

## Jeux de mots



Erez Aiden et Jean-Baptiste Michel

### Uncharted

#### Big Data as a Lens on Human Culture

Riverhead Books, New York, 2013, 288 pages, 27,95 dollars (toilé).

Ce livre laisse de vastes territoires inexplorés. Mais, avec Erez Aiden et Jean-Baptiste Michel, inexploré ne veut pas dire laissé de côté; la *terra incognita* évoquée dans le titre serait plutôt une cartographie.

Ce livre raconte comment on peut collecter les milliards de mots contenus dans tous les livres du monde, tous ces mots qui étaient jusque-là perdus dans la signification du texte, autant dire dans des contrées inconnues, et que l'on peut désormais cartographier à l'envi. Les auteurs espèrent que ce processus de défrichage, qu'ils ont baptisé «culturomique», mettra au jour des aspects intéressants de notre culture. Ils se sont alliés à Google pour concevoir un outil Internet puissant, mais les perspectives qu'ils laissent entrevoir sont peut-être extravagantes.

Ce qui ne veut pas dire que leur livre ne soit pas drôle. Du reste, la page des remerciements est tout un programme : Aiden, qui remercie ses trois enfants, nous dévoile au passage le deuxième prénom d'une de ses filles, Banana. (Au moins, il est constant dans l'excentricité puisque son fils s'appelle Galileo!)

J'adore m'amuser, mais Aiden et Michel réalisent un travail scientifique

important, et ce n'est pas pour se faire plaisir qu'ils choisissent des exemples rigolos. Pas besoin de bases de données géantes pour nous convaincre du fait que le mot «chupacabra» (désignant une créature buveuse de sang aperçue à Porto Rico en 1995) est largement moins répandu que «Sasquatch» ou «monstre du Loch Ness». Il semble aussi ridicule de retracer l'évolution de l'usage de «argh» et «aargh» dans les livres publiés entre les années 40 (difficile de préciser leur date d'apparition à partir du graphique reproduit dans l'ouvrage) et l'année 2000. Avec une certaine audace, la jaquette du livre cite d'ailleurs le commentaire de la revue *Mother Jones* concernant leur application Ngram Viewer : «C'est l'outil le plus chronophage de toute l'histoire de l'Internet».

Témoigner du patrimoine culturel en chargeant des robots de lire chaque mot de chaque livre publié dans le monde est ambitieux. Alors, pourquoi ai-je dit que ce travail restait très partiel? Aiden et Michel reconnaissent que leurs recherches ne portent que sur un échantillon de mots minuscule et, même s'ils déclarent que Google a déjà scanné 30 millions d'ouvrages (et sans doute plus à ce jour), il en reste encore une centaine de millions à traiter.

En outre, si l'usage d'un mot constitue bien un indice de notre histoire culturelle, bon nombre de sources sont omises ici : les articles de journaux et de magazines, les lettres, les films, les interviews télévisées et radiophoniques, les transcriptions, les conférences, bref tout ce qui est écrit ou parlé, mais non publié dans un livre. De surcroît, après avoir été rédigés, les livres sont souvent relus et révisés, leur grammaire et leur orthographe corrigées, et ils peuvent même être traduits dans d'autres langues. Chaque auteur sait que l'éditeur adapte son texte au style de la maison d'édition. Je me demande si le langage utilisé dans les livres, même dans 30 millions de livres, constitue une source fiable pour étudier l'évolution des usages.

J'imagine que Aiden et Michel diraient que les ouvrages numérisés par

Google sont le seul matériau qu'ils avaient à leur disposition, mais, compte tenu de tout ce qui n'y est pas, leur théorie sur les possibilités d'examiner notre culture «à la loupe» est peut-être un tantinet excessive.

Quant aux graphiques, ils sont nombreux, austères, en noir et blanc (et gris), et sont la stricte interprétation des données. Je n'ai rien contre les graphiques en tant que tels, mais le

## Ce livre raconte comment collecter les milliards de mots contenus dans tous les livres du monde.

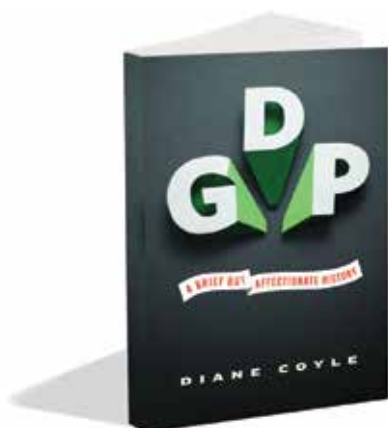
manque de vignettes et de pointillés et les courbes proprement dites (jusqu'à six dans un même graphique), imprimées dans des nuances de gris assez proches, rendent la lecture parfois difficile. Les auteurs nous parlent du Web, où tous ces problèmes ont été résolus : on utilise une couleur différente pour chaque courbe, on clique pour faire apparaître une vignette ou une date. C'est un exemple de ce qui différencie un graphique imprimé d'un graphique en ligne.

Mais soyons positifs. Les auteurs abordent, sans chercher à les esquiver, les questions de droits d'auteur posées par la numérisation et les moyens «virtuels» apparemment peu éthiques de les contourner. On apprend un détail charmant sur une expérience réalisée en 2002 par Larry Page et Marissa Mayer, qui avaient tenté de déterminer combien de temps pourrait prendre la numérisation de tous les livres du monde. Leur réponse était apparemment «des milliers d'années, voire une éternité». Pour savoir comment les auteurs ont surmonté ce problème, lisez leur livre; vous allez rire!

**Nigel Holmes**

Directeur d'*Explanation Graphics*, auteur de *Wordless Diagrams*, publié dernièrement, et de *The Book of Everything*

## Un avis classique à l'intention des économies modernes



Diane Coyle

### GDP

#### A Brief but Affectionate History

Princeton University Press, Princeton, New Jersey, 2014, 168 pages, 19,95 dollars (toilé).

**P**ourquoi Aristote n'a-t-il pas inventé le concept de produit intérieur brut (PIB) il y a deux mille ans, alors que le terme «économie» vient du grec *oikos* (maison)? Parce que sa science se concentrait sur les corps en mouvement.

Or voilà 50.000 ans que le PIB humain n'a pas bougé. À l'époque d'Aristote, la télévision n'agitait pas les marchés en annonçant des fluctuations de PIB. Personne n'espérait vivre mieux que ses parents. L'argent et les dettes n'étaient évidemment pas sans importance. Dans les années 1500, Henri VIII demandait à ses trésoriers de surveiller son ardoise au pub et les dépenses engagées dans les guerres contre la France, mais jamais il n'aurait demandé si le PIB par habitant avait augmenté.

Lucide et intelligent, cet ouvrage raconte l'histoire de cette mesure née au 20<sup>e</sup> siècle et qui, tous les trois mois, menace de faire chuter des premiers ministres. Coyle commence par nous rappeler quels sont les enjeux, pas dans la Grèce antique, mais de nos jours, à Athènes, où le directeur de l'institut national de statistique parle de son travail comme d'un «sport de combat». Elle nous relate l'expérience d'une de ses collègues, Paola Subacchi, qui s'était rendue à l'institut et s'attendait à voir

des supercalculateurs ou, au moins, un boulier. Mais elle s'était retrouvée dans une pièce poussiéreuse, sans ordinateur et avec une poignée de personnes.

Comme les instituts statistiques doivent produire des résultats, ils arrangent souvent les données pour vendre des obligations et obtenir des aides de pays tiers. D'après Coyle, les autorités chinoises se vanteraient parfois d'avoir un PIB élevé et le minoreraient en d'autres circonstances, pour pouvoir bénéficier de dons. Après l'effondrement de l'URSS, je suis allé à Saint-Petersbourg. Mes vieux manuels d'économie laissaient entendre que l'URSS avait connu une croissance soutenue pendant l'ère communiste. Même un lauréat du Prix Nobel comme Paul Samuelson avait publié ce genre d'informations discutables. Pourtant, il m'a suffi de respirer l'atmosphère moisie du musée de l'Hermitage pour comprendre que le communisme n'était pas seulement en retard par rapport aux pays de l'Ouest, mais même par rapport à la Russie de 1917!

Coyle rappelle que le calcul même du PIB incite les dirigeants à dépenser plus, parce que la hausse des dépenses publiques fait forcément augmenter la somme  $C + I + G + [X - M]$ . Les dirigeants n'ont qu'à ouvrir le robinet des dépenses et leurs gentils petits comptables feront grossir le PIB. De plus, la valeur des dépenses publiques est calculée sur la base des traitements versés aux fonctionnaires, et non sur la valeur qu'ils produisent. À Harvard, l'un de mes étudiants m'a dit un jour que, compte tenu de cette tautologie de départ, les dirigeants prêts à envisager ne serait-ce qu'un bref épisode d'austérité budgétaire mériteraient une médaille.

Coyle examine aussi d'autres inconvénients du PIB, par exemple le paradoxe qui veut qu'un veuf qui épouse sa femme de ménage fait baisser le PIB parce qu'il ne la paie plus. Il est particulièrement difficile d'estimer correctement la valeur monétaire d'un service dans une économie de l'information. Imaginons un tableau de chiffres pour aider les enfants à apprendre l'arithmétique. Quand les enfants s'initient aux additions grâce à ce

tableau baptisé Math Arrow, ils augmentent leurs perspectives de revenus disons de 100.000 dollars. Pourtant, l'application ne coûte que 4,99 dollars. Est-ce que chaque téléchargement du tableau crée une valeur de 100.000 ou de juste quelques dollars?

Coyle se demande aussi s'il serait possible de faire mieux et parcourt la liste des candidats : indice de développement humain, mesure du bien-être économique et autres indices de bonheur. Elle a raison d'être sceptique, surtout à l'égard des instruments proposés par les marchands de bonheur et les démagogues. Hugo Chavez a qualifié le PIB de «conspiration capitaliste». Mais il est encore plus facile de faire dire aux autres mesures ce que l'on veut. En 2009, le Costa Rica s'est retrouvé en tête du classement établi selon l'indice planète heureuse, Cuba arrivant en septième position. D'après cet indice, les habitants des zones sous autorité palestinienne étaient plus heureux et mieux portants que les Israéliens. Un porte-parole du mouvement pro-sioniste qui prétendrait que les Palestiniens sont mieux lotis que les Israéliens serait ridiculisé ou lapidé. Au fait, les États-Unis étaient classés au 114<sup>e</sup> rang. Bizarrement, je n'ai jamais vu un raideau partir de Miami pour rejoindre Cuba. C'est donc à juste titre que Coyle analyse à la fois les défauts du PIB et les raisons de s'inquiéter de sa disparition.

Je n'ai noté qu'une seule omission dans cet opuscule magistral. Quand je m'interroge sur le niveau de vie d'un pays, je me demande souvent combien d'heures doit travailler un ouvrier moyen pour acheter un poulet. Dans les années 20, le Président Hoover avait fait campagne en promettant un poulet dans chaque casserole. À l'époque, il fallait à peu près deux heures et demie pour gagner de quoi acheter un poulet. Aujourd'hui, il faut moins d'un quart d'heure. C'est tout de même un progrès, non? Sauf pour les poulets, bien sûr.

**Todd G. Buchholz**

Ancien directeur de la politique économique à la Maison Blanche, auteur de *New Ideas from Dead Economists* et *PDG de Sproglit*, un éditeur de logiciels