

0 La statistique littéraire

Tout le monde croit en la loi des grands nombres.

Pour être plus précis, tout le monde a l'intuition que dans un grand échantillon, enlever une valeur ne peut pas changer grand-chose à la moyenne. De là à s'imaginer qu'il existe toujours une constante idéale vers laquelle des moyennes empiriques doivent converger, c'est autre chose. Cela peut être le cas quand il s'agit de mesures physiques. C'est beaucoup moins évident dans le domaine humain, et en particulier littéraire.

1 la Maison de la Sagesse à Bagdad

Pourtant, c'est par là que la statistique a commencé, pour autant que l'on sache. Le premier ouvrage connu parlant de statistique a été écrit à Bagdad, à la Maison de la Sagesse. L'auteur est un des plus grands savants arabes, peut être le premier savant universel, al-Kindi.

2 de l'ordre de 300 traités sur...

Il a écrit de l'ordre de 300 traités. Sur l'éthique, la pharmacologie, la zoologie. Sur les bijoux, le verre, les marées. Aussi sur la linguistique, la phonologie, et la poésie.

Al-Kindi est connu comme le philosophe du bonheur, à cause de son « Art de combattre la tristesse ».

histoires de statistique

La statistique littéraire

croire en la loi des grands nombres



hist-math.fr

Bernard YCART

la Maison de la Sagesse à Bagdad

Al Kindi (ca. 800-880)



de l'ordre de 300 traités sur...

Al Kindi (ca. 800-880)

- métaphysique, éthique, logique,
- médecine, pharmacologie, psychologie,
- linguistique, phonologie, poésie,
- arithmétique, géométrie, astronomie, musique
- physique, chimie, zoologie,
- parfums, épées, bijoux, verre, teintures, marées, miroirs, météo, tremblements de terre...

7 Fréquences sur 3667 lettres

C'est le premier recueil de données connu de l'histoire de la statistique. Le plus étonnant là-dedans, ce n'est pas que al-Kindi ait cru en la réalité tangible de la fréquence de chaque lettre en arabe, ni qu'il ait eu foi en ses comptes sur 3667 lettres pour approcher ces fréquences absolues.

Non, le plus étonnant est qu'on ait continué à croire en cette même chimère pendant des siècles après al-Kindi, jusqu'au vingtième siècle. Voici des extraits d'un manuel de cryptologie militaire datant de 1916.

Fréquences sur 3667 lettres

Al Kindi (ca. 800-880)

| Letters | Frequency | Letters | Frequency | Letters | Frequency | Letters | Frequency |
|------------------------|-----------|---------|-----------|---------|-----------|---------|-----------|
| ā (ا) | 600 | n (ن) | 221 | k (ك) | 112 | l (ل) | 35 |
| l (ل) | 437 | r (ر) | 155 | d (د) | 92 | s (س) | 32 |
| m (م) | 320 | ' (ع) | 131 | s (س) | 91 | h (ح) | 20 |
| h (ه) | 273 | f (ف) | 122 | q (ق) | 63 | t (ت) | 17 |
| ^(*) ū+w (و) | 262 | t (ت) | 120 | h (ح) | 57 | ʔ (ظ) | 15 |
| ^(*) i+y (ي) | 252 | b (ب) | 112 | z (ز) | 46 | z (ز) | 15 |
| | | | | | | z (ز) | 8 |

8 Manual for the solution of military cyphers (1916)

« Si on compte dix mille lettres consécutives d'un texte en notant la fréquence d'occurrence de chaque lettre, les nombres trouvés seront pratiquement identiques à ceux obtenus à partir d'un autre texte de dix mille lettres dans la même langue. La proportion relative d'occurrence des différentes lettres sera à peu près identique, même pour des textes très courts. Un tel décompte d'un grand nombre de lettres, mis sous forme de table, s'appelle une table de fréquences. Chaque langue a sa propre table de fréquences et, pour chaque langue, la table de fréquences est presque aussi fixe que l'alphabet. »

Tout juste si l'auteur reconnaît des différences entre les styles, pour les minimiser aussitôt.

Manual for the solution of military cyphers (1916)

Parker Hitt (1878-1971)

Si on compte dix mille lettres consécutives d'un texte en notant la fréquence d'occurrence de chaque lettre, les nombres trouvés seront pratiquement identiques à ceux obtenus à partir d'un autre texte de dix mille lettres dans la même langue. La proportion relative d'occurrence des différentes lettres sera à peu près identique, même pour des textes très courts. Un tel décompte d'un grand nombre de lettres, mis sous forme de table, s'appelle une table de fréquences. Chaque langue a sa propre table de fréquences et, pour chaque langue, la table de fréquences est presque aussi fixe que l'alphabet.

9 Manual for the solution of military cyphers (1916)

« Il existe des différences mineures entre des tables de fréquences préparées à partir de textes sur des sujets différents. Par exemple si le texte est un article de journal, la table de fréquences s'écartera légèrement d'une table préparée à partir d'ordres militaires, ainsi que d'une table préparée à partir de messages télégraphiques. Mais ces écarts sont mineurs comparés aux écarts entre les tables de deux langues différentes. »

Pourquoi est-il étonnant que cette croyance ait perduré? Eh bien parce que depuis longtemps on savait que dans toutes les langues, on pouvait écrire des textes entiers qui n'utilisaient pas une lettre donnée. On appelle ces jeux littéraires des « lipogrammes ». Le plus célèbre en français est « La disparition » de Georges Perec. Perec s'amuse à décrire l'objet de la disparition : précisément la lettre E.

Manual for the solution of military cyphers (1916)

Parker Hitt (1878-1971)

Il existe des différences mineures entre des tables de fréquences préparées à partir de textes sur des sujets différents. Par exemple si le texte est un article de journal, la table de fréquences s'écartera légèrement d'une table préparée à partir d'ordres militaires, ainsi que d'une table préparée à partir de messages télégraphiques. Mais ces écarts sont mineurs comparés aux écarts entre les tables de deux langues différentes.

10 La disparition (1969)

« Ainsi, parfois, un rond, pas tout à fait clos, finissant par un trait horizontal : on aurait dit un grand G vu dans un miroir. Ou, blanc sur blanc, surgissant d'un brouillard cristallin, l'hautain portrait d'un roi brandissant un harpon. Ou, un court instant, sous trois traits droits, l'apparition d'un croquis approximatif, insatisfaisant : substituts saillants, contours bâtards profilant, dans un vain sursaut d'imagination, la Main à trois doigts d'un Sardon ricanant. »

Et en plus c'est du beau français ! Pérec était loin d'être le premier. Un roman sans la lettre E avait été écrit en anglais en 1939 par Ernest Wright.

11 Gadsby (1939)

« If youth, throughout all history, had had a champion to stand up for it; to show a doubting world that a child can think; and, possibly, do it practically; you wouldn't constantly run across folks today who claim that "a child don't know anything." »

Les lipogrammes existaient déjà chez les Grecs. On a trouvé des réécritures entières de l'Iliade sans A, dans E, sans I.

12 Varios effetos de amor (1671)

Voici la page de titre d'un livre de 1671 écrit en espagnol. Il est fait de cinq nouvelles. La première est écrite sans la lettre A, la seconde sans la lettre E, etc.

Que l'on puisse biaiser à loisir les fréquences des lettres aurait dû être une évidence depuis de nombreux siècles. Pourtant ça ne l'était pas, et on a cru dès le début en l'existence des probabilités que l'on pouvait approcher par des fréquences. Voici comment John Arbuthnot l'exprime dans le premier traité en anglais sur les probabilités.

13 Of the Laws of Chance (1692)

« Or, il y a un calcul des probabilités fondées sur l'expérience, dont on peut faire usage dans les paris à propos de toute chose. Imaginons qu'il s'agisse des chances qu'une femme accouche d'un garçon, et que vous vouliez connaître les chances exactes, vous devez considérer la proportion des garçons sur les filles dans les registres de naissance. »

Vous le voyez, il n'est pas question de se demander quel sens a la probabilité qu'une femme accouche d'un garçon, ni si un tel objet existe.

La disparition (1969)

Georges Pérec (1936-1982)

Ainsi, parfois, un rond, pas tout à fait clos, finissant par un trait horizontal : on aurait dit un grand G vu dans un miroir.

Ou, blanc sur blanc, surgissant d'un brouillard cristallin, l'hautain portrait d'un roi brandissant un harpon.

Ou, un court instant, sous trois traits droits, l'apparition d'un croquis approximatif, insatisfaisant : substituts saillants, contours bâtards profilant, dans un vain sursaut d'imagination, la Main à trois doigts d'un Sardon ricanant.

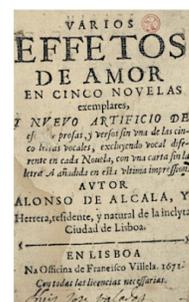
Gadsby (1939)

Ernest V. Wright (1872-1939)

If youth, throughout all history, had had a champion to stand up for it; to show a doubting world that a child can think; and, possibly, do it practically; you wouldn't constantly run across folks today who claim that "a child don't know anything."

Varios effetos de amor (1671)

Alonso de Alcalá y Herrera (1599-1682)



Of the Laws of Chance (1692)

John Arbuthnot (1667-1735)

There is likewise a Calculation of the Quantity of Probability founded on Experience, to be made use of in Wagers about any thing; it is odds if a Woman is with child but it shall be a Boy, and if you would know the just odds, you must confider the Proportion in the Bills that the Males bear to the Females.

14 Jacob Bernoulli (1655–1705)

C'est Jacques Bernoulli qui a le premier démontré la loi des grands nombres. Mais qu'a-t-il fait exactement et qu'est ce que c'était pour lui que la loi des grands nombres ?

Jacob Bernoulli (1655–1705)



15 Ars Conjectandi (1713)

L'énoncé se trouve dans la quatrième partie de l'Ars Conjectandi, paru en 1713. Bernoulli était très fier de ce résultat. Il le comparait à la quadrature du cercle et estimait qu'il lui avait coûté dix ans d'efforts. C'est un peu plus fin que ce qu'on entend intuitivement par loi des grands nombres. Voici comment il explique la différence.

Ars Conjectandi (1713)

Jacob Bernoulli (1655–1705)



16 peu de mérite à démontrer ce que personne n'ignore

« Même quelqu'un de très stupide, sans aucune instruction, guidé par un instinct naturel (ce qui est assez miraculeux) est sûr que plus grand sera le nombre d'observations, plus faible sera le risque de s'éloigner du but. Bien que tout le monde sache cela naturellement, la démonstration dérivée de principes scientifiques n'est pas triviale, nous devons donc l'exposer ici. Cependant, j'estimerai avoir peu de mérite à démontrer ce que personne n'ignore. Il reste autre chose que probablement personne jusqu'ici n'a jamais envisagé, même dans ses pensées. Il reste en effet à rechercher, quand le nombre d'observations augmente continûment, si la probabilité d'atteindre le vrai rapport entre les nombres de cas où un événement peut se produire ou non, augmente de telle sorte qu'elle dépasse n'importe quel degré de certitude. »

Eh oui, Bernoulli ne se contente pas de montrer que les fréquences expérimentales convergent vers la probabilité, il donne une borne sur la probabilité de trouver la fréquence expérimentale dans un intervalle donné autour de la probabilité, pour un nombre d'expériences fixé.

peu de mérite à démontrer ce que personne n'ignore

Jacob Bernoulli (1655–1705)

Même quelqu'un de très stupide, sans aucune instruction, guidé par un instinct naturel (ce qui est assez miraculeux) est sûr que **plus grand sera le nombre d'observations, plus faible sera le risque de s'éloigner du but**. Bien que tout le monde sache cela naturellement, la démonstration dérivée de principes scientifiques n'est pas triviale, nous devons donc l'exposer ici. Cependant, j'estimerai avoir peu de mérite à démontrer ce que personne n'ignore. Il reste autre chose que probablement personne jusqu'ici n'a jamais envisagé, même dans ses pensées. Il reste en effet à rechercher, quand le nombre d'observations augmente continûment, si **la probabilité d'atteindre le vrai rapport entre les nombres de cas où un événement peut se produire ou non, augmente de telle sorte qu'elle dépasse n'importe quel degré de certitude**.

17 Loi des Grands Nombres (1713)

Essayons de voir le résultat comme l'exprime Bernoulli. Pour lui, il y a des cas fertiles qui sont nos cas favorables, et des cas stériles. Comprenez qu'on a un événement de probabilité r/t , qui a été observé X fois sur n expériences.

Bernoulli veut calculer la probabilité que la fréquence X/n soit dans l'intervalle $(r-1)/t$, $(r+1)/t$. Pour une même valeur de la probabilité, si r et t sont grands, l'intervalle est étroit.

Voici le résultat. Bernoulli trouve deux nombres n_1 et n_2 , qui se calculent comme il est écrit ici. Si le nombre d'expériences les dépasse, alors la probabilité cherchée est plus grande que c sur $c+1$, c'est-à-dire proche de 1 si c est grand.

Ça paraît tordu, mais c'est bien le résultat souhaité.

18 Loi des Grands Nombres (1713)

Voici l'application numérique qu'il donne. Prenons $r = 30$, $s = 20$, donc une probabilité de trois cinquièmes, et $c = 1000$.

Le résultat de Bernoulli dit que si le nombre d'expériences dépasse 25550, alors la probabilité que la fréquence soit entre $29/50$ et $31/50$ dépasse 1000 sur 1001.

19 Loi des Grands Nombres (1713)

En réalité, pour $n = 25550$, la probabilité en question est beaucoup plus proche de 1 : regardez ce qu'en dit R. L'écart de cette probabilité à un est de l'ordre de 10^{-11} . Mais peu importe, c'est le premier résultat qui justifie mathématiquement l'intuition de la loi des grands nombres.

Bernoulli explique sa vision de la loi.

Loi des Grands Nombres (1713)

Jacob Bernoulli (1655-1705)

cas fertiles : r , cas stériles : s , total : $t = r + s$
expériences : n , cas fertiles observés : X . Soit :

$$P = \text{Prob} \left(\frac{r-1}{r+s} < \frac{X}{n} < \frac{r+1}{r+s} \right).$$

Soit $c \gg 1$. Posons

$$m_1 = \frac{\log(c(s-1))}{\log(r+1) - \log(r)}; \quad m_2 = \frac{\log(c(r-1))}{\log(s+1) - \log(s)};$$
$$n_1 = m_1 t + \frac{m_1 r t - r t}{s+1}; \quad n_2 = m_2 t + \frac{m_2 s t - s t}{r+1}.$$

Si $n > \max\{n_1, n_2\}$, alors :

$$\frac{P}{1-P} > c \iff P > \frac{c}{c+1}.$$

Loi des Grands Nombres (1713)

Jacob Bernoulli (1655-1705)

Exemple : $r = 30$, $s = 20$, $c = 1000$.

Si $n > 25550$, alors :

$$P = \text{Prob} \left(\frac{29}{50} < \frac{X}{n} < \frac{31}{50} \right) > \frac{1000}{1001}.$$

Loi des Grands Nombres (1713)

Jacob Bernoulli (1655-1705)

Exemple : $r = 30$, $s = 20$, $c = 1000$.

Si $n > 25550$, alors :

$$P = \text{Prob} \left(\frac{29}{50} < \frac{X}{n} < \frac{31}{50} \right) > \frac{1000}{1001}.$$

```
> n <- 25550
> pbinom(29/50*n,n,3/5)
[1] 3.948986e-11
> pbinom(31/50*n,n,3/5,lower.tail=FALSE)
[1] 2.862587e-11
```

20 tout est gouverné par des rapports précis

« De là se voit la conséquence singulière, que si les observations de tous les événements continuaient éternellement (la probabilité se transformant en parfaite certitude), tout dans le monde est gouverné par des rapports précis et une loi de changement constante, de sorte que même pour ce qui est au plus haut point contingent et fortuit, nous devrions admettre une certaine nécessité et j'oserais dire, une fatalité. »

Cette vision est bien dans l'air du temps. Un demi-siècle auparavant, Pascal avait qualifié de « stupéfiant » le fait de contruire une géométrie du hasard. Mathématiques et hasard étaient vus comme contradictoires, et la loi des grands nombres en quelque sorte réconciliait les deux.

tout est gouverné par des rapports précis

Jacob Bernoulli (1655–1705)

De là se voit la conséquence singulière, que si les observations de tous les événements continuaient éternellement (la probabilité se transformant en parfaite certitude), tout dans le monde est gouverné par des rapports précis et une loi de changement constante, de sorte que même pour ce qui est au plus haut point contingent et fortuit, nous devrions admettre une certaine nécessité et j'oserais dire, une fatalité.

21 Adolphe Quetelet (1796–1874)

Un siècle et demi plus tard, Adolphe Quetelet continuait à y croire. Quetelet est un statisticien belge, qui a fait beaucoup pour le développement de la discipline en Europe.

Il applique les statistiques non seulement à l'anthropométrie, mais aussi à la sociologie. Quand en 1831 il écrit ses « Recherches sur le penchant au crime », il définit un objet théorique nouveau, qu'il appelle « L'Homme moyen ».

Adolphe Quetelet (1796–1874)



22 Recherches sur le penchant au crime (1831)

« L'homme que j'ai considéré est dans la société l'analogue du centre de gravité dans les corps ; c'est un être fictif pour lequel toutes les choses se passent conformément aux résultats moyens obtenus pour la société. Si l'homme *moyen* était déterminé pour une nation, il présenterait le type de cette nation, s'il pouvait être déterminé d'après l'ensemble des hommes, il présenterait le type de l'espèce humaine toute entière. »

Recherches sur le penchant au crime (1831)

Adolphe Quetelet (1796–1874)

L'homme que j'ai considéré est dans la société l'analogue du centre de gravité dans les corps ; c'est un être fictif pour lequel toutes les choses se passent conformément aux résultats moyens obtenus pour la société. Si l'homme *moyen* était déterminé pour une nation, il présenterait le type de cette nation, s'il pouvait être déterminé d'après l'ensemble des hommes, il présenterait le type de l'espèce humaine toute entière.

23 Recherches sur le penchant au crime (1831)

« D'après tout ce qui vient d'être dit, je crois que non-seulement il n'est pas absurde, mais qu'il est même *possible* de déterminer l'homme moyen d'une nation ou de l'espèce humaine ; l'absurdité apparente d'une semblable recherche ne provient que du manque d'observations faites avec précision et en assez grand nombre pour que les résultats présentent le plus de probabilité possible qu'ils sont conformes à la vérité. »

Recherches sur le penchant au crime (1831)

Adolphe Quetelet (1796–1874)

D'après tout ce qui vient d'être dit, je crois que non-seulement il n'est pas absurde, mais qu'il est même *possible* de déterminer l'homme moyen d'une nation ou de l'espèce humaine ; l'absurdité apparente d'une semblable recherche ne provient que du manque d'observations faites avec précision et en assez grand nombre pour que les résultats présentent le plus de probabilité possible qu'ils sont conformes à la vérité.

24 Charles Babbage (1791–1871)

Charles Babbage était un ami de Quetelet. Oui, c'est bien le Babbage de la Machine Analytique. Il s'est aussi occupé de statistique. Et comme Quetelet, il croyait en la loi des grands nombres.

Babbage était allé jusqu'à planifier la fabrication d'une « table des constantes de la nature et des arts », où il mettait exactement sur le même plan des constantes de type astronomique, et des quantités beaucoup plus variables.

Charles Babbage (1791–1871)



25 Constants of Nature and Arts (1832)

« Elle doit contenir tous ces faits qui peuvent être exprimés par des nombres dans les sciences et les arts. On en donnera une meilleure idée en précisant le plan.

Premièrement, toutes les constantes de notre système, comme les distances de chaque planète, leur période de révolution, l'inclinaison de leur orbite, la proportion de lumière reçue du soleil, la force de la gravité à leur surface.

Deuxièmement : le poids atomique de chaque corps.

Troisièmement : la liste des gravités de chaque corps. »

Jusque là, ça va, ce sont bien des constantes, ou peu s'en faut.

Constants of Nature and Arts (1832)

Charles Babbage (1791–1871)

It ought to contain all those facts which can be expressed by numbers in the various sciences and arts. A better idea will be formed by giving an outline of its proposed contents.

1. All the constant quantities belonging to our system; – as distance of each planet, – period of revolution, – inclination of orbit, &c. – proportion of light received from sun, – force of gravity on surface of each.
2. The atomic weights of bodies.[...]
3. List of specific gravities of all bodies.

26 Constants of Nature and Arts (1832)

« Septièmement; de l'homme. Le poids moyen et la taille moyenne à différentes périodes de son existence, les tables de mortalité en différents endroits, les durées moyennes des règnes des souverains

Huitièmement : La puissance de l'homme et des animaux. Un homme travaillant dix heures par jour sciera tant de pieds carrés de tel bois. »

Et sur sa lancée, Babbage arrive aux langues.

Constants of Nature and Arts (1832)

Charles Babbage (1791–1871)

7. Of man. Average weight at various periods of existence, height of ditto, tables of mortality in various places, average durations of reigns of sovereigns; [...]
8. Power of man and animals. A man laboring ten hours per day will saw () square feet of [...]

27 Constants of Nature and Arts (1832)

« 18. Tables de fréquence des lettres de l'alphabet dans les différentes langues, fréquences de ces lettres en début et en fin de mot, le nombre des lettres redoublées dans les différentes langues, la proportion des lettres qui débutent les noms de famille des différentes nations. »

Constants of Nature and Arts (1832)

Charles Babbage (1791–1871)

18. Tables of the frequency of occurrence of the various letters of the alphabet in different languages; of the frequency of occurrence of the same letters at the beginnings or endings of words; of the numbers of double letters occurring in different languages; of the proportion of letters commencing surnames amongst different nations.

28 Augustus De Morgan (1806–1871)

Toujours en Angleterre, Augustus De Morgan était reconnu comme un des tout meilleurs spécialistes anglais de la théorie des probabilités, en tout cas un de ceux qui avaient compris la théorie de Laplace, et au dix-neuvième siècle, ils étaient peu nombreux.

En 1851 De Morgan écrit à un ami pour lui proposer une application.

Augustus De Morgan (1806–1871)



29 Letter to Rev. W. Heald Aug. 18 1851

« J'ai toujours pensé qu'une petite dépense d'argent suffirait à décider de questions de paternité d'œuvres littéraires, de cette façon.

Je m'attendrais à ce que les légères différences entre écrivains restent cohérentes d'un auteur à l'autre.

Si les spécialistes de littérature connaissaient la loi des moyennes aussi bien que les mathématiciens, il serait aisé de rassembler un peu d'argent pour tenter l'expérience à grande échelle. J'essaierais le grec, le latin, et l'anglais, et je m'attendrais à ce qu'un même homme écrivant sur deux sujets différents, soit plus en accord avec lui-même que deux hommes différents écrivant sur le même sujet. Un de ces jours des faussaires seront démasqués par ce test. »

L'occasion de tester la méthode de De Morgan n'a pas tardé. Vers la moitié du dix-neuvième siècle, une controverse a éclaté. En voici la trace dans une lettre de Carlyle à son frère.

Letter to Rev. W. Heald Aug. 18 1851

Augustus De Morgan (1806–1871)

It has always run in my head that a little expenditure of money would settle questions of authorship in this way.

[...]

I should expect the slight differences between one writer and another to be well maintained against each other, and very well agreeing with themselves.

[...]

If scholars knew the law of averages as well as mathematicians, it would be easy to raise a few hundred pounds to try this experiment on a grand scale. I would have Greek, Latin, and English tried, and I should expect to find that one man writing on two different subjects agrees more nearly with himself than two different men writing on the same subject. Some of these days, spurious writings will be detected by this test.

30 Carlyle to his brother (1853)

« Nous avons en ce moment un dame yankee envoyée par Emerson, qui a découvert que l'« homme Shakespeare » est un mythe, et qu'il n'a pas écrit les pièces qui portent son nom. Au contraire elles ont été écrites par un « associé secret » dont le nom est inconnu. En fait elle est venue en Angleterre dans le but d'examiner, et si possible de prouver cela, à partir du British Museum, et d'autres sources de preuves. »

Carlyle to his brother (1853)

Thomas Carlyle (1795–1881)

For the present, we have (occasionally) a yankee lady, sent by Emerson, who has discovered that the "Man Shakespeare" is a *Myth*, & did not write those Plays that bear his name, who were on the contrary written by a "secret associate" (names *unknown*) : she has actually come to England for the purposes of examining that, and if possible, proving it, from the British Museum and other sources of evidence. *Ach Gott!*

31 Delia Salter Bacon (1811–1859)

La Yankee Lady, la voici : Delia Salter Bacon. Elle est l'auteur d'un livre de sept cent pages, où il est démontré que Shakespeare n'a pas écrit les pièces de Shakespeare. Je sens que je vais vous décevoir, mais je n'ai pas eu le courage de tout lire, et je n'ai pas compris grand-chose au peu que j'ai lu. Voici un extrait de sa prose.

Delia Salter Bacon (1811–1859)



32 The philosophy of the plays of Shakspere unfolded (1857)

« À l'esprit qui est capable de recevoir dans n'importe quelle condition la vérité démontrée, et lui donner son poids entier, à l'esprit pour lequel la vérité est religion, ce livre est dédié. Les faits qu'il contient parlent d'eux-mêmes ou au moins parleront d'eux-mêmes par la suite. Ils ne dépendront pas finalement des modalités de leur exposition ici. Car ils ont la grande qualité, ils ont la solidité et les dimensions de la vérité historique, et ils sont accessibles par plus d'un côté. »

Les soi-disant preuves accumulées, visent à prouver que les œuvres de Shakespeare étaient en fait de Francis Bacon. Aucun rapport avec son propre nom de famille. Delia Salter Bacon n'était qu'une des premières d'une longue série, qui entre autres contient un mathématicien célèbre, Georg Cantor. En 1886, une société baconienne est fondée.

The philosophy of the plays of Shakspere unfolded (1857)

Delia Salter Bacon (1811–1859)

To the mind which is able to receive under all conditions the demonstrated truth, and give to it its full weight, – to the mind to which truth is religion, this book is dedicated. The facts which it contains are able to assert themselves, will be, at least, hereafter. They will not be dependent ultimately upon the mode of their exhibition here. For they have the large quality, they have the solidity and dimensions of historical truth, and are accessible on more sides than one.

33 The Baconian Society (founded 1886)

« L'affirmation des baconiens est que William Shakespeare n'a rien à voir du tout dans la production ni des pièces ni des poèmes, que c'était un homme peu éduqué, qui savait à peine écrire son propre nom; qu'il n'y a aucun début de preuve qu'il ait jamais écrit, ou qu'il ait pu écrire quoi que ce soit d'autre. »

Le romantisme avait fait naître en Angleterre une idolâtrie de Shakespeare sans commune mesure avec la relative indifférence des siècles précédents. Il n'était pas possible que ce soit un homme issu du peuple qui ait pu produire ces chefs-d'œuvre absolus, ces trésors nationaux de la fierté britannique. Il fallait nécessairement que ce soit un noble, et pourquoi pas Bacon. Il y avait d'ailleurs d'autres candidats.

The Baconian Society (founded 1886)

The contention of the Baconians is that William Shakspere had no hand whatsoever in the production of either the plays or the poems – that he was an uneducated man, who could just manage to write his own name; that there is not a particle of evidence that he ever wrote, or could write, anything else.

34 Thomas Corwin Mendenhall (1841–1924)

Thomas Corwin Mendenhall (1841–1924)



Mendenhall était un Américain, qui avait eu connaissance de la proposition de De Morgan, et qui avait décidé de la mettre en pratique.

35 The characteristic curves of composition (1887)

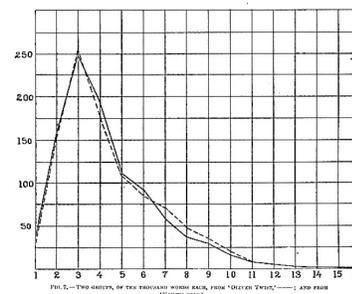
En 1887, il avait écrit un article sur les « courbes caractéristiques de la composition ». L'idée était pour un texte donné de relever la fréquence des longueurs de mots, et d'en faire un diagramme en barre, même si Mendenhall le présente sous forme de courbe.

Si De Morgan a raison, la courbe comme son nom l'indique, doit être caractéristique d'un auteur donné. Le graphique que vous avez sous les yeux est un peu inquiétant. Parce que les deux courbes, qui sont pratiquement confondues, proviennent l'une de *Oliver Twist*, de Charles Dickens, l'autre d'une série d'articles dans *Vanity Fair*.

Mais qu'à cela ne tienne, un riche américain décide de financer un projet pour trancher la controverse Bacon-Shakespeare. Le nouvel article de Mendenhall est paru en 1901. Il est intitulé « Une solution mécanique d'un problème littéraire ».

The characteristic curves of composition (1887)

Thomas Corwin Mendenhall (1841–1924)



36 A mechanical solution of a literary problem (1901)

« La méthode a été appliquée enfin à cette question des plus curieuses et pourtant des plus intéressantes, grâce à la générosité de M. Augustus Heminway de Boston, qui a gentiment offert de couvrir les dépenses du travail, qui consistait à employer des personnes pour compter et classer presque deux millions de mots. À part l'expression de ma reconnaissance à Mr Heminway, je tiens à remercier aussi les dames qui ont entrepris la lourde tâche du comptage Madame Eichard Mitchell et Mademoiselle Amy C. Whitman, de Worcester, dans le Massachussetts. »

Que c'était des dames qui avaient fait le comptage, vous l'auriez parié. Mais alors la conclusion ?

A mechanical solution of a literary problem (1901)

Thomas Corwin Mendenhall (1841–1924)

That the method has been applied at last to this most curious and yet most interesting question is entirely due to the liberality of Mr. Augustus Heminway, of Boston, who kindly offered to defray the expenses of the work, that is, to employ persons to count and classify nearly two millions of words. Besides expressing my indebtedness to Mr. Heminway, I wish to make grateful acknowledgment of the excellent and entirely satisfactory manner in which the heavy task of counting was performed by the ladies who undertook it, Mrs. Eichard Mitchell and Miss Amy C. Whitman, of Worcester, Massachusetts.

41 The statistical study of literary vocabulary (1944)

« Quand j'ai eu avancé quelque peu dans cette étude particulière, il m'est apparu trop clairement dans quel champ d'épines j'étais allé traîner. Il s'est avéré que les statistiques du domaine littéraire avaient leurs propres problèmes spéciaux, leurs difficultés particulières, et leurs sources d'erreurs, et que personne jusque-là n'avait fait la moindre tentative pour les explorer. »

The statistical study of literary vocabulary (1944)

George Udny Yule (1871–1951)

When I had advanced some way in that particular study, it became only too clear into how thorny a field of statistics I had strayed. [Statistics of literary vocabulary proved to have their own special problems](#), their own peculiar difficulties and sources of fallacy, which no one apparently had made any attempt systematically to explore.

42 références

Prouver que c'est pas moi qui ai écrit cette histoire, ça devrait être facile. Cent pour cent de mes histoires se terminent par une vanne, le plus souvent foireuse. Pas celle-ci. Alors vous voyez bien !

références

- J. Bernoulli (1713) *Ars Conjectandi*, Bâle : Thurnis
- I. Hacking (1990) *The taming of chance*, Cambridge University Press
- C. Laporte (2007) The Bard, the Bible, and the Victorian Shakespeare question, *English Literary History*, 74(3), 609–628
- S. M. Stigler (1999) *Statistics on the table. The history of statistical concepts and methods*, Cambridge MA : Harvard University Press
- G. U. Yule (1944) *The statistical study of literary vocabulary*, Cambridge University Press