

JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

J.-B. ESTOUP

Les mots usuels

Journal de la société statistique de Paris, tome 58 (1917), p. 137-140

http://www.numdam.org/item?id=JSFS_1917__58__137_0

© Société de statistique de Paris, 1917, tous droits réservés.

L'accès aux archives de la revue « Journal de la société statistique de Paris » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/legal.php>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

V

VARIÉTÉ

LES MOTS USUELS (1),

LEUR NOMBRE ET LEUR FRÉQUENCE

Le nombre des mots d'une langue comme la nôtre est certainement très élevé. Il est difficile de donner un chiffre, même approximatif. Il y a des dictionnaires de 80.000, de 100.000 mots et plus. Mais il s'en faut que tous soient d'un usage courant. Max Muller évalue à 3.000 ou 4.000 le nombre des mots employés habituellement par un Anglais cultivé.

Il peut être intéressant à divers points de vue, pour les sténographes, d'avoir des précisions à cet égard, de connaître non seulement combien de mots sont d'un usage courant, forment la trame ordinaire du discours, mais encore quels sont ces mots et quel est leur ordre de fréquence.

J'ai fait, dans ce but, une statistique qui a porté sur un texte d'une longueur totale de 30.000 mots. Afin que l'expérimentation fût plus probante, j'ai eu soin d'opérer non sur un texte suivi d'un même auteur traitant un même sujet, mais sur des extraits de discours et d'articles (75 extraits) portant sur les sujets les plus variés.

J'ai d'abord découpé le texte entier en tranches de 1.000 mots. J'ai recherché, en considérant successivement chaque tranche :

1^o Combien chaque tranche contient de mots différents et ne figurant pas dans les tranches précédentes;

2^o Combien de fois revient chaque mot répété.

En procédant ainsi, j'ai compté, dans la première tranche de 1.000, 336 mots différents.

Dans la deuxième, j'ai compté 233 mots différents et ne figurant pas dans la première.

J'ai opéré de même et successivement pour les tranches suivantes, comptant toujours combien y sont contenus de mots différents et nouveaux, c'est-à-dire ne figurant pas dans les précédentes.

Voici le résultat de ces décomptes (Voir tableau, page 138).

Ce tableau montre clairement que le pourcentage des mots différents et nouveaux trouvé dans chaque tranche décroît très rapidement au fur et à mesure que l'on avance dans le décompte. Dans le 2^e mille, la proportion des mots nouveaux est de 23 %; dans le 10^e mille, elle est d'environ 9 %; dans le 20^e de 6 %; dans le 30^e de 4 %.

Il n'est pas téméraire d'affirmer que, si l'on poursuivait le décompte jusqu'à un 50^e ou à un 60^e mille, la proportion descendrait à 1 % ou peut-être à moins, c'est-à-dire qu'à ce point on aurait épuisé à peu près le vocabulaire entier qui constitue la trame habituelle de tout discours.

Pour former le texte entier d'une longueur de 30.000 mots, 2.780 vocables différents ont suffi.

(1) Nous devons cette étude à M. J.-B. Estoup, sténographe de la Chambre des Députés, qui a bien voulu nous autoriser à reproduire cet intéressant travail extrait de son ouvrage : *Gammes sténographiques* - Méthode pour l'acquisition de la vitesse applicable à tous les systèmes de sténographie (chez l'auteur, 3 francs).

Tranches	Nombre de mots différents et nouveaux	Pour 100	Moyenne de tranches
1 ^{re}	336	33,6	
2 ^e	233	23,3	
3 ^e	201	20,1	
4 ^e	152	15,2	
5 ^e	134	13,4	
6 ^e	121	12,1	
7 ^e	105	10,5	
8 ^e	92	9,2	} 9,80 %
9 ^e	87	8,7	
10 ^e	88	8,8	
11 ^e	89	8,9	
12 ^e	87	8,7	} 8,10
13 ^e	66	6,6	
14 ^e	82	8,2	
15 ^e	66	6,6	
16 ^e	50	5,0	} 7,20
17 ^e	105	10,5	
18 ^e	67	6,7	} 5,90
19 ^e	53	5,3	
20 ^e	59	5,9	
21 ^e	81	8,1	
22 ^e	43	4,3	} 5,20
23 ^e	44	4,4	
24 ^e	56	5,6	
25 ^e	39	3,9	
26 ^e	69	6,9	} 4,32
27 ^e	51	5,1	
28 ^e	30	3,0	
29 ^e	56	5,6	
30 ^e	36	3,6	
Total	2.778		

Si le décompte était poursuivi jusqu'au 60^e mille, nous pouvons admettre que l'augmentation par tranche de 10.000 mots serait la suivante :

Tranche du 31 ^e au 40 ^e mille.	3,00 %	soit 300 mots.
— 41 ^e au 50 ^e mille.	2,00	— 200 —
— 51 ^e au 60 ^e mille.	1,50	— 150 —
Total		650 mots.

Ainsi, ces 30 nouvelles tranches de mille ne doivent pas compter plus de 650 mots différents et nouveaux. Additionnés aux 2.780 précédents, nous obtenons un chiffre de 3.430, disons en nombre rond 3.500 mots différents, pour un texte d'une longueur totale de 60.000 mots.

Examinons maintenant l'ordre de fréquence de ces mots.

Ils se rangent ainsi (Le décompte ne porte plus ici que sur 20.000 mots) :

	Nombre des répétitions		Nombre des répétitions
Le, la les	1.949	dans	181
de, du, des	1.712	pas.	179
que, qui.	766	nous	167
à, au.	743	plus	162
et, est, ai.	741	pour	157
un, une.	413	par.	155
ce, se.	393	on	155
il.	282	son, sont	141
ne	279	je.	137
en	258	faire (et temps)	136
ces, ses, cette	232	tout, toute.	119
être (et temps).	200	avoir (et temps).	118

	Nombre des répétitions		Nombre des répétitions
nos	105	vouloir	37
vous	118	droit	36
ou, où	100	autre	36
leur	94	produit	36
si	93	environ	35
elle	92	falloir (et temps)	35
même	84	donner (et temps)	34
sur	82	voir (et temps)	33
mais	82	savoir (et temps)	33
y	70	nation	32
pouvoir (et temps)	66	développer	31
celui, celle, ceux	62	économique	31
dont, donc	62	messieurs	30
devoir (et temps)	60	sa	30
travailler (et temps)	60	France	29
peut, peu	58	quelque	29
dire	57	trouver	29
comme	56	intérêt	29
bien	53	aller (et temps)	28
industrie	50	lui	27
français	48	nombre	27
certain	44	point	26
grand	44	public	26
avec	41	année	25
sans	40	moins	25
venir	39	demander (et temps)	25
quel	39	tant, temps	24
commerce	38	chose	23
nouveau	37	État	23
tres	37	me	23

Sont répétés 22 fois :
depuis, deux, enseigner, moyen, ni, toujours.

Sont répétés 21 fois :
considérable, entre, petit, premier.

Sont répétés 20 fois :
bon, consommer, général, homme, non, obliger, possible, production.

Sont répétés 19 fois :
actuel, fabriquer, jour, jusque, loi, mieux, question.

Sont répétés 18 fois :
cela, connaître, industriel, qualité, richesse, seul.

Sont répétés 17 fois :
arriver, besoin, croire, effort, école, esprit, importation, mesure, parler, vin, vos.

Sont répétés 16 fois :
aujourd'hui, cause, chaque, effet, mon, ouvrier, parce que, politique, surtout, vie.

Sont répétés 15 fois :
ainsi, alors, budget, chambre, devenir, élever, goût, jeunesse, marché, ministre, personne, porter, prix, prime, société.

Sont répétés 14 fois :
car, commission, emploi, lorsque, orient, partie, passer, permettre, prendre, progrès, situation, spécial, vraiment.

Sont répétés 13 fois :
cas, différent, étranger, exister, former, important, mettre, nécessaire, perdre, rendre, concurrence, eux, facile, heure, lieu, monnaie, particulier, tel.

Sont répétés 12 fois :
aucun, aussi, compter, compagnie, condition, diminuer, énergie, époque, étude, exemple, force, fournir, million, penser, préparer, priver, prospérité, quand, résultat, rester, souvent, tenir, technique, usine.

Sont répétés 11 fois :

agriculture, après, appel, apporter, atteindre, augmenter, beaucoup, colonie, crise, dernier, fortement, Gouvernement, jamais, largeur, longtemps, manquer, marchand, moment, monde, nécessité, place, pratique, présent, principal, représenter. République, servir, simple, sous, sorte, trop.

Sont répétés 10 fois :

d'ailleurs, alimenter, assurer, charger, chez, client, culture, démontrer, défendre, direct, douane, enfin, essayer, façon, grave, idée, immédiat, matériel, moral, repos, œuvre, presque, profession, rapport, service, suffire, suivre, utile, ville, vue.

Sont répétés 9 fois :

acheter, article, actif, autant, avantage, beau, chiffre, comprendre, continuer, côté, démocrate, devant, dépense, dix, entreprise, exact, fer, finance, guère, haut, loin, méthode, multiplier, raison, rien, siècle, somme, suivant, vers, voyage.

Sont répétés 8 fois :

bénéfice, but, commencer, constater, contre, cours, créer, disparaître, égal, entrer, entendre, épuiser, exporter, extrême, famille, finir, frapper, groupe, grâce, huile, ici, indigène, intelligence, marché, or, partout, Paris, parmi, peuple, protéger, puisque, rapide, rechercher, reconnaître, revenu, révolution, suite, toucher, trois, valeur.

Sont répétés 7 fois :

action, adresse, admettre, affaire, agir, Allemagne, Allemand, Amérique, apparaître, apprenti, assez, atelier, avenir, cependant, chemin, civiliser, colon, combien, constituer, déjà, dessin, divers, doute, échange, entier, établir, extérieur, fois, fromage, garder, houille, impôt, instruction, instituer, justice, liberté, long, lutte, mal, manger, maison, naturel, opération, organiser, patron, pendant, pièce, point de vue, près, preuve, procéder, producteur, profond, propre, ressource, réserver, retard, rival, rôle, social, supposer, véritable, volonté, wagon.

Puis viennent :

62 mots répétés 6 fois (*);

82 mots répétés 5 fois;

131 mots répétés 4 fois;

194 mots répétés 3 fois;

329 mots répétés 2 fois;

Enfin 922 ne se présentent qu'une fois.

(*) La liste n'a pas d'intérêt.

J.-B. ESTOUP.