

0 La proportion des voyelles

La proportion des voyelles dans une langue : ce petit détail statistique a accompagné l'histoire de la discipline depuis ses débuts, jusqu'au vingtième siècle. Cette même proportion des voyelles a aussi alimenté les préjugés sur les langues.

histoires de statistique

La proportion des voyelles

d'Alberti à Markov



hist-math.fr

Bernard YCART

1 Natya Shastra (±500)

Des préjugés sur les langues, il y en a toujours eu. Voici par exemple un extrait d'un traité indien sur les arts de la scène, où le Prakrit, qui est la langue vulgaire, est opposé au Sanskrit, qui est la langue noble.

« Le Prakrit devrait être réservé aux enfants, aux personnes de caractères les plus bas, et je vous laisse traduire les autres catégories. Aux Bouddhistes et aux gens instruits qui portent des costumes en accord avec leur position, le Sanskrit devrait être réservé. »

Voici quelqu'un qui est beaucoup plus proche de nous :...

Natya Shastra (±500)

Traité indien sur les arts de la scène

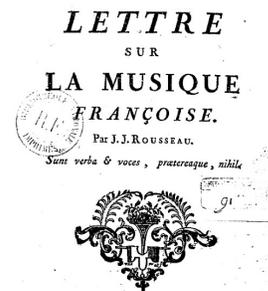
Prakrit should be assigned to children, persons possessed of spirits of lower order, women in feminine character persons of low birth, lunatics and phallus-worshippers. But to inherent recluses, sages, Buddhists, pure Śrotriyas and others who have received instruction [in the Vedas] and wear costumes suitable to their position (*vinigastha*) should be assigned Sanskrit Recitation.

2 Lettre sur la musique française (1753)

Jean-Jacques Rousseau. Dans un contexte polémique que je vous raconterai ailleurs, il a écrit en 1753 une « Lettre sur la musique française ». Voici ce qu'il dit.

Lettre sur la musique française (1753)

Jean-Jacques Rousseau (1712-1778)



3 les quatre qualités les plus convenables au chant

« Or s'il y a en Europe une langue propre à la musique, c'est certainement l'italienne ; car cette langue est douce, sonore, harmonieuse, et accentuée plus qu'aucune autre, et ces quatre qualités sont précisément les plus convenables au chant.

Elle est douce, parce que les articulations y sont peu composées, que la rencontre des consonnes y est rare et sans rudesse, et qu'un très grand nombre de syllabes n'y étant formées que de voyelles, les fréquentes élisions en rendent la prononciation plus coulante. »

La proportion de voyelles a toujours été considérée comme un critère : une langue qui a plus de voyelles est forcément plus musicale.

4 ce sera tant pis pour eux

« Je crois avoir fait voir qu'il n'y a ni mesure ni mélodie dans la musique française, parce que la langue n'en est pas susceptible ; que le chant français n'est qu'un aboiement continu, insupportable à toute oreille non prévenue ; que l'harmonie en est brute, sans expression et sentant uniquement son remplissage d'écolier ; que les airs français ne sont point des airs, que le récitatif français n'est point du récitatif. D'où je conclus que les Français n'ont point de musique et n'en peuvent avoir ; ou que si jamais ils en ont une, ce sera tant pis pour eux. »

Ça, c'est dit !

Après Rousseau, voici l'opinion de Voltaire, sur ce qu'il appelle « le génie d'une langue ».

5 Dictionnaire philosophique (1764)

« On appelle *génie d'une langue* son aptitude à dire de la manière la plus courte et la plus harmonieuse ce que les autres langages expriment moins heureusement.

Le latin, par exemple, est plus propre au style lapidaire que les langues modernes, à cause de leurs verbes auxiliaires qui allongent une expression et qui l'énervent.

Le grec, par son mélange mélodieux de voyelles et de consonnes, est plus favorable à la musique que l'allemand et le hollandais.

L'italien, par des voyelles beaucoup plus répétées, sert peut-être encore mieux la musique efféminée.

Le latin et le grec étant les seules langues qui aient une vraie quantité, sont plus faites pour la poésie que toutes les autres langues du monde. »

Voici deux citations attribuées à Charles Quint.

les quatre qualités les plus convenables au chant

Rousseau, Lettre sur la musique française (1753)

Or s'il y a en Europe une langue propre à la Musique, c'est certainement l'Italienne ; car **cette langue est douce, sonore, harmonieuse, & accentuée plus qu'aucune autre, & ces quatre qualités sont précisément les plus convenables au chant.**

Elle est douce, parce que les articulations y sont peu composées, que **la rencontre des consonnes y est rare** & sans rudesse, & qu'un très grand nombre de syllabes n'y étant formées que de voyelles, les fréquentes élisions en rendent la prononciation plus coulante.

ce sera tant pis pour eux

Rousseau, Lettre sur la musique française (1753)

Je crois avoir fait voir qu'il n'y a ni mesure ni mélodie dans la Musique Française, parce que la langue n'en est pas susceptible ; que le chant François n'est qu'un aboiement continu, insupportable à toute oreille non prévenue ; que l'harmonie en est brute, sans expression & sentant uniquement son remplissage d'Ecolier ; que les airs François ne sont point des airs, que le récitatif François n'est point du récitatif. D'où je conclus que **les François n'ont point de Musique & n'en peuvent avoir** ; ou que si jamais ils en ont une, ce sera tant pis pour eux.

Dictionnaire philosophique (1764)

Voltaire (1694-1778)

On appelle *génie d'une langue* son aptitude à dire de la manière la plus courte et la plus harmonieuse ce que les autres langages expriment moins heureusement.

Le Latin, par exemple, est plus propre au style lapidaire que les langues modernes, à cause de leurs verbes auxiliaires qui allongent une expression et qui l'énervent.

Le grec, par son mélange mélodieux de voyelles et de consonnes, est plus favorable à la musique que l'allemand et le hollandais.

L'italien, par des voyelles beaucoup plus répétées, sert peut-être encore mieux la musique efféminée.

Le latin et le grec étant les seules langues qui aient une vraie quantité, sont plus faites pour la poésie que toutes les autres langues du monde.

6 citations... ?

« Je parle espagnol à Dieu, italien aux femmes, français aux hommes et allemand à mon cheval. »

ou bien : « J'ai appris l'italien pour parler au pape ; l'espagnol pour parler à ma mère ; l'anglais pour parler à ma tante ; l'allemand pour parler à mes amis ; le français pour me parler à moi-même. »

La seconde citation est un peu plus réaliste que la première, mais on n'a pas de preuve qu'il ait prononcé aucune des deux.

citations... ?

Charles Quint (1500–1558)

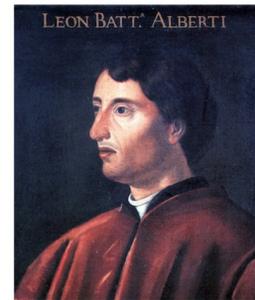
- Je parle espagnol à Dieu, italien aux femmes, français aux hommes et allemand à mon cheval.
- J'ai appris l'italien pour parler au pape ; l'espagnol pour parler à ma mère ; l'anglais pour parler à ma tante ; l'allemand pour parler à mes amis ; le français pour me parler à moi-même.

7 Leon Battista Alberti (1404–1472)

Dans un contexte biaisé par les préjugés, un humaniste de la Renaissance, Leon Battista Alberti a fait une remarque tout à fait insolite sur la proportion des voyelles en latin. Cette remarque anodine va à l'encontre des préjugés répandus. Mais d'abord qui est cet Alberti, et quel est son rapport avec les mathématiques ?

Il est connu comme théoricien de la perspective. Dans l'introduction de son livre sur la peinture, il prend bien soin de préciser l'apport des mathématiques.

Leon Battista Alberti (1404–1472)



8 de la peinture

« Voulant écrire sur la peinture en ces brefs commentaires, nous emprunterons aux mathématiciens, afin de rendre notre parler plus clair, tout ce qui paraîtra s'appliquer à notre sujet. Cela bien su, autant que notre intelligence y suffira, nous expliquerons cet art d'après les principes naturels. Toutefois, je souhaite vivement que, dans tout le cours de ce propos, on considère que je n'entends pas traiter la question en mathématicien, mais bien en peintre. »

de la peinture

Alberti (1404–1472)

Voulant écrire sur la peinture en ces brefs commentaires, nous emprunterons aux mathématiciens, afin de rendre notre parler plus clair, tout ce qui paraîtra s'appliquer à notre sujet. Cela bien su, autant que notre intelligence y suffira, nous expliquerons cet art d'après les principes naturels. Toutefois, je souhaite vivement que, dans tout le cours de ce propos, on considère que je n'entends pas traiter la question en mathématicien, mais bien en peintre.

9 l'architecture

Alberti est surtout connu comme architecte. Il existe encore en Italie des monuments bâtis sur ses plans. Il a écrit ce traité d'architecture, qui a été traduit en français. Là encore, il n'oublie pas ses connaissances mathématiques.

l'architecture

Alberti (1404–1472)



14 poètes et orateurs

« Ainsi en effet, il me semble avoir noté que chez les poètes les voyelles ne sont pas surpassées en nombre par les consonnes de plus d'un huitième ; mais chez les orateurs, les consonnes ne les surpassent pas, environ, plus que dans la proportion qu'on appelle sesquiterce. En effet, si les voyelles d'une page de cette sorte, dénombrées et rassemblées, sont par exemple au nombre de 300, le nombre total de toutes les consonnes restantes sera environ de 400. »

Il y a trois affirmations dans cette phrase.

15 poètes et orateurs

La première porte sur la proportion de voyelles chez les poètes. Appelons-la P . Ce que dit Alberti c'est que la différence entre $1 - P$, la proportion de consonnes, et P , la proportion de voyelles, est plus petite que un huitième. Donc P est plus grand que sept seizièmes.

16 poètes et orateurs

Dans la deuxième affirmation sesquiterce veut dire quatre tiers. Disons que la proportion de voyelles chez les orateurs est R . Le rapport de $1 - R$ à R est inférieur à quatre tiers. Donc R est supérieur à trois septièmes.

Regardez la différence entre sept seizièmes et trois septièmes : c'est moins de un pour cent. Comment Alberti a-t-il fait pour détecter une différence aussi faible ?

poètes et orateurs

Alberti, De cyfris (ca. 1467)

Ainsi en effet, il me semble avoir noté que chez les poètes les voyelles ne sont pas surpassées en nombre par les consonnes de plus d'un huitième ; mais chez les orateurs, les consonnes ne les surpassent pas, environ, plus que dans la proportion qu'on appelle sesquiterce. En effet, si les voyelles d'une page de cette sorte, dénombrées et rassemblées, sont par exemple au nombre de 300, le nombre total de toutes les consonnes restantes sera environ de 400.

poètes et orateurs

Alberti, De cyfris (ca. 1467)

- 1 Ainsi en effet, il me semble avoir noté que chez les poètes les voyelles ne sont pas surpassées en nombre par les consonnes de plus d'un huitième ;

$$(1 - P) - P < 1/8 \implies P > 7/16 \simeq 43.75\%$$

poètes et orateurs

Alberti, De cyfris (ca. 1467)

- 1 Ainsi en effet, il me semble avoir noté que chez les poètes les voyelles ne sont pas surpassées en nombre par les consonnes de plus d'un huitième ;

$$(1 - P) - P < 1/8 \implies P > 7/16 \simeq 43.75\%$$

- 2 mais chez les orateurs, les consonnes ne les surpassent pas, environ, plus que dans la proportion qu'on appelle sesquiterce.

$$(1 - R)/R < 4/3 \implies R > 3/7 \simeq 42.86\%$$

17 poètes et orateurs

La troisième affirmation donne une indication. Dire que si on compte 300 voyelles il y aura 400 consonnes est une redite de l'affirmation précédente. Or le De Cyfris d'Alberti est un tout petit texte, il n'y a pas un mot de trop. Donc si Alberti dit cela, à mon avis c'est une indication statistique : il a dû compter des pages de 700 lettres.

D'une part, Alberti avait remarqué que la proportion de voyelles pouvait dépendre du style, bien avant Rousseau et Voltaire, qui croyaient qu'elle était caractéristique de la langue. D'autre part, trois siècles avant les premiers tests statistiques, Alberti avait été capable de détecter une différence de fréquence en comptant des lettres.

Après tout, il disait peut-être n'importe quoi au hasard. Alors, j'ai voulu en avoir le cœur net. J'ai téléchargé 10 textes de poètes et 10 textes d'orateurs, et j'ai demandé à R de compter.

18 Dix textes de poètes

Il y a les métamorphoses d'Ovide, l'Enéide de Virgile, et tous les poètes latins célèbres que j'ai pu trouver. En tout deux millions six cents mille lettres.

19 Dix textes d'orateurs

Voici les textes d'orateurs, pour presque autant de lettres. Il y a Cicéron, Horace, Sénèque, Salluste. La guerre des Gaules de César n'est pas un discours, mais César était bien un orateur. J'ai mis aussi l'architecture de Vitruve, parce qu'Alberti le cite et dit qu'il s'en est inspiré.

poètes et orateurs

Alberti, De cyfris (ca. 1467)

- 1 Ainsi en effet, il me semble avoir noté que chez les poètes les voyelles ne sont pas surpassées en nombre par les consonnes de plus d'un huitième;

$$(1 - P) - P < 1/8 \implies P > 7/16 \approx 43.75\%$$

- 2 mais chez les orateurs, les consonnes ne les surpassent pas, environ, plus que dans la proportion qu'on appelle sesquiterce.

$$(1 - R)/R < 4/3 \implies R > 3/7 \approx 42.86\%$$

- 3 En effet, si les voyelles d'une page de cette sorte, dénombrées et rassemblées, sont par exemple au nombre de 300, le nombre total de toutes les consonnes restantes sera environ de 400.

Dix textes de poètes

410620 mots, 2617488 lettres

| | | mots | lettres |
|----|----------------------------|-------|---------|
| CA | Catullus, Poems | 13120 | 71747 |
| JS | Juvenal, Saturae | 24983 | 142645 |
| LN | Lucretius, De rerum Natura | 48508 | 274355 |
| ME | Martial, Epigrams | 53914 | 299099 |
| OM | Ovid, Metamorphoses | 77971 | 446848 |
| PE | Propertius, Elegiae | 24553 | 134719 |
| SP | Silius Italicus, Punica | 74035 | 449903 |
| ST | Statius, Thebaid | 42286 | 365326 |
| TE | Tibullus, Elegiae | 11283 | 66062 |
| VE | Virgil, Aeneid | 39967 | 366784 |

Dix textes d'orateurs

412182 mots, 2570344 lettres

| | | mots | lettres |
|----|--------------------------------------|-------|---------|
| AM | Apuleius, Metamorphoses | 36466 | 332617 |
| CG | Caesar, De Bello Gallico | 50661 | 317056 |
| CP | Cicero, Catilinarian and Philippics | 64799 | 370521 |
| HS | Horace, Sermones | 8893 | 77859 |
| LP | Lactantius, De Mortibus Persecutorum | 10956 | 65616 |
| PP | Pliny the younger, Panegyricus | 19205 | 112992 |
| QD | Quintillian, Declamatio Major | 66329 | 390261 |
| SC | Seneca the Elder, Controversiae | 88783 | 499280 |
| SO | Sallust, Orations | 8607 | 50676 |
| VA | Vitruvius, De Architectura | 57483 | 353466 |

20 Poètes et orateurs

Voici le résultat. En bleu, les textes de poètes, en rouge les textes d'orateurs. En abscisse, le pourcentage cumulé de A, de E, et de Y. En ordonnée, le pourcentage cumulé de I, de O, et de U. La somme abscisse plus ordonnée, X plus Y , est la proportion de voyelles. La droite rouge est $X + Y = 7/16$, la droite bleue $X + Y = 3/7$. Comme vous le voyez, la proportion de A,E,Y est supérieure chez les poètes, et les poètes sont majoritairement au-dessus de la droite bleue. Les orateurs sont majoritairement au-dessus de la droite rouge. C'est exactement ce que disait Alberti.

21 Poètes et orateurs

Quant aux pages de 700 lettres, j'ai tiré 10 000 paquets de 700 lettres consécutives dans chacun des textes, et j'ai représenté pour chaque texte la distribution des 10 000 proportions de voyelles. Voici les représentations de ces distributions pour les poètes, toujours en bleu, et les orateurs, en rouge. Les lignes horizontales correspondent aux prédictions d'Alberti.

Alberti avait donc raison, mais il avait raison beaucoup trop tôt : quatre bons siècles avant le début de la statistique linguistique. Alors personne n'a fait attention à ce qu'il avait dit.

22 William Owen Pughe (1759–1835)

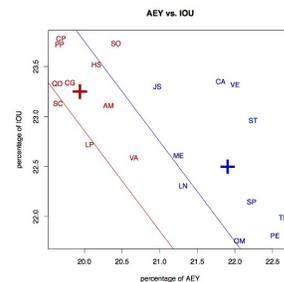
William Pughe, signait ses articles sous le nom de Meirion. C'était un défenseur acharné de la langue galloise, et il envoyait régulièrement aux journaux des articles sur le gallois. Il a écrit une grammaire, un dictionnaire gallois-anglais, et il a contribué à codifier l'orthographe du gallois.

23 Monthly Magazine (Jan 1, 1799)

Voici ce qu'il écrit au Monthly Magazine le premier janvier 1799. 1799, c'est le début de la statistique d'état. Le « Statistical Account of Scotland » de Sinclair, est paru entre 1791 et 1799 justement.

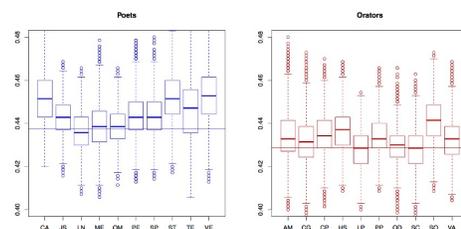
Poètes et orateurs

pourcentages par groupes de voyelles



Poètes et orateurs

10000 pages de 700 lettres



William Owen Pughe (1759–1835)



Monthly Magazine (Jan 1, 1799)

William Owen Pughe (1759–1835)

THE MONTHLY MAGAZINE.

No. XLII.] FOR JANUARY, 1799. [Vol. VII.

ORIGINAL COMMUNICATIONS.

To the Editor of the Monthly Magazine.

Should not trouble you with any further observations relative to Mr. Good, if the point in discussion between us, did not appear to me to be one in which the public are at this time particularly interested. Mr. Good asserts, that it is impossible to peruse the Shrew-

rated the cost of the poor at Norwich in provisions, at 2s. 6d. and at Shrewsbury at 3s. 13d. each, per week. It appears, however, that the poor at Norwich were actually supported the last year for 12s. 11d. each; and, I am enabled to add, are now supported in the Shrewsbury House, for 11. 9d. each, per week. Mr. Good may perhaps be inclined again to "de-

24 un préjugé vulgaire

« En regardant la nouvelle édition de l'Encyclopédie de Chambers, il y a quelque temps, je suis tombé par hasard sur une remarque à propos d'un sujet ayant trait au langage, selon laquelle le *gallois* et le *hollandais* étaient désignés comme abondant plus en consonnes que la plupart des langues européennes, sinon toutes. Je savais bien qu'une telle affirmation était proverbiale, un préjugé vulgaire : mais cela m'a mis un peu en colère de me rendre compte qu'il avait trouvé place dans l'un des premiers dictionnaires de ce temps ; et, incapable d'effacer l'impression de mon esprit, j'eus l'idée de trouver une certitude arithmétique tolérablement exacte, quant à la fausseté d'une telle observation. »

un préjugé vulgaire

Pughe, Monthly Magazine (1799)

En regardant la nouvelle édition de l'Encyclopédie de Chambers, il y a quelque temps, je suis tombé par hasard sur une remarque à propos d'un sujet ayant trait au langage, selon laquelle le *gallois* et le *hollandais* étaient désignés comme abondant plus en consonnes que la plupart des langues européennes, sinon toutes. Je savais bien qu'une telle affirmation était proverbiale, un préjugé vulgaire : mais cela m'a mis un peu en colère de me rendre compte qu'il avait trouvé place dans l'un des premiers dictionnaires de ce temps ; et, incapable d'effacer l'impression de mon esprit, j'eus l'idée de trouver une certitude arithmétique tolérablement exacte, quant à la fausseté d'une telle observation.

25 montrer les résultats dans une table

« La méthode, adoptée comme la plus appropriée, fut d'établir le nombre moyen de voyelles pour cent consonnes, dans différentes langues ; et de montrer les résultats dans une table. Comme la conclusion que l'on en tire tend à établir un fait, sinon d'importance, au moins de quelque curiosité, vous pourriez être amené, Monsieur, à l'insérer dans votre estimable recueil. »

montrer les résultats dans une table

Pughe, Monthly Magazine (1799)

La méthode, adoptée comme la plus appropriée, fut d'établir le nombre moyen de voyelles pour cent consonnes, dans différentes langues ; et de montrer les résultats dans une table. Comme la conclusion que l'on en tire tend à établir un fait, sinon d'importance, au moins de quelque curiosité, vous pourriez être amené, Monsieur, à l'insérer dans votre estimable recueil.

26 Nombre de voyelles pour 100 consonnes

Voici cette table. Comme vous vous en doutiez, le gallois arrive en tête avec autant de voyelles que de consonnes, et l'anglais bon dernier, ce qu'il fallait démontrer. Ce que Pughe ne précise pas, c'est que le gallois écrit utilise beaucoup de voyelles redoublées, comme il l'avait lui-même décidé.

Ce qui est plus intéressant, c'est que Pughe se montre conscient de la variabilité d'échantillon, même s'il ne va pas jusqu'à quantifier ce qu'il appelle une « certitude arithmétique tolérablement exacte. »

Nombre de voyelles pour 100 consonnes

Pughe, Monthly Magazine (1799)

| | | |
|------------|-----|-----|
| Gallois | 100 | 100 |
| Grec | 95 | 100 |
| Espagnol | 92 | 100 |
| Italien | 88 | 100 |
| Latin | 86 | 100 |
| Français | 78 | 100 |
| Allemand | 74 | 100 |
| Hollandais | 66 | 100 |
| Anglais | 61 | 100 |

27 l'anglais est très variable

« L'anglais est très variable en ce qui concerne la proportion de voyelles et de consonnes : celle des consonnes est beaucoup plus élevée dans le style des Écritures que dans le langage élégant, plus particulièrement le langage scientifique, qui contient plus de mots dérivés des langues savantes.

[...]

La variabilité en grec est considérable. J'ai eu trouvé 150 voyelles pour 100 consonnes ; et fréquemment aussi peu que 86. Les autres langues sont proches du nombre moyen, donné dans la table : le gallois dévie de la moyenne par rarement plus de 3 voyelles. »

Donc Pughe se montre conscient des différences stylistiques, mais ça ne l'empêche pas d'affirmer n'importe quoi.

l'anglais est très variable

Pughe, Monthly Magazine (1799)

L'anglais est très variable en ce qui concerne la proportion de voyelles et de consonnes : celle des consonnes est beaucoup plus élevée dans le style des Écritures que dans le langage élégant, plus particulièrement le langage scientifique, qui contient plus de mots dérivés des langues savantes.

[...]

La variabilité en grec est considérable. J'ai eu trouvé 150 voyelles pour 100 consonnes ; et fréquemment aussi peu que 86. Les autres langues sont proches du nombre moyen, donné dans la table : le gallois dévie de la moyenne par rarement plus de 3 voyelles.

28 Frantz (Francis) Lieber (1798–1872)

Frantz Lieber est plus sérieux. C'est un Allemand, formé à Iéna, qui a émigré aux États-Unis. Il est considéré comme le fondateur de la linguistique comparative.

Frantz (Francis) Lieber (1798–1872)



29 Encyclopaedia Americana (1830)

Il est co-auteur de cette Encyclopédie Américaine, parue en 1830, qui contient, comme vous le voyez, une copieuse collection d'articles originaux. Pour cette encyclopédie, il rédige l'article Consonnes, où il reprend le même vieux préjugé sur la musicalité des langues, avec un peu plus de prudence tout de même.

Encyclopaedia Americana (1830)

Francis Lieber (1798–1872)

ENCYCLOPÆDIA AMERICANA.
A
POPULAR DICTIONARY
OF
ARTS, SCIENCES, LITERATURE, HISTORY, POLITICS AND
BIOGRAPHY,
BROUGHT DOWN TO THE PRESENT TIME;
INCLUDING
A COPIOUS COLLECTION OF ORIGINAL ARTICLES

30 la musique d'une langue

« Le son mélodieux ou la musique d'une langue dépend, en partie, de la proportion des voyelles aux consonnes, une langue devenant trop dure s'il y a trop de consonnes. Nous ne disons pas que l'euphonie d'une langue dépend entièrement de cette proportion, et qu'elle devient d'autant plus mélodieuse si la proportion de voyelles est plus grande. Ici, comme ailleurs, la plus grande partie de l'effet dépend de la répartition des éléments. La disposition même des voyelles par rapport aux consonnes, le mélange heureux des voyelles longues et courtes, accentuées et non accentuées, produisent la douce harmonie d'une langue. »

la musique d'une langue

Lieber, Encyclopedia Americana (1830) article Consonnes

Le son mélodieux ou la musique d'une langue dépend, en partie, de la proportion des voyelles aux consonnes, une langue devenant trop dure s'il y a trop de consonnes. Nous ne disons pas que l'euphonie d'une langue dépend entièrement de cette proportion, et qu'elle devient d'autant plus mélodieuse si la proportion de voyelles est plus grande. Ici, comme ailleurs, la plus grande partie de l'effet dépend de la répartition des éléments. La disposition même des voyelles par rapport aux consonnes, le mélange heureux des voyelles longues et courtes, accentuées et non accentuées, produisent la douce harmonie d'une langue.

31 proportions de consonnes et voyelles

Néanmoins, il propose une table donnant la proportion des voyelles et des consonnes. En fin linguiste, il sait qu'il faut distinguer la proportion orthographique de la proportion phonétique. Il le fait pour le français, l'anglais et l'allemand.

Il en déduit un véritable programme de statistique linguistique, avant l'heure.

proportions de consonnes et voyelles

Lieber, Encyclopedia Americana (1830) article Consonnes

| | Cons. | Vowels. |
|-------------------------|-------|---------------------------|
| Sandwich islands . . . | 1 | : 1.8 |
| Greek { Ionic dial. . . | 1 | : 1.333 |
| { Attic dial. . . | 1 | : 1.006 |
| Portuguese | 1.02 | : 1 |
| Common Arabic . . . | 1.06 | : 1* |
| Italian | 1.1 | : 1 |
| Seneca Indians . . . | 1.18 | : 1 |
| Chahia Indians . . . | 1.2 | : 1 |
| Sanscrit | 1.2 | : 1* |
| Latin | 1.2 | : 1 } † |
| Hebrew | 1.2 | : 1* |
| Spanish | 1.34 | : 1 |
| Persian | 1.33 | : 1* |
| Malay | 1.33 | : 1* † |
| French, phonic prop. | 1.34 | : 1 orthographic 1.27 : 1 |
| Dutch | 1.5 | : 1 † |
| English, phonic prop. | 1.51 | : 1 orthographic 1.52 : 1 |
| Swedish | 1.64 | : 1 |
| German, phonic prop. | 1.7 | : 1 orthographic 1.54 : 1 |

32 Comparer les proportions de consonnes et voyelles

« Comparer les proportions de consonnes et voyelles, dans des familles différentes de langues ; montrer les proportions de gutturales, labiales, etc. . . , des différents idiomes, [. . .] pourrait mener à des conclusions intéressantes. Cette branche de la philologie pourrait être comparée au nouveau domaine de la *stochiométrie* en chimie, qui traite de la proportion des quantités d'éléments en état de neutralisation ou solution – une branche de la science qui devient chaque jour de plus en plus importante. . . »

La comparaison avec la chimie est intéressante de la part de quelqu'un qui vient de Iéna, et qui a connu dans sa jeunesse le développement de la chimie allemande.

Comparer les proportions de consonnes et voyelles

Lieber, *Encyclopedia Americana* (1830) article *Consonnes*

Comparer les proportions de consonnes et voyelles, dans des familles différentes de langues ; montrer les proportions de gutturales, labiales, etc. . . , des différents idiomes, [. . .] pourrait mener à des conclusions intéressantes. Cette branche de la philologie pourrait être comparée au nouveau domaine de la *stochiométrie* en chimie, qui traite de la proportion des quantités d'éléments en état de neutralisation ou solution – une branche de la science qui devient chaque jour de plus en plus importante. . .

33 Andrei Andreievich Markov (1856–1922)

Un pour qui la leçon n'a pas été perdue, c'est Markov. En 1913, le pouvoir tsariste en Russie fête le tricentenaire de la dynastie Romanov. Un peu par défi, Markov décide d'organiser un congrès pour fêter le bicentenaire de la parution de l'*Ars Conjectandi* de Bernoulli. C'est le livre fondateur des probabilités, dans lequel on trouve la première loi des grands nombres.

L'idée de Markov depuis quelque temps est que l'hypothèse d'indépendance des épreuves n'est pas une condition nécessaire pour la loi des grands nombres. C'est pour cela qu'il introduit ce qu'il appelle des événements en chaîne, ce que nous appelons les chaînes de Markov. À l'occasion de ce congrès de 1913, il publie un exemple d'application.

Andrei Andreievich Markov (1856–1922)



34 Un exemple d'étude statistique. . . (1913)

L'article s'intitule « Un exemple d'étude statistique sur le texte d'*Eugène Onéguine* illustrant la liaison des épreuves en chaîne. »

« Cette étude examine un extrait de texte contenant 20 000 lettres de l'alphabet Russe [. . .] du roman de Pouchkine Eugène Onéguine – le premier chapitre entier et seize paragraphes du second.

Cette suite nous fournit 20 000 épreuves connectées, qui sont soit une voyelle, soit une consonne. »

Markov divise ses 20 000 lettres en 200 paquets de 100 lettres consécutives, et écrit ces 100 lettres en 10 lignes par 10 colonnes. Puis il compte le nombre de voyelles dans chaque colonne, et regroupe les colonnes distantes de 5 : la première et la cinquième, la seconde et la sixième, etc.

Un exemple d'étude statistique. . . (1913)

Andrei Markov (1856–1922)

sur le texte d'*Eugène Onéguine* illustrant la liaison des épreuves en chaîne

Cette étude examine un extrait de texte contenant 20 000 lettres de l'alphabet Russe [. . .] du roman de Pouchkine Eugène Onéguine – le premier chapitre entier et seize paragraphes du second.

Cette suite nous fournit 20 000 épreuves connectées, qui sont soit une voyelle, soit une consonne.

35 Un exemple d'étude statistique... (1913)

Il regroupe ces résultats par 5, et ça donne quarante tables qui résument chacune 500 lettres. Il compte aussi les fréquences des couples consécutifs. La probabilité d'une voyelle est de 0.66 quand la lettre précédente est une consonne, elle est de 0.13 quand la lettre précédente est une voyelle. Il y a donc bien une dépendance.

La proportion des voyelles est le seul exemple que Markov ait jamais donné de ses événements en chaîne. Il recommencera sur 100 000 lettres d'un roman d'Aksakov pour la quatrième édition de son livre d'enseignement sur la théorie des probabilités.

36 références

Et vous ne demandez pas pourquoi les proportions de voyelles ne sont pas les mêmes chez les poètes et chez les orateurs? Vous l'avez déjà demandé? Ah bon, pardon, j'avais pas entendu.

Voici ce que je pense, mais je ne l'ai pas démontré. La poésie latine était écrite pour être sinon chantée, au moins scandée. Demandez à qui vous voulez, pour chanter il faut ouvrir la bouche. Donc il se peut qu'un texte écrit pour être chanté ait plus de voyelles dites ouvertes, comme A et È que de voyelles fermées comme I. C'est bien ce qu'on observe dans les textes latins. Je serais prêt à parier que c'est vrai aussi en français et dans n'importe quelle langue. Pas vous?

Un exemple d'étude statistique... (1913)

Andrei Markov (1856-1922)

```
6 8 11 11 13 49 16 11 9 8 7 51 14 12 7 8 6 42 5 11 10 6 10 42
12 11 7 7 5 42 4 8 9 11 10 42 5 5 11 9 11 41 12 8 8 11 7 46
6 6 6 7 13 38 9 9 9 7 10 44 8 10 6 10 7 41 7 7 12 10 9 45
8 10 11 9 4 42 12 9 6 10 7 44 11 11 8 3 10 43 8 12 7 9 9 45
10 11 6 10 8 44 3 8 10 8 8 38 4 4 11 14 8 41 12 8 10 9 8 47
42 46 40 44 43 15 44 45 43 44 43 19 42 42 43 39 42 8 44 46 47 45 43 25

8 7 8 7 10 40 11 11 8 7 7 44 11 10 10 12 6 49 12 9 8 10 10 49
10 9 9 8 8 44 9 6 10 11 11 47 4 4 9 7 9 33 3 10 12 9 10 44
8 9 8 8 8 41 12 9 9 5 6 41 11 13 6 9 10 49 11 11 6 11 10 49
10 6 13 6 12 47 10 8 6 11 11 46 6 7 11 8 6 35 10 8 11 6 7 42
8 12 5 13 8 44 7 6 8 9 8 38 8 6 10 7 12 43 6 8 7 9 6 36
44 43 43 42 44 16 49 40 41 43 43 10 40 40 46 43 43 12 42 46 44 45 43 20

7 7 7 7 9 37 12 7 7 6 8 40 7 4 11 5 7 34 5 5 7 5 9 31
9 13 6 8 4 40 6 8 7 10 8 39 11 14 9 11 9 54 12 6 10 10 8 46
9 7 11 12 14 53 9 10 10 8 7 44 7 6 9 8 9 39 8 14 11 11 10 54
7 11 8 9 7 42 9 5 6 7 7 34 10 9 8 10 5 42 4 8 9 5 9 30
8 10 10 11 9 48 7 11 9 13 7 47 11 10 8 9 11 49 18 14 9 11 7 54
40 48 42 47 43 20 43 41 39 44 37 4 46 43 45 43 41 18 42 42 46 42 43 16
```

références

- P. Grzybek (2013) Historical remarks on the Consonant-Vowel proportion - From cryptanalysis to linguistic typology. The concept of phonological stoichiometry (Francis Lieber, 1800-1872), *Glottometrics*, 26, 96-103
- D. Link (2006) Traces of the mouth : Andrei Andreyevich Markov's mathematization of writing, *History of Science*, 44(145) 321-348
- M. Petruszewycz (1979) A. A. Markov, ses probabilités en chaîne et les statistiques linguistiques, *Mathématiques et Sciences Humaines*, 66, 5-42
- B. Y. (2014) Alberti's letter counts, *Lit. Linguist. Computing* 29(2), 255-265
- B. Y. (2013) Letter counting : a stem cell for Cryptology, Quantitative Linguistics, and Statistics, *Historia Linguistica* 40(3) 303-329