

On trouve ainsi de nouvelles formes de la loi de Benford (voir l'encadré 1). Le mot *un* est non seulement un chiffre et un adjectif numéral, c'est aussi un article. Pas étonnant donc qu'il domine sans la moindre équivoque tous les autres chiffres *deux, trois, ..., neuf*. En revanche, on reste émerveillé de la régularité des courbes pour *deux, trois, ..., neuf* et de leurs positions relatives.

Une autre curiosité numérique attend une explication. Si on compare les taux d'usage de *100 francs, 200 francs, 300 francs, 400 francs, 500 francs*, on trouve sans surprise que viennent en tête *100 francs* et *500 francs* (ce sont des chiffres ronds) suivis dans l'ordre de *200 francs, 300 francs, 400 francs*.

Cela est conforme à ce que la loi de Benford laisse attendre, en la corrigeant du fait que les chiffres ronds sont poussés en avant par les usages commerciaux qui dérangent un peu l'ordre naturel. La surprise vient de la croissance générale de tous les taux de citations (pour les cinq expressions) jusqu'en 1910, suivie d'une décroissance générale, elle aussi régulière. S'est-on plus particulièrement intéressé à l'argent vers 1910 ? Pourquoi donc ? Ou alors, l'explication est-elle liée à un changement dans les usages lexicaux, éditoriaux ou typographiques ? Étrange.

Une démarche qui a ses limites

Aussi impressionnantes que soient les quantités de données prises en compte pour produire les courbes, cela ne suffit pas à en faire un domaine scientifique nouveau. D'une part, la base reste quand même particulière (absence des journaux et revues). D'autre part, l'impossibilité d'accéder au corpus lui-même interdit souvent d'étudier les raisons de ce qu'on découvre. Par exemple, il faudrait disposer de statistiques sur la composition du corpus pour comprendre d'où vient l'augmentation de l'usage des chiffres isolés, où de celles des sommes comme *100 francs* vers 1910. La vue donnée par les *n-grammes* est certes intéressante, mais mille questions se posent dont les réponses ne peuvent pro-

venir que de tests menés directement et spécifiquement sur les cinq millions de livres numérisés, ou sur des sous-ensembles soigneusement constitués, dont il faudrait pouvoir disposer.

Notons encore que d'autres bases numériques seraient intéressantes à exploiter pour le linguiste, le sociologue ou l'historien, et que comme celle du grand corpus de livres, elles sont aujourd'hui interdites d'accès, et seraient encore plus difficiles à manipuler du fait de leur volume.

C'est le cas de la colossale base d'informations, qui enchanterait un sociologue, constituée par les milliards de fichiers des *datacenters* de Facebook dont on sait qu'elle concerne 500 millions d'inscrits et contient 30 milliards de photos. Les ingénieurs de Facebook indiquent recevoir 12 téraoctets d'informations nouvelles chaque jour, c'est-à-dire trois fois plus que le corpus géant dont nous avons parlé. Cela donne un total plusieurs centaines de fois plus gros que le corpus géant de livres, déjà difficile à manipuler. Le nombre de 30 000 téraoctets (3×10^{15} octets) a été cité pour le total des données de Facebook. Les données collectées par les moteurs de recherche Google, Bing, etc., ou de Gmail, seraient aussi intéressantes à explorer.

Le début de cet article mentionnait la difficulté d'évaluer certaines variables du monde social, économique ou politique que les sondages ne permettent pas de connaître. Les masses d'informations que certaines entreprises sur Internet constituent aujourd'hui sont de nouveaux objets du monde. On aurait pu croire qu'étant stockés sous format numérique, on les exploiterait facilement, contrairement aux données du monde réel non numérisées et difficiles à parcourir, car éparpillées. Ce n'est pas vrai. Outre les impossibilités liées aux droits de la propriété intellectuelle et aux règles de la protection de la vie privée qui limitent (heureusement !) l'accès des chercheurs aux bases de données des firmes sur Internet, les masses constituées sont aujourd'hui si volumineuses que, techniquement, elles ne sont exploitables que très partiellement et sans doute uniquement... par sondage.

L'AUTEUR



Jean-Paul DELAHAYE est professeur à l'Université de Lille et chercheur au Laboratoire d'informatique fondamentale de Lille (LIFL).

BIBLIOGRAPHIE

J. Evans et J. Foster, *Metaknowledge, Science*, vol. 331, pp. 721-725, 2011.

J.-B. Michel et al., *Quantitative analysis of culture using millions of digitized books, Science*, vol. 331, pp. 176-182, 2010 (<http://mfi.uchicago.edu/publications/papers/ScienceCulturomics.pdf>); **Supporting online material** : http://dericbownds.net/uploaded_images/Michel.SOM.pdf

J. Bohannon, *Google opens books to new cultural studies, Science*, vol. 330, p. 1600, 2010.

J. Bohannon, *The science hall of fame, Science*, vol. 331, p. 143, 2011 : <http://www.sciencemag.org/content/331/6014/143.3.full>

N. Gauvrit et J.-P. Delahaye, *Scatter and regularity implies Benford's law*, dans H. Zenil (ed.), *Randomness through complexity*, World Scientific, 2011.

N. Gauvrit, *Dépressivité et biais de positivité, Cahiers Romains de Sciences Cognitives*, vol. 3, n° 3, pp. 63-71, 2009.

SUR LE WEB

Site Internet de la culturomique : <http://www.culturomics.org/>

Page permettant d'engendrer des courbes : <http://ngrams.googlelabs.com/>

□ ART & SCIENCE

Le cerveau caché de Michel-Ange

Dans deux panneaux du plafond de la chapelle Sixtine, la Création d'Adam et la Séparation de la lumière et des ténèbres, Michel-Ange aurait dissimulé des planches anatomiques du système nerveux.

Loïc MANGIN

En 1505, Michel-Ange (1475-1564) quitte Florence et rejoint Rome à la demande du nouveau pape Jules II (1443-1513). Ce dernier confie au sculpteur le soin de lui concevoir un tombeau monumental dans la basilique Saint-Pierre : le mausolée ne sera jamais achevé... Après un exil de quelques mois à Bologne, Michel-Ange s'attelle à un autre projet, le plafond de la chapelle Sixtine qu'il peindra entre 1508 et 1512. Le projet fut proposé à l'artiste à l'instigation de l'architecte Bramante qui était persuadé de l'échec de son rival. Mal lui en prit, le plafond est un des plus célèbres chefs-d'œuvre de la Renaissance italienne !

Le plafond, de 40 mètres de longueur et 14 de largeur, est situé à 20 mètres de hauteur. Il est constitué de différentes surfaces en demi-lunes, en triangles, en rectangles... où Michel-Ange devait, selon son contrat, représenter les 12 apôtres. L'artiste vit plus grand et proposa d'y peindre des épisodes de la Genèse (voir la figure a), de la Séparation de la lumière et des ténèbres jusqu'à l'ivresse de Noé en passant par la Création d'Adam. Les murs latéraux étaient déjà ornés de fresques où le Pérugin, Botticelli et d'autres ont raconté les vies du Christ et de Moïse.

Intéressons-nous à la Séparation. Les historiens ont longtemps été intrigués par les irrégularités du cou de Dieu dans ce panneau (voir les figures b et c) et par le traitement de la lumière dans cette région du corps. Le personnage, Dieu, est éclairé en diagonale, du bas à gauche vers le haut à droite, mais le

cou semble illuminé par une lumière venant de la droite. Est-ce une erreur ? Ian Suk et Rafael Tamargo, experts en neuroanatomie à l'École de médecine de l'Université Johns Hopkins, à Baltimore, y voient plutôt le signe d'un message caché...

D'abord, ils ont vérifié qu'aucun autre cou peint par Michel-Ange n'est aussi étrange et « grumeleux ». Ensuite, ils ont montré que les détails du cou se superposent à l'anatomie d'un cerveau humain vu du dessous (voir la figure d). De fait, la correspondance est étonnante.

Les détails du cou de Dieu se superposent à l'anatomie du cerveau humain.

Les deux neurologues vont plus loin et discernent la moelle épinière dans les replis du drapé qui traversent le torse (voir la figure e). De plus, au niveau de l'abdomen, un pli en Y semble se prolonger par deux protubérances sphériques situées sous le thorax (voir la figure f) : ce schéma coïncide précisément avec un dessin du nerf optique et des yeux, effectué par Léonard de Vinci en 1487. Or les deux artistes sont contemporains et s'apprécient mutuellement.

En 1990, Frank Meshberger avait déjà remarqué que, dans la Création d'Adam, le rideau entourant les anges et Dieu reproduit une coupe transversale de cerveau humain

(voir les figures g et h). Précisons que d'autres y virent un rein, un organe qui fit beaucoup souffrir Michel-Ange...

De même que nous voyons un visage dans le moindre nuage, ces ressemblances trahissent peut-être la trop grande habitude qu'ont les deux Américains des planches neuroanatomiques. Pourtant, Michel-Ange disposait des connaissances requises pour « cacher » un cerveau. En 1493, il offre au prêtre de la basilique Santa Maria del Santo Spirito, à Florence, un crucifix en bois peint. En échange, les autorités ecclésiastiques l'autorisent à examiner, et à disséquer, le corps de défunts en provenance de l'hôpital du couvent. Plus tard, il illustrera un traité d'anatomie de son ami, le chirurgien Realdo Colombo.

Si la ressemblance n'est pas fortuite, que signifie cette planche anatomique cachée dans la Séparation ? Selon I. Suk et R. Tamargo, Michel-Ange aurait cherché à rehausser le sens de ce panneau. Le neurologue américain Douglas Fields s'interroge : n'y aurait-il pas là une illustration de la lutte entre la religion et la science ? Ou bien Michel-Ange postule-t-il que l'intelligence – et l'organe qui en est le support, comme Hippocrate le pensait déjà – suffit au croyant pour communiquer avec Dieu, rendant l'Église inutile ? On comprend alors les précautions prises pour dissimuler la planche anatomique ! ■

I. Suk et R. Tamargo, *Concealed neuroanatomy in Michelangelo's Separation of light from darkness in the Sistine Chapel*, *Neurosurgery*, vol. 66(5), pp. 851-861, 2010.