


LA FRANCE SELON LES RECHERCHES GOOGLE

France | 

Recherche

J'ai de la chance



*Comment elles révèlent nos mœurs, nos peurs,
nos divisions et nos aspirations*

David Dubois

David Dubois

La France selon les recherches Google

Comment elles révèlent nos mœurs, nos peurs, nos divisions et nos aspirations

© David Dubois, 2022

ISBN numérique : 979-10-405-0325-5

Librinova”

www.librinova.com

Le Code de la propriété intellectuelle interdit les copies ou reproductions destinées à une utilisation collective. Toute représentation ou reproduction intégrale ou partielle faite par quelque procédé que ce soit, sans le consentement de l’auteur ou de ses ayants cause, est illicite et constitue une contrefaçon sanctionnée par les articles L335-2 et suivants du Code de la propriété intellectuelle.

Préambule

Durant l'Antiquité, les Romains avaient pris pour habitude d'adopter une représentation bicéphale du dieu du temps et de la transition, Janus – un visage tourné vers le passé et l'autre vers le futur. Dans notre monde très moderne, les technologies numériques ont pris la place convoitée de Janus : elles sont devenues des outils transitionnels – en particulier grâce à leur capacité à réduire les distances, faisant du monde un village. Nos nombreuses interactions au quotidien avec ces technologies nous les font d'abord voir comme des outils de connectivité : ce sont des ponts qui nous permettent de parler plus vite, plus loin, plus fort, avec plus de gens. C'est la face visible des technologies, celle qui nous est la plus familière. Cet ouvrage s'intéresse à la partie moins visible mais non moins importante de nos activités en ligne : les données que nous laissons derrière nous et ce qu'elles disent de nous. C'est l'autre visage du Janus numérique. En effet, si les outils numériques sont des ponts (et Google, le Golden Gate), nous y laissons mécaniquement nos empreintes à chaque passage. Loin d'être un sol dans lequel nos empreintes se fondraient à jamais, les supports digitaux (et ceux des GAFAM – Google, Apple, Facebook, Apple et Microsoft – en particulier) ont une remarquable capacité à imprimer et conserver celles-ci dans notre nouvelle mémoire numérique collective.

Comme Janus, dieu du futur et du passé, les outils numériques sont donc créateurs d'expériences, tout en cristallisant notre passé dans la roche digitale. Ils *connectent* et *collectent*, en même temps. Formidable avancée ou dangereuse aventure ? Cet ouvrage invite chacun à découvrir concrètement de quoi il retourne, en se focalisant sur les données des recherches faites sur Google. Celles-ci ont fait leurs preuves en tant qu'outils de connaissance et de prédiction dans de nombreux domaines des sciences sociales, et sont relativement accessibles, par exemple, grâce à l'outil gratuit Google Trends utilisé tout au long de l'ouvrage.

Pour beaucoup de lecteurs, Google est un remarquable outil de connaissance et de mesure *objectif* – Larry Page, co-fondateur de Google ne s'est-il pas lui-même fixé pour but « d'organiser toute l'information du monde » ? Pour ce faire, Google s'est doté de puissants algorithmes fondés sur des critères statistiques, rationnels, avec lesquels le grand public interagit au quotidien. Ainsi, Google

Maps se veut une reproduction fidèle des moindres recoins de notre planète. De même, le moteur de recherche Google s'appuie sur l'algorithme *Pagerank* qui mesure quantitativement la popularité d'une page web via le nombre de liens qui y renvoient à partir d'autres pages du Web. Mais en plus d'être un outil de mesure objectif, Google est aussi devenu un outil de mesure majeur de notre *subjectivité*. Celle-ci se manifeste au travers de nos requêtes journalières dans son moteur de recherche et, plus récemment, dans sa plateforme de vidéos YouTube : du contenu textuel, visuel ou des vidéos qui répondent à nos questionnements, reflètent nos besoins, dévoilent nos aspirations. C'est cet aspect de Google – comme portrait de notre subjectivité – que cet ouvrage dévoile.

Quelques mots sur l'auteur avant de commencer : je suis « tombé » dans la marmite du *big data* un peu par hasard. Ni ma formation originelle de rat de laboratoire (un doctorat en marketing, à l'intersection de la psychologie sociale et de la sociologie avec comme outil principal d'analyse l'expérience contrôlée en laboratoire), ni ma nature (je ne suis pas accro au digital) ne me prédestinaient à m'intéresser aux données numériques. Je n'ai compris leur importance qu'en 2012, alors que je développais un cours sur ce sujet à l'INSEAD qui m'a amené à rencontrer un grand nombre de dirigeants des GAFAM ou des BATX (Baidu, Alibaba, Tencent et Xiaomi). J'ai ainsi pris conscience du rôle croissant des données dans les stratégies commerciales de grandes entreprises, comme l'Oréal ou Air Liquide, mais aussi de PME et de start-ups, comme le leader du covoiturage Blablacar en France. Curieux par nature, je me suis alors demandé quel rôle de révélateur les données digitales pouvaient jouer au-delà des comportements de consommation, dans notre compréhension de phénomènes sociaux, économiques ou politiques plus larges.

L'idée de cet ouvrage a germé après mon départ pour Singapour. J'ai alors pris pour habitude – sans doute par nostalgie du pays ! – de suivre d'un peu plus près qu'auparavant les débats passionnés qui traversent nos médias nationaux. Lecteur ou auditeur attentif, je remarquais, lors de ces débats, un déficit de données quantifiées, seulement citées occasionnellement sans vraiment préciser leur source, ce qui me disais-je participait à une certaine cristallisation des débats. Je décidai alors de constituer une petite équipe de jeunes collaborateurs formés à scruter, collecter et analyser les données de recherche en France, sur lesquels je me suis beaucoup appuyé pour cet ouvrage. Qu'ils en soient ici remerciés chaleureusement, en particulier Mohar Chaudhuri, Eric Atega, Louis

Mangold et Ingrid Valle. Un grand merci également à Marie-Laure Cahier pour son aide précieuse tout au long de la rédaction de l'ouvrage, ainsi qu'à mes nombreux relecteurs amicaux : Mathieu Trepanier, Caroline Thibodeau, Bruno Kocher, Perrine Desmichel, Jean-Francois Mignot, Laurent Cesari, Emilie Couton et bien d'autres. Cet ouvrage doit beaucoup à leur regard aiguisé. Il va sans dire que je fais mienne toutes les erreurs restantes. Sans autre préambule, poussons ensemble la porte des données numériques !

Introduction

Il y a un milliard d'heures, l'Homo Sapiens moderne émergeait.

Il y a un milliard de minutes, la Chrétienté naissait.

Il y a un milliard de secondes, l'ordinateur d'IBM devenait disponible.

Il y a un milliard de recherches Google.... C'était ce matin.

Hal Varian, *Chief Economist* de Google

20 décembre 2013

Elles font tellement parties de notre quotidien que nous n'y pensons plus. Et pourtant, la plupart d'entre nous effectuent des recherches sur Google plusieurs fois par jour, et pour quelques-uns d'entre nous, des dizaines de fois par jour. Qu'est-ce qu'une recherche Google ? C'est un échange : le moteur de recherche nous donne accès à de l'information – sous forme textuelle, visuelle, vidéo ou même audio – qu'il a organisée pour nous en échange de l'empreinte de notre pensée, de notre (re)quête que constitue le mot-clé rentré. En agrégeant, anonymement bien sûr, ces mots-clés associés à un lieu et un moment précis, on peut quantifier et suivre en temps réel ce qu'est devenu notre poulx collectif : nos recherches en ligne, facilement accessibles depuis l'outil partagé Google Trends¹. Pour la première fois dans son histoire, l'humanité peut sonder sa mémoire commune. Vertigineux ! En pratique, l'observation des dynamiques d'intérêts autour de nous – dans nos villes, régions, pays – peut se révéler riche d'enseignements et ainsi éclairer nombre de décisions dans des domaines très divers.

Prenons quelques exemples concrets. Un boulanger de quartier s'interroge sur les nouveaux pains à intégrer à son assortiment pour satisfaire la demande locale, et le moment le plus approprié pour le faire ? Agrégées localement, les recherches en ligne effectuées par les internautes sur différents types de pains et graines le renseigneront sur les tendances récentes, par exemple au cours des derniers mois². Un responsable politique local souhaite sonder le sentiment

d'insécurité chez ses concitoyens ? Les recherches Google portant sur les vols, les agressions, les commissariats et les dépôts de plainte constitueront un baromètre *en temps réel* de l'insécurité (surtout depuis qu'un dispositif de dépôt de plainte en ligne a été mis en place en 2013). Un particulier hésite à investir dans une crypto-monnaie récemment mise en circulation ? Les recherches Google lui révéleront les dynamiques d'intérêt pour cette monnaie et lui donneront de précieux renseignements pour optimiser ses choix. Un patron de PME ou un artisan souhaitant exporter se questionne sur les marchés où ses produits et ses savoir-faire auront le plus de chance d'être valorisés ? Les recherches Google liées à son activité révéleront les pays ou régions du monde dans lesquels l'intérêt pour ses produits se manifeste avec le plus de fréquence³. Un commercial se demande comment affûter ses arguments de vente ? Un examen des recherches associées pour les produits ou les solutions qu'il propose lui communiquera de précieuses informations sur le niveau de connaissance du marché mais aussi sur les besoins, les attentes, et éventuellement les problèmes rencontrés par ses clients, qu'il pourra par la suite prendre en compte pour étoffer son offre et améliorer l'expérience client. Un journaliste prépare l'interview d'une personnalité ? Il pourra, en se penchant sur le type de requêtes sur cette personnalité, découvrir les aspects de celle-ci qui intéressent le plus le public et ainsi orienter ses questions. Enfin, un étudiant en sciences sociales réfléchit sur le développement de mouvements anti-vaccination ? Une plongée, même brève, dans les recherches Google constituera pour lui une piste de départ, lui permettant de localiser les « poches de résistance » aux vaccins dans l'espace et dans le temps. On comprend donc au travers de ces quelques exemples que l'étude des données de recherche constitue un outil décisionnel révolutionnaire, sur le plan professionnel comme personnel. Il permet soit de valider une hypothèse et d'apporter plus de certitude dans la prise de décision, soit de dévoiler des phénomènes économiques, politiques ou sociaux plus contre-intuitifs, à contre-courant des idées reçues et peu explorés faute de données.

Pourquoi introduire et plébisciter l'utilisation de ces données, en France en particulier ? Parce que l'approche empirique – c'est-à-dire la place accordée à l'expérience et à l'observation par rapport à la théorie et aux opinions – y est peut-être un peu moins développée que dans d'autres pays comme dans les pays anglo-saxons. Les données de recherche, qui sont par essence des observations faites sur les expériences quotidiennes de chacun, nous donnent justement un moyen d'y remédier, d'enrichir nos grilles d'analyses sur des sujets par

définition subjectifs car remplis sinon de passions, en tout cas d'émotions : nos mœurs, nos peurs, nos divisions et nos aspirations.

Le manque d'attrait relatif de la France pour l'empirisme, notamment dans le débat public sur des questions économiques ou sociales, dans lequel les idées priment souvent sur les données, tient à des raisons à la fois contextuelles, historiques, et philosophiques.

En premier lieu, nombre de sujets pourtant sensibles, comme l'immigration ou le communautarisme, échappent à la statistique faute de données fiables ou en raison de limitations juridiques, et nourrissent de ce fait beaucoup de fantasmes. Ainsi, la loi *Informatique et Libertés* du 6 janvier 1978, plus restrictive que dans beaucoup d'autres pays, interdit de « collecter ou de traiter des données à caractère personnel qui font apparaître, directement ou indirectement, les origines raciales ou ethniques, les opinions politiques, philosophiques ou religieuses ou l'appartenance syndicale des personnes, ou qui sont relatives à la santé ou à la vie sexuelle de celles-ci. » Bien qu'ayant des avantages évidents sur le plan de la protection des libertés individuelles à laquelle les Français sont très attachés, cette approche limite aussi notre capacité à quantifier ces sujets et peut conduire à une certaine crispation du débat. Les recherches en ligne – anonymes et agrégées localement – peuvent justement redonner à ces sujets une certaine transparence. Par exemple, les variations dans le temps et entre territoires des recherches pour la Bible ou le Coran éclaireront sous un nouvel angle les dynamiques confessionnelles ou communautaires en France (voir Chapitre 6). Contrairement aux données des médias sociaux ou des études déclaratives soumises à un biais de désirabilité sociale, les données de recherche reflètent des comportements privés, émanant d'un individu seul. Elles reflètent donc mieux les véritables désirs, valeurs, opinions politiques, orientations sexuelles ou comportements de consommation. En tant que telles, les données de recherche prédisent remarquablement des dynamiques économiques réelles, comme la confiance des consommateurs, l'effet de la publicité sur les achats de biens ou d'expériences et bien d'autres phénomènes politiques, sociaux ou économiques⁴.

Par ailleurs, en France – pays de Durkheim et de Bourdieu – les décideurs économiques et politiques font souvent la part belle aux données structurelles, accessibles grâce à des instituts comme l'INSEE qui quadrillent la société en fonction de critères sociodémographiques « froids », tels que l'âge, le genre, le métier, le niveau de revenu et d'éducation, et leurs dérivés comme le taux de chômage. Les indicateurs psychologiques et comportementaux mesurant les

expériences quotidiennes de nos concitoyens « à chaud », typiquement les enquêtes et sondages, sont certes abondants mais souvent critiqués à juste titre parce que portant sur des échantillons de petites tailles. De fait, la plupart des sondages portent sur des populations très limitées, dépassant rarement 10 000 individus (c'est-à-dire moins de 0,015 % des 67 millions de Français). De plus, les méthodes pour s'assurer de la représentativité des échantillons – comme la méthode des quotas qui s'appuie sur les données structurelles – sont de moins en moins pertinentes dans une France « archipelisée », dans laquelle les notions classiques de classe sociale ou d'appartenance politique perdent leur sens. Les instituts de sondages sont ainsi de plus en plus décriés pour leur incapacité à prédire les résultats électoraux sur le plan national comme sur le plan régional. À ces problèmes s'ajoutent des doutes croissants sur la méthode de recrutement des panels. Par exemple, le journaliste Luc Bronner du *Monde* a réussi très facilement à participer, sous des identités factices, à plus de 200 sondages proposés par Ipsos, IFOP, Kantar ou BVA⁵. Par contraste, les recherches Google, notre poulx collectif en temps réel, remédient en partie à nos angles morts analytiques, en prenant la température à chaud sur un échantillon jamais égalé dans l'histoire – l'échantillon *est* la population (en tout cas, la population connectée) ! Ayant un fort pouvoir d'évocation de nos projets (par exemple, « réussir une grande école », « s'expatrier », « fonder une famille »), nos peurs (par exemple « fin du monde », « installer une alarme contre cambriolage », ou bien « acheter un taser »), mais aussi de notre mal-être (par exemple « je me sens seul » ou « je suis boulimique »), elles constituent autant d'indicateurs de notre état d'esprit à portée de main (ou de clic !) susceptibles de déconstruire les mythes et de quantifier des tendances émergentes.

Enfin, notre tropisme pour la théorie fait des recherches en ligne un outil particulièrement novateur et pertinent. Nous pensant, du moins depuis Descartes, le pays de la logique, de la démonstration et de la linéarité de la preuve, nous semblons avoir – en partie – oublié ces deux premières composantes au profit de la preuve, après lui avoir ôté toutefois ses oripeaux de linéarité. De fait, la « preuve » flirte souvent avec le sensationnel, l'anecdotique, le cas particulier « visuel » censé valider nos croyances, au détriment de données plus quantitatives capables de donner à voir la complexité du réel.

Le tweet ou le fait divers sont ainsi souvent érigés en argument de poids, surtout depuis l'avènement des médias en continu comme BFMTV et CNEWS. Ils engagent souvent notre regard sur l'écume des choses plutôt que sur les