trois sur l'Internet. En 2017, les huit instituts ont fait leurs enquêtes en ligne. Cela résulte de la difficulté à joindre les gens par d'autres méthodes. On note par exemple que, dans les quartiers très bourgeois ou très populaires, l'accès physique ou téléphonique aux personnes est de plus en plus difficile.

Plus de 90 % de la population disposent d'un téléphone mobile mais les personnes de plus de 70 ans sont moins équipées. De plus, tous les numéros des téléphones mobiles ne sont pas référencés et, de toute manière, le téléphone mobile n'est pas adapté aux questionnaires longs.

Par ailleurs, peut-on penser que celles des personnes âgées qui utilisent Internet (environ la moitié de cette population) ont le même profil sociologique et électoral que l'autre moitié ? Probablement : non !

Tous ces facteurs font que, sur un échantillon idéal défini avant l'enquête, on n'obtient qu'un taux de réponse faible. Aux États-Unis, ce phénomène est scientifiquement observé depuis 1992. Cette année-là, 40 % des personnes jointes par les enquêteurs répondaient effectivement. La proportion était tombée à 15 % en 2012... Sur ce point, les améliorations à venir sont douteuses. Le temps est compté pour tous et les gens répugnent de plus en plus à répondre à une enquête un peu longue. Et puis le système politique est beaucoup plus mouvant aujourd'hui qu'il y a une quarantaine d'années. En 2012, année pourtant peu éloignée, la moitié des gens savaient pour qui ils allaient voter assez longtemps avant la date du scrutin. Cette proportion est tombée à 28 % en 2017. Le baromètre sur la confiance envers les politiques, qui montre un niveau de confiance de 9 % seulement, peut expliquer la mobilité des intentions de vote. Certes, les principaux mouvements se font à l'intérieur du bloc de gauche ou de celui de droite mais il est très difficile d'estimer le comportement des gens à l'égard d'un candidat qui se veut les deux à la fois. Les choses risquent de se dégrader davantage encore du fait de l'instabilité actuelle des systèmes politiques. Ces phénomènes sont moins accentués en Allemagne mais sont très forts aussi en Italie ou en Espagne.

En France, la technique consistant à mesurer et pondérer en se servant des votes passés est un système assez conservateur qui se trouve fragilisé par les mobilités actuelles en matière d'intention de vote. C'est une difficulté supplémentaire qui n'est pas réglée.

Globalement, observe l'invité, c'est la troisième fois que les sondages d'intentions de vote produisent de bonnes estimations en matière d'élection présidentielle et, en 2017, contrairement à ce qui a été observé dans le passé, le propos est valable pour tous les instituts de sondage. Il n'empêche que des règles de conduite s'imposent. Du point de vue d'Opinion Way :

- il faut habituer la population à l'idée qu'un sondage est une mesure de ce qui va se passer, mais une mesure incertaine à cause des biais d'échantillonnage et des marges d'erreur, ces dernières étant élargies par la difficulté croissante à respecter un échantillon idéal. Il faut donc regarder les estimations en tendance, alerter sur les incertitudes et réfléchir en permanence aux moyens de faire comprendre comment les instituts de sondage travaillent. En effet, les indicateurs autres que l'intention de vote, qui découlent des questions posées lors de l'enquête et seraient susceptibles d'éclairer les estimations proprement dites, sont ignorés par les médias et par conséquent par le public ;

- il faut travailler sur les méthodes. Des essais de modélisation ont été réalisés en s'appuyant sur la connaissance fine de la population que permettent les recensements. On s'est ainsi interrogé avec ces modèles sur la question de savoir si un Emmanuel Macron sorti de nulle part pouvait atteindre certains scores, ou si la baisse continue des intentions de vote en faveur de Mme Le Pen était

réaliste, ou encore s'il était possible d'avoir un candidat du Front national à 25 % des intentions de vote. Ce genre de questionnement a évité pas mal de chausse-trappes¹⁰ tout en soulignant la fragilité croissante des sondages.

Quelle est l'influence des sondages ? Elle existe mais Opinion Way pense que les gens sont plongés dans la confusion par l'abondance même des sondages. Le poids des sondages est indiscutable sur le comportement des candidats et sur celui des médias de toute espèce. L'invité rêve d'une expérience de pensée : couper le pays en deux avec une moitié de la population exposée à des sondages électoraux et l'autre moitié qui n'y aurait nul accès. Les comportements électoraux seraient-ils distincts dans ces deux univers ? C'est évidemment impossible à réaliser. Les sondages apportent une information parmi bien d'autres, qui jouent elles aussi leur rôle.

Débat :

Les sondages en question

Quelle est l'utilité des sondages électoraux ? demande un participant. Un autre va plus loin en dénonçant ce qu'il tient pour un jeu pathologique des médias voulant faire accroire qu'une photographie de l'opinion aurait une vertu prédictive à un instant donné alors que personne ne peut prévoir ! Selon lui, les instituts de sondage ne feraient pas preuve de suffisamment de déontologie en se prêtant à ce jeu et auraient un devoir de pédagogie et d'humilité qu'ils rempliraient mal.

Conforté en cela par un membre d'un autre institut de sondage, l'invité confirme tout d'abord ce qu'il a déjà dit, à savoir que les enquêtes sur les intentions de vote mesurent l'état de l'opinion au temps t. Les gens sont évidemment libres de changer d'avis à tout instant. Si les instituts de sondage font des enquêtes d'intentions de vote, c'est parce que tout le monde veut savoir ce qui va se passer. Il est universel de vouloir savoir ce qui va se passer. C'est donc la demande qui fait l'offre.

Grâce aux enquêtes électorales, on approche cette connaissance de l'avenir, et cela d'autant plus précisément qu'on est plus proche du scrutin. Elles permettent aussi, grâce à la palette des questions posées, de comprendre le comportement des électeurs et leur processus de décision et d'expliquer ainsi pourquoi les choses changent. Si elles ne sont pas nécessaires pour la connaissance des sociologies du vote, elles le sont pour examiner les liens entre le vote et des questions thématiques comme l'immigration, l'emploi ou la sécurité.

Et puis, note l'invité, les instituts de sondage ont des clients. C'est une réalité économique. Cela ne les empêche pas d'être rigoureux dans leur travail sur les intentions de vote. Le commanditaire impose une obligation de moyens et l'institut de sondage fait le travail, sous le contrôle de la Commission des sondages. Le commanditaire a la main sur les sujets (qu'il définit) mais pas sur les questions.

Un participant pense qu'il faut nuancer cette affirmation : certes, le commanditaire ne choisit pas les questions, mais il doit sans doute en discuter avec l'institut de sondage ? L'invitée précise qu'au Cevipof le commanditaire (qui est parfois le financeur) et les chercheurs discutent mais c'est toujours le Cevipof qui choisit les questions qui seront posées et ces questions doivent être aussi neutres que possible dans leur formulation.

¹⁰ NDR : Mon logiciel de reconnaissance vocale a écrit « chausse-trappes ». J'aurais écrit (étant âgé) « chausse-trape ». Le Petit Robert et le Petit Larousse donnent *chausse-trape* en entrée avec *chausse-trappe* comme variante orthographique. Contrairement à sa huitième édition de 1935 qui attestait *chausse-trape*, le Dictionnaire de l'Académie française donne, dans sa neuvième édition de 1992, *chausse-trappe* (pluriel : chausse-trappes), avec pour nouvelle orthographe *chaussetrappe*, qui, comme toutes les rectifications orthographiques du 6 décembre 1990, doit être soumise à l'épreuve du temps. Je n'ai donc pas corrigé mon logiciel...

N'y a-t-il pas, demande un participant, des commanditaires qui refusent de publier des résultats qui ne leur plaisent pas ? En dehors des sondages électoraux, indique l'invité, les résultats des sondages sont publiés ou non selon ce que souhaite le client. Dans le cas de résultats qui ne sont pas destinés à être publiés, si le client veut absolument poser une question idiote, l'institut de sondage s'y résoudra. En revanche il ne l'acceptera pas si le résultat du sondage doit être publié. Celui qui finance ne choisit pas les résultats. Ce ne serait d'ailleurs pas l'intérêt de l'institut de sondage, qui doit assurer sa crédibilité auprès d'autres clients potentiels.

Peu de sondages ne sont pas publiés aujourd'hui. En matière électorale, ce serait d'ailleurs trop visible ! Imaginons un institut de sondage publiant tous les jours une évolution des intentions de vote : s'il manquait une journée, ce serait une avalanche d'étonnements et de protestations.

En résumé, un institut de sondage n'a aucune raison de refuser de faire une enquête sur les intentions de vote. De nos jours, l'institut de sondage fait les mises au point nécessaires, le média prend la responsabilité de ses articles et chacun est ainsi dans son rôle. Cela dit, les instituts de sondage ne sont jamais assez entendus sur leurs mises en garde et les médias peuvent en effet surinterpréter les résultats produits par les instituts de sondage.

Quel est le modèle économique des instituts de sondage ? demande un participant. Si seules les intentions de vote intéressent, et dès lors que le produit est à peu près identique pour tous les instituts, pourquoi n'assiste-t-on pas à une concentration dans le secteur ? L'invité indique en réponse que le chiffre d'affaires des instituts de sondage - entreprises privées - résulte pour 90 % d'enquêtes marketing et pour moins de 5 % de sondages électoraux¹¹. Les instituts de sondage jouent sur leur notoriété et certains acceptent dans ce but de perdre de l'argent sur les sondages électoraux.

Qui finance ? Les médias ne paient quasiment plus les instituts de sondage. Ce sont des marques qui paient et elles explorent, à l'occasion des sondages électoraux, les domaines de leur action. Ainsi, quand des agences immobilières commanditent une enquête électorale, elles en profitent pour faire poser de nombreuses questions sur le logement. Ce que confirmera un participant qui, interrogé en ligne, s'était étonné que le questionnaire s'intéresse aussi à son logement.

En dehors des entreprises privées, les instituts de sondage sont sollicités aussi par des villes, des départements ou des régions, dans le cadre de marchés publics.

Toutes ces précisions entendues, un participant restera sur l'opinion que les instituts de sondage posent parfois des questions que les gens ne se posent pas, avec des effets sur l'opinion susceptibles d'être de grande ampleur, comme ce qui s'est passé selon lui au moment du référendum sur le traité de Maastricht (septembre 1992), quand l'accumulation des sondages a débouché sur un oui à 51 % alors que le oui l'emportait largement au début de la campagne. L'invité se montre d'accord avec Bourdieu dans sa critique des sondages qui imposent des débats mais il note que, de fait, le débat médiatique existe en général avant d'être un débat public.

Un commentaire émanant d'un autre participant relance la question des effets des sondages électoraux : si en 2002, dit cette personne, des sondages avaient été publiés juste avant le 1^{er} tour du scrutin, on aurait vu que Le Pen était en train de dépasser Jospin et le vote aurait pu en être transformé. Ce que conteste l'invité : en 2002, on pouvait déjà publier des résultats de sondages l'avant-veille de l'élection ; en réalité, personne n'a prévu l'accès de Le Pen au second tour et il y a eu là une véritable erreur technique des sondeurs.

Pour sa part, l'invitée mentionne une étude du Cevipof sur l'effet des sondages. Cet organisme a posé dans le cadre de l'Enquête électorale française la question de savoir si l'électeur tenait beaucoup ou pas du tout compte des résultats publiés des sondages. Les personnes interrogées devaient situer leur position sur une échelle de zéro à 10. Un dixième de ces personnes se cotaient entre 7 et 10 et manifestaient donc leur influençabilité aux sondages. Les plus certaines d'aller voter manifestaient plus souvent cette influençabilité que les autres. Or, les électorats potentiels de Le

¹¹ Au cours du débat, le coût des enquêtes sur les intentions de vote réalisées du 6 février au 6 mai 2017 par un institut de sondage a été estimé à environ 100 000 €.

Pen, Macron et Mélenchon étaient également plus sensibles aux résultats publiés que les autres, ce qui accrédi-terait l'idée que les sondages peuvent contribuer, minoritairement, à former l'opinion.

Questions techniques et juridiques

Qu'en est-il des **marges d'erreur** dans les enquêtes par quotas ? demande un participant. Le calcul des marges d'erreur, mathématiquement assuré dans le cas des enquêtes auprès d'échantillons tirés au hasard, l'est moins pour les enquêtes par quotas. Peut-on néanmoins extrapoler de l'aléatoire aux quotas ?

L'invité indique qu'en 2010 le Sénat a publié des papiers de statisticiens sur le sujet. Les marges d'erreur sont en effet calculées pour les sondages aléatoires mais semblent être calculables aussi pour les enquêtes par quotas, dans lesquelles une part d'aléatoire intervient avec la stratification géographique. On peut même hasarder que la méthode des quotas n'accroît pas la marge d'erreur. Quoi qu'il en soit, la loi oblige à publier des marges d'erreur. Certes, la loi citée plus haut a été votée en urgence avec une partie sur les sondages ajoutée au dernier moment, mais il faut bien que les instituts de sondage la respectent.

Ce débat est l'occasion pour un participant d'exprimer l'idée que pour le deuxième tour il aurait fallu que les échantillons soient en adéquation avec la population qui allait voter ; or, il y avait une grande hésitation chez beaucoup d'électeurs. Une autre personne rebondit pour dire que l'abondance du nombre d'abstentions et de bulletins blancs et nuls ne pouvait manquer d'entraîner un problème d'estimation puisqu'on est face à un rapport de deux variables aléatoires, le dénominateur n'étant plus quelque chose de certain. Un troisième participant, qui travaille dans un institut de sondage, précisera que le calcul de la marge d'erreur se fait sur l'intégralité des échantillons, y compris les abstentions, les votes blancs et les votes nuls. Une autre personne enfin souhaiterait qu'on sache distinguer l'abstention habituelle et l'abstention de caractère exceptionnel.

Les **redressements** ont soulevé quelques questions, surtout d'ordre juridique, étant rappelé au préalable que les techniques de redressement sont conservatrices. Dès lors, quand il y a un changement de structuration des intentions de vote, lié notamment à la mobilité du vote, la référence aux votes passés ne fonctionne pas bien ; mais il n'y a pas d'autre moyen de caler les redressements.

Les instituts de sondage ont-ils l'obligation de dire ce qu'ils font ? Quelle est sur ce point l'interprétation de la loi ? Aux yeux de l'invité, la loi est mal rédigée sur ce point et chacun peut publier ses calculs de redressement ou ne pas le faire. Opinion Way publie pour sa part les résultats bruts de ses enquêtes et ses calculs de redressement. Les méthodes de redressement étant en fait les mêmes pour tous les instituts de sondage, l'invité pense que la demande croissante de transparence dans tous les domaines les amènera tous à en dire davantage sur leurs méthodes. D'ailleurs, la loi sera peut-être aménagée un jour pour être plus précise dans ses préconisations.

Cette interprétation de la loi est contestée par un participant qui en trouve le texte clair et contraignant, même si la loi a été adoptée de manière précipitée en ce qui concerne sa partie relative aux sondages. Le texte dit en effet¹² : « *Avant la publication ou la diffusion de tout sondage ... l'organisme qui l'a réalisé procède au dépôt auprès de la commission des sondages ... d'une notice précisant au minimum : ... 7° S'il y a lieu, les critères de redressement des résultats bruts du sondage.... Dès la publication ou la diffusion du sondage, toute personne a le droit de consulter auprès de la commission des sondages la notice prévue par le présent article ; cette commission rend publique cette notice sur son service de communication au public en ligne.* »

Mais, selon l'invitée, la loi parle des critères de redressement, sans autre précision. Qu'en est-il au juste de ces critères et surtout des calculs effectués ? Il reste à son avis des marges de manœuvre pour les instituts de sondage.

¹² Loi n°2016-508 du 25 avril 2016.

Pourrait-on procéder autrement que par des sondages électoraux ?

L'invité apporte sur ce point des précisions relatives aux **travaux de modélisation** conduits par Opinion Way. Les structures démographiques et sociales du corps électoral sont connues avec précision, à des niveaux géographiques fins, grâce aux statistiques publiques. Opinion Way les utilise en travaillant sur les 36 000 communes regroupées en 6 000 entités : les 32 000 communes les plus petites sont regroupées sur une base géographique, les autres restant traitées isolément. On applique ensuite, à chacun des territoires élémentaires ainsi définis, les résultats des dernières élections et des sondages les plus récents pour explorer l'élection à venir, en tenant compte de la sociologie des votes exprimés. Puis on agrège au niveau national. Dans le cas du premier tour de la présidentielle 2017, on n'a pas utilisé dans un premier temps les résultats des dernières élections régionales : le résultat n'a pas été bon, qui mettait en tête Macron et Fillon. Dans un deuxième temps on a entré dans le modèle les résultats des dernières élections régionales : le résultat a été bon puisqu'il donnait Macron et Le Pen en tête, dans cet ordre. Le modèle n'est donc pas vain et ces réflexions permettent de conforter les résultats des sondages. Ce travail permettrait par exemple de considérer qu'un score de 23 à 25 % en faveur de Macron était possible au 1^{er} tour. Cette modélisation est encore à un stade exploratoire mais semble prometteuse pour améliorer la qualité des sondages sur les intentions de vote.

Un participant s'étonne qu'Opinion Way ait déjà produit une estimation du **nombre de sièges par parti aux prochaines législatives**¹³. N'est-il pas risqué de publier cela alors que l'incertitude est pourtant maximale ? De fait, le modèle qui vient d'être évoqué a été utilisé pour les prochaines élections législatives, mais en repassant au niveau de la circonscription électorale. On ne connaît pas, certes, tous les candidats mais on connaît les partis qui seront en lice. Est-il absurde de penser qu'Emmanuel Macron puisse avoir une majorité présidentielle ? Le modèle répond que non. Mais on appréhende mal « l'effet sortants », notamment en ce qui concerne les socialistes, puisqu'il faut remonter en ce qui les concerne à 1993 pour trouver des effectifs de sortants comparables à ce à quoi on s'attend pour 2017. Cela dit, les enquêtes habituelles sur les intentions de vote vont reprendre et on devrait savoir si les électeurs voudront donner une majorité au nouveau président. Dès le 19 mai, on connaîtra le nom de tous les candidats et on pourra donc proposer des noms aux personnes enquêtées.

Pour l'élaboration du résultat de ces enquêtes, le premier tour des présidentielles 2017 sera pris en compte, avec cette difficulté que l'électorat d'Emmanuel Macron est hétéroclite à cause du vote utile : même dans des départements comme le Val-de-Marne ou les Hauts-de-Seine, qui ont porté Emmanuel Macron largement en tête, il y a eu des poches très favorables à Mme Le Pen. Les « non votes d'adhésion » posent problème en effet¹⁴. Assistera-t-on à une cristallisation du vote Macron au bénéfice de son parti ?

Chaque institut de sondage dispose de deux élections de référence : 2012 et les dernières élections régionales : il appartient à chacun d'arbitrer entre ces deux références. Face à plusieurs hypothèses, chaque institut cherche celle qui marche le mieux de son point de vue¹⁵. En général, c'est la dernière élection présidentielle pour une présidentielle. En tout cas la Commission des sondages vérifie que l'institut de sondage ne modifie pas ses critères en cours de route.

L'essai de modélisation fait par Opinion Way intéresse beaucoup un participant, qui serait friand d'en avoir une connaissance fine pour pouvoir le critiquer (positivement, s'entend !).

¹³ Voir : <https://www.opinion-way.com/fr/sondage-d-opinion/sondages-publics/politique/la-lettre-de-l-opinion-legislatives-2017-vers-une-recomposition-inedite-juin-2016.html>

¹⁴ Cela dit, le vote en faveur de Macron est sociologiquement plutôt homogène.

¹⁵ Cette manière de faire trouble un participant qui se demande s'il n'y a pas matière à interrogation déontologique dans le fait de chercher au coup par coup la meilleure méthode pour présenter les résultats.

Une autre personne demande s'il a été procédé à un appariement des quartiers Insee et des bureaux de vote, par exemple grâce aux fichiers Cartelec¹⁶ ? Il existe en effet un projet de recherche de quelques laboratoires qui consiste à appairer les bureaux de vote et les Iris¹⁷. C'est compliqué à faire parce que les découpages évoluent de part et d'autre. Il existe des bases nationales de bureaux de vote mais leur utilisation est rendue complexe par ces évolutions.

Plusieurs participants demandent s'il était pertinent de **faire des sondages sur le deuxième tour de l'élection avant d'avoir les résultats du premier tour** ? Ils notent que cette pratique semble condamnée par tel commentateur ayant du métier en matière de sondages.

Le Cevipof indique pratiquer de tels sondages dans une optique de recherche mais ne publier de résultats que pour les sondages réalisés une fois connus les résultats du premier tour. Pour sa part, l'invité s'amuse des positions actuelles du commentateur évoqué, qui pratiquait naguère à grande échelle des enquêtes de cette sorte. Celles-ci disent en fait des choses intéressantes. Ainsi, les sondages sur le deuxième tour avant que le premier tour n'ait eu lieu permettaient de savoir avec certitude que Mme Le Pen ne gagnerait pas cette élection. A contrario, il était certainement réducteur en 2002 de s'interroger sur le deuxième tour avant de connaître les résultats du premier tour en n'opposant que Chirac et Jospin...

Que penser de l'**exploitation des données massives**, et notamment de celle des informations recueillies sur les réseaux sociaux ? demandent deux participants, qui citent, l'un les travaux d'un organisme canadien (qui ont débouché sur une lourde erreur), l'autre une entreprise lyonnaise (qui a produit à partir de mégadonnées des estimations de vote avec deux chiffres après la virgule et sans marge d'erreur !). Du point de vue de l'invité, l'exploitation des données collectées sur les réseaux sociaux ne doit pas être écartée par principe, tout en sachant que pour le moment elle fournit avant tous des éléments de contexte. Mais la traduction en pourcentages de ce qu'on trouve au sortir de telles exploitations est plus que discutable selon lui, et encore davantage quand on prétend fournir des estimations avec deux chiffres après la virgule. Pour sa part, Opinion Way n'a pas voulu d'estimation comportant des demi-points (ce qui a posé quelques difficultés pour les candidats recueillant très peu d'intentions de vote).

D'autres participants se déclarent troublés par la **faible variabilité des résultats des enquêtes d'intention de vote réalisées par les différents instituts de sondage**. Certes, les techniques mises en œuvre doivent être homogènes mais la convergence des estimations était frappante. Pourtant il y a eu dans cette campagne électorale plusieurs événements exogènes d'importance¹⁸ et, entre les deux tours, un débat inhabituel dans son déroulement. On maîtriserait donc à ce point, non seulement la prise en compte des votes inavoués, mais aussi l'impact des événements ?

¹⁶ Le projet **Cartelec** développe un système d'information géographique permettant d'analyser les résultats des élections de 2007 (présidentielle et législatives), 2009 (européennes) et 2010 (régionales) au niveau des bureaux de vote pour la France métropolitaine. Grâce aux outils de cartographie assistée par ordinateur, les périmètres des bureaux de vote sont mis en relation, d'une part, avec les résultats enregistrés lors de ces scrutins et, d'autre part, avec les caractéristiques démographiques et socio-économiques des territoires concernés. Voir <http://www.cartelec.net/>

¹⁷ Afin de préparer la diffusion du recensement de la population de 1999, l'Insee avait développé un découpage du territoire en mailles de taille homogène appelées IRIS2000. Un sigle qui signifiait « Ilots regroupés pour l'information statistique » et qui faisait référence à la taille visée de 2 000 habitants par maille élémentaire. Depuis, l'Iris (appellation qui se substitue désormais à IRIS2000) constitue la brique de base en matière de diffusion de données infra-communales. Il doit respecter des critères géographiques et démographiques et avoir des contours identifiables sans ambiguïté et stables dans le temps (mais le découpage en Iris peut être affecté par les modifications de la géographie communale). Les communes d'au moins 10 000 habitants et une forte proportion des communes de 5 000 à 10 000 habitants sont découpées en Iris. Ce découpage constitue une partition de leur territoire. La France compte environ 16 100 Iris dont 650 dans les DOM. Par extension, afin de couvrir l'ensemble du territoire, on assimile à un Iris chacune des communes non découpées en Iris.

¹⁸ Les révélations du Canard enchaîné le 25 janvier, puis la mise en examen de François Fillon, ont été des événements marquants.

L'invité indique que la faible variabilité des résultats entre instituts de sondage est un élément nouveau. Les votes passés étant communément utilisés en France comme élément de redressement (contrairement aux pratiques des Anglo-Saxons), on peut avoir là un début d'explication. Par ailleurs, si un institut en copiait d'autres, cela se trahirait par des mouvements brutaux sur ses courbes d'évolution des intentions de vote. Cela dit, ajoute l'invité, il faut avoir en tête plusieurs éléments :

- le vote « inavoué » *a priori* se révèle nécessairement *a posteriori* et on en a donc une bonne connaissance ;
- le vote est différent de l'intention de vote mais plus on se rapproche de la date de l'élection, plus on limite l'écart entre l'intention et la réalisation ;
- les abstentions ont été bien estimées et les votes blancs ou nuls ont été estimés à un niveau correct. Mais les instituts ont rencontré des problèmes dans la structure des populations de votants et de non-votants ;
- les estimations pour le deuxième tour ont été moins bonnes que pour le premier à cause de tout ce qui s'est passé mais aussi pour des raisons techniques, liées au fait qu'il est plus facile de joindre un électeur de Macron qu'un électeur de Le Pen. Pour les électeurs de Le Pen, le monde est source de dangers et les instituts de sondage participent de cette malveillance envers eux. Dans les données brutes des enquêtes, il y avait donc beaucoup trop de votes en faveur de Macron à cause de la surreprésentation de ses électeurs ;
- il reste difficile de bien établir le lien entre un événement et une intention de vote¹⁹. Entre les deux tours, des évolutions importantes ont eu lieu à la suite du débat du mercredi soir ; il fallait un événement aussi fortement marqué pour que les mesures en révèlent l'impact.

Ce dernier point rejoint une remarque d'un participant, qui note que les mécanismes de l'intention de vote et ceux du vote ne sont pas forcément les mêmes et que le passage des uns aux autres peut être fortement influencé par le débat social. Ce participant pense que les médias devraient mieux expliquer ce qui se passe chez les gens, comme le quotidien *Le Monde* a pu le faire par un long article consacré aux déchirements intervenus entre les deux tours au sein des familles et des groupes d'amis.

Cela amène un autre participant à demander si on fait des **enquêtes post électorales** pour comprendre ce qui s'est passé ? A quoi l'invité répond que de telles enquêtes post électorales n'auraient d'intérêt que si tous les instituts de sondage acceptaient de mettre en commun l'ensemble de leurs redressements, ce qui ne semble pas près de se produire.

Prenant du recul sur les sondages, un participant note le rôle d'intermédiaires que jouent les instituts de sondage et les médias dans la chaîne des opérations qui vont de la collecte des informations jusqu'à la restitution au public des résultats chiffrés. Chacun joue sa partie dans un cadre complexe et mouvant. Les doutes sur les sondages résultant des surprises de l'élection de Donald Trump et du Brexit n'ont pas été confortés par l'élection présidentielle française. D'ailleurs, à y regarder de près, les sondages aux États Unis sont loin d'avoir failli puisqu'ils avaient prévu l'avance en voix de Hillary Clinton²⁰. Mais, dans chacun des huit États clés (les « Swing States »), les sondages effectués les jours précédant le 8 novembre 2016 donnaient une chance sur deux de gagner à Trump et à Clinton. Le raisonnement fait par les commentateurs a sans doute été le suivant : la probabilité que l'un ou l'autre des deux principaux candidats gagne chacun de ces États était de 1/2 ; mais la probabilité qu'un des deux candidats gagne simultanément dans les huit États considérés était très faible (0,5⁸) et donc, puisqu'il suffisait qu'Hillary Clinton gagne dans seulement deux ou trois de ces « Swing States » pour obtenir la majorité des grands électeurs au niveau fédéral, les commentateurs ont pu estimer que sa probabilité de gagner l'élection au niveau fédéral était sensiblement supérieure à 0,5. Mais ce raisonnement aurait supposé pour être exact que les événements dans les huit urnes que constituent les huit « Swing States », aient été indépendants. Ce qui n'était évidemment pas le

¹⁹ Au moins, il n'y aura pas d'effet Fillon aux législatives, remarque un participant !

²⁰ Avance estimée, il est vrai, à 2 millions de voix seulement contre près de 2,9 millions décomptées.

cas, et la probabilité que l'un ou l'autre des deux candidats gagne simultanément tous ces États était donc en fait plus voisine de 0,5 que de zéro. Les commentateurs auraient dû être prudents ; ce sont eux, plutôt que les instituts de sondage américains, qui ont failli.

De façon générale, des résultats différents de ceux que laissent entrevoir les sondages peuvent se produire avec tout système électoral quand les intentions de vote sont proches les unes des autres, et ce ne sont pas les sondages qu'il faut alors incriminer mais l'utilisation qui en est faite, en particulier dans le cas américain.

Enfin, ajoute ce participant, le métier de sondeur n'est pas exempt de risques politiques. En Équateur, un institut de sondage a produit des résultats d'intentions de vote défavorables au président sortant, ce qui a entraîné une forte répression à l'égard de ses membres. Une solidarité internationale des statisticiens est en train de s'organiser sur ce dossier.

Un message important des invités aura été que, malgré le succès des estimations des intentions de vote enregistré à l'occasion de l'élection présidentielle de 2017, bien des questions techniques nouvelles se posent aux instituts de sondage eu égard aux conditions de vie modernes, aux comportements citoyens, et notamment à la mobilité des opinions politiques des électeurs.



Annexes : Documents préparatoires

Société française de statistique
Groupe Statistique et enjeux publics
Café de la statistique du 9 mai 2017

Sondages électoraux : peut-on leur faire confiance ?

Documents préparatoires

Voici trois extraits d'articles sur le sujet.

Document 1 : extrait de « Un point de vue sur les sondages d'intentions de vote en 2017 », article de Jean Chiche paru dans Variances le 31/03/2017

Article complet : <http://variances.eu/?p=2105>

Si jusqu'au milieu des années 2000, les sondages étaient le plus souvent administrés en face à face ou par téléphone, depuis l'élection de 2007[6], les enquêtes par Internet sont devenues quasiment la règle en France comme dans bien des pays. Leur coût de production bien moindre a encore augmenté le nombre d'intentions de vote publiés. Outre l'avantage du faible coût économique, l'auto-administration des questionnaires réduit la désirabilité sociale liée à la présence d'un enquêteur. Donner des opinions extrémistes ou jugées non convenables est plus facile face à un écran. Mais qui sont les personnes interrogées ? Le plus souvent, les instituts ont leur propre panel d'internautes qu'ils entretiennent le mieux possible. Mais comment sont constitués ces « viviers » et qui accepte ou n'accepte pas d'y participer ? Le recrutement des panélistes est crucial. Les enquêtes de recrutement donnent des informations précieuses. Mais elles n'existent pas toujours ou sont confidentielles et non disponibles. Quand les panélistes sont simplement recrutés via des pop-up ou des incitations par mail ou des fichiers achetés à des brokers, que vaut vraiment la base de sondage ? On pourra élaborer le meilleur plan de sondage et interroger le plus bel échantillon possible, les profils sociologiques de ceux qui ne sont pas dans le panel ne seront jamais atteints. On apprend, dans le rapport du CREDOC pour l'ARCEP[7], que la population n'utilisant pas ou très peu Internet pour des raisons de couverture, de moyens financiers, ou par manque d'intérêt peut atteindre jusqu'à 19% de la population. De même la population la plus aisée à fort patrimoine est moins disposée à être volontaire pour répondre régulièrement aux enquêtes proposées par un institut. Ce ne sont pas les très faibles incentives qui peuvent les attirer. L'analyse de la composition socioéconomique des panels utilisés par les instituts permettrait de mieux prendre en compte la non-réponse totale. Et peut-être de la corriger si on avait une meilleure connaissance de ces populations. Ces questions sont majeures pour élaborer des plans de sondages corrects et donc pour tirer des échantillons de qualité.

Document 2 : extrait de « Sondages : peut-on (encore) leur faire confiance ? », article de Sylvain Tronchet paru sur franceinter.fr le 16/03/2017

Article complet : <https://www.franceinter.fr/politique/sondages-peut-on-encore-leur-faire-confiance>

En période électorale près de cent sondages politiques sont publiés tous les mois en France. Pourtant, les méthodes des instituts sont de plus en plus contestées.

Le « oui » des Britanniques lors du référendum sur le Brexit, l'élection de Donald Trump aux Etats-Unis, les désignations de François Fillon et de Benoît Hamon lors des primaires en France... Quel est le point commun entre tous ces scrutins ? A chaque fois, les instituts de sondage ont été incapables de prédire correctement leur issue. Evidemment, les électors, les méthodes des instituts de sondage différant trop d'un pays à l'autre, il est difficile de tirer une conclusion définitive de cette succession d'échecs. Mais le problème est bien là. Admis par une partie de la profession.

L'opinion est de plus en plus difficile à mesurer et les méthodes qui semblaient fonctionner jusque-là se révèlent défailtantes. Le constat est partagé par de nombreux spécialistes : la qualité des sondages s'est dégradée ces dernières années. Guenaëlle Gault, responsable du "digital" pour Kantar Sofres et ex-directrice du département "Opinion", constate que les humeurs des Français sont de plus en plus difficiles à capturer :

'Le centre de gravité de la société s'est déplacé du groupe à l'individu. Au sein d'un groupe, l'opinion avait tendance à ne pas beaucoup changer. Alors que les individus sont versatiles, de sorte qu'aujourd'hui tout va extrêmement vite, ce n'est plus du tout stable.'

Document 3 : extrait de « Filteris, Enigma... Face aux instituts de sondage, la défaite des prévisions « alternatives », article paru sur lemonde.fr le 24/04/2017

Article complet : http://www.lemonde.fr/pixels/article/2017/04/24/big-data-buzz-sur-les-reseaux-sociaux-la-defaite-des-previsions-alternatives_5116570_4408996.html#

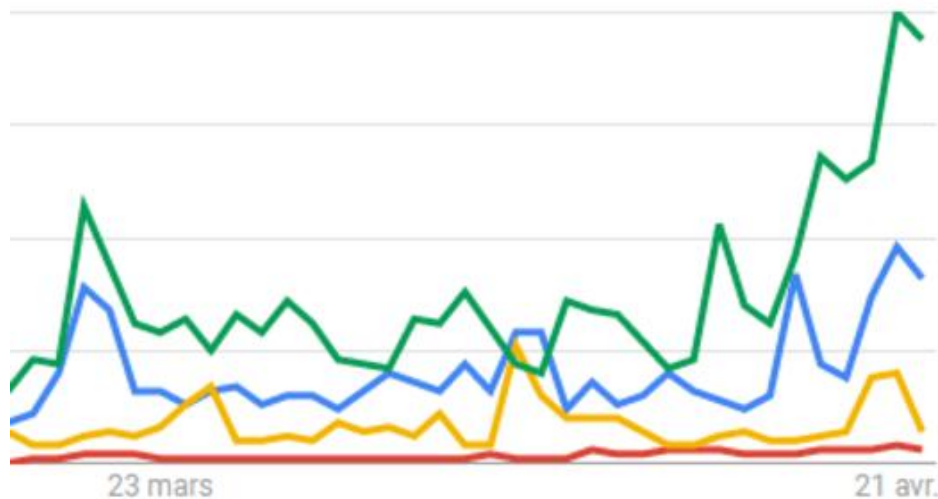
Elles étaient particulièrement prisées des militants : depuis des mois, des études se basant non pas sur des sondages mais sur le « big data », l'analyse des messages publiés sur les réseaux sociaux ou encore du profil sociodémographique des Français, prétendaient mesurer plus efficacement que les instituts de sondage les intentions de vote au premier tour de la présidentielle.

Ces études donnaient pour la plupart des résultats très différents des sondages classiques. L'entreprise canadienne Filteris affirmait ainsi que le second tour opposerait François Fillon et Marine Le Pen. Vigiglobe, une start-up française, ne donnait pas de classement prédictif, mais affirmait que François Fillon serait probablement au second tour, et jugeait Marine Le Pen en perte de vitesse. Plus affirmatif, Predict my president, un programme d'étudiants de l'école Telecom Paris Tech, donnait des scores très précis pour les quatre principaux candidats : Marine Le Pen (24,13 %), François Fillon (21,77 %), Emmanuel Macron (20,32 %), et Jean-Luc Mélenchon (18,66 %).

Elles avaient prévu la victoire de Trump

Toutes ces prédictions se sont donc révélées fausses, tout comme celles d'entreprises parues dans la presse étrangère. Le fonds d'investissement Leonie Hill Capital avançait ainsi que son programme d'intelligence artificielle voyait Marine Le Pen emporter 28 % des voix au premier tour, devançant Emmanuel Macron à 19 % ou 20 % et François Fillon à 16,4 %. Un peu plus prudente dans ses affirmations, l'agence suisse Enigma envisageait, elle, « l'hypothèse d'un duel Fillon-Mélenchon », qui « paraît tout à fait cohérente avec d'autres analyses, qualitatives cette fois ».

Surfant sur la méfiance envers les sondages, après les victoires de Donald Trump aux Etats-Unis et le Brexit britannique, les entreprises spécialisées dans le « big data » ont suscité l'engouement. La plupart en utilisant un argument massue : elles avaient, contrairement aux sondages, prévu la victoire de Donald Trump. C'était notamment l'un des points sur lesquels s'appuyait Filteris, dont les prédictions avaient la faveur de certains supporteurs de François Fillon, que l'entreprise donnait en tête à l'issue du premier tour. L'entreprise mettait aussi volontiers en avant ses prédictions passées en France, et notamment le fait qu'elle avait prévu le bon classement à la présidentielle de 2012, ou encore la victoire de François Fillon à la primaire de la droite et du centre.



Détail d'une recherche Google Trends sur des noms des candidats de l'élection présidentielle.

Mais l'entreprise oubliait aussi de préciser que son modèle prédictif avait connu des ratés majeurs – elle imaginait que François Fillon affronterait Nicolas Sarkozy au second tour de la primaire, et avait largement sous-estimé le score de Marine Le Pen en 2012 – sans oublier qu'elle s'était totalement trompée dans sa prédiction pour la primaire de la gauche gagnée par Benoît Hamon, voyant une victoire de Manuel Valls face à Arnaud Montebourg au deuxième tour.