Jean-Baptiste Estoup et les prémices de la loi de Zipf : un sténographe à l'esprit scientifique - 1868-1950.

Alain Lelu

Alain Lelu
Professeur, Université de Franche-Comté
Associé à l'équipe KIWI, LORIA (Nancy)
alain.lelu@univ-fcomte.fr

Séminaire d'histoire du calcul des probabilités et de la statistique
EHESS - Séance du 7 décembre 2007

Jean-Baptiste Estoup et les prémices de la loi de Zipf : un sténographe à l'esprit scientifique - 1868-1950.

- Avec la collaboration active de
  - Jacques et Geneviève Estoup
  - Jean-Paul Lelu, Bruno Delprat, Denise Delprat
- sans oublier les travaux de Mme M. Petruszewycz [1]

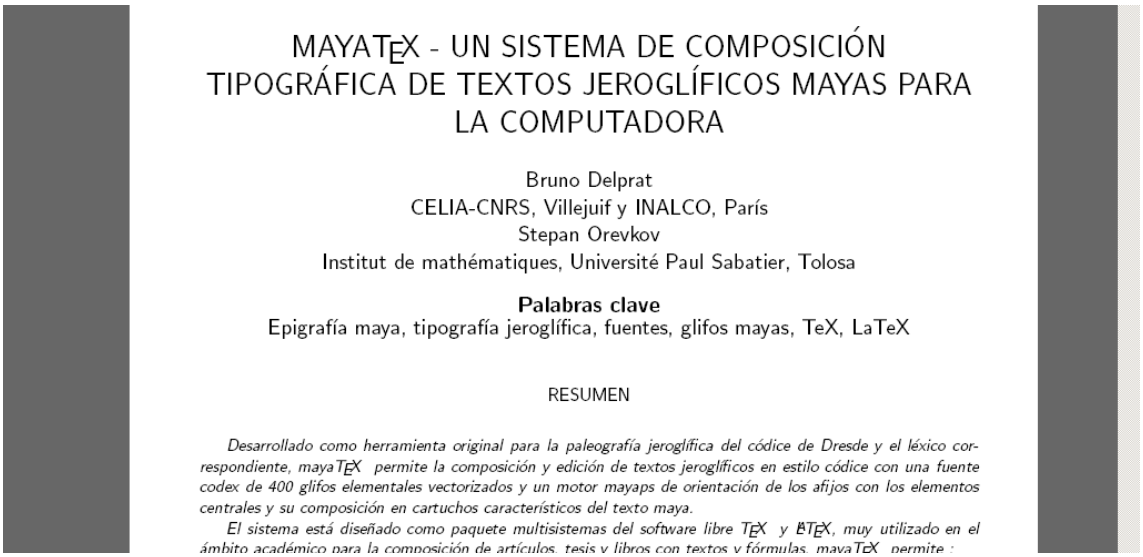
Une touche personnelle en guise d'introduction

Jean-Baptiste Estoup est aussi mon grand-père maternel[1].

L'intérêt pour la langue et les langues est répandu dans sa descendance :

Ma sœur Denise : carrière de professeur d'espagnol

Mon neveu Bruno s'est passionné pour l'étude du chinois, puis de l'écriture Maya



MAYATEX - UN SISTEMA DE COMPOSICIÓN TIPOGRÁFICA DE TEXTOS JEROGLÍFICOS MAYAS PARA LA COMPUTADORA

Bruno Delprat
CELIA-CNRS, Villejuif y INALCO, París
Stepan Orevkov
Institut de mathématiques, Université Paul Sabatier, Tolosa

Palabras clave
Epigrafía maya, tipografía jeroglífica, fuentes, glifos mayas, TeX, LaTeX

RESUMEN

Desarrollado como herramienta original para la paleografía jeroglífica del códice de Dresde y el léxico correspondiente, mayaTEX permite la composición y edición de textos jeroglíficos en estilo códice con una fuente codex de 400 glifos elementales vectorizados y un motor mayaps de orientación de los afijos con los elementos centrales y su composición en cartuchos característicos del texto maya.

El sistema está diseñado como paquete multisistemas del software libre TEX y LATEX, muy utilizado en el ámbito académico para la composición de artículos, tesis y libros con textos y fórmulas. mayaTEX permite :

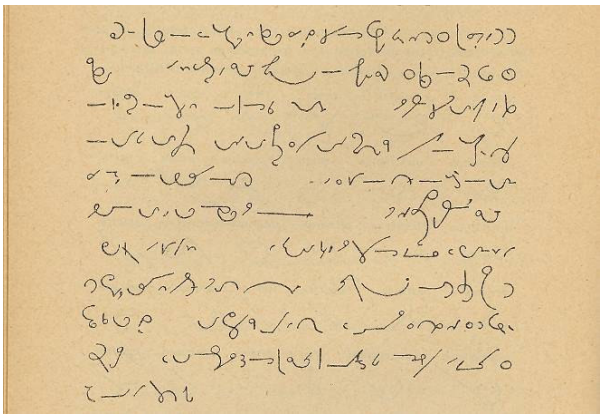
Mon frère Jean-Paul, féru de toponymie, entre autres.

Des tantes et un oncle sténographes. Cet oncle a aussi mis au point le premier prototype de télex français vers 1939, commercialisé par SAGEM en 1948.

Et moi qui ai consacré ma carrière universitaire à l'analyse de données textuelles.

[1] Je tiens à remercier tout particulièrement mon cousin Jacques Estoup qui a su conserver la mémoire de l'itinéraire familial, professionnel et scientifique de notre grand-père, et qu'un grave handicap empêche d'être parmi nous. Il a pu fournir en abondance documents et photos, partiellement présentés ici.

J'ai très peu de souvenirs de mon grand-père : j'avais 5 ans et je le revois allongé, malade, dans son lit, avec sa grande moustache blanche et son port de tête caractéristique.
Mon enfance a été baignée par ces signes mystérieux avec lesquels ma mère écrivait ses listes de courses ou ses recettes de cuisine, par des mots obscurs comme Métagraphie directe, Prévost-Delaunay, Duployé intégrale, par de lointains échos de brouilles familiales incompréhensibles.



Je savais que mon grand-père avait été un grand sténographe, sans plus.
Bien des années passent, jusqu'au jour de 1994 où je découvre, en lisant le livre *Statistique textuelle* de Ludovic Lebart et André Salem [2], qu'ils le citent, et qu'ils citent aussi Benoît Mandelbrot le citant comme un des pères fondateurs des études de fréquences de mots dans la langue.
« *... Il s'est trouvé que le premier – à notre connaissance – à s'occuper des fréquences relatives des mots dans le discours a été* ***un sténographe à l'esprit scientifique****, Jean-Baptiste Estoup. Ses résultats ont été incomparablement étendus par George Kingsley Zipf, qui enseigna à Harvard University un bizarre mélange de folles élucubrations et de faits très importants et négligés par ses contemporains, parce que trop difficiles à classer. ...* » [3]

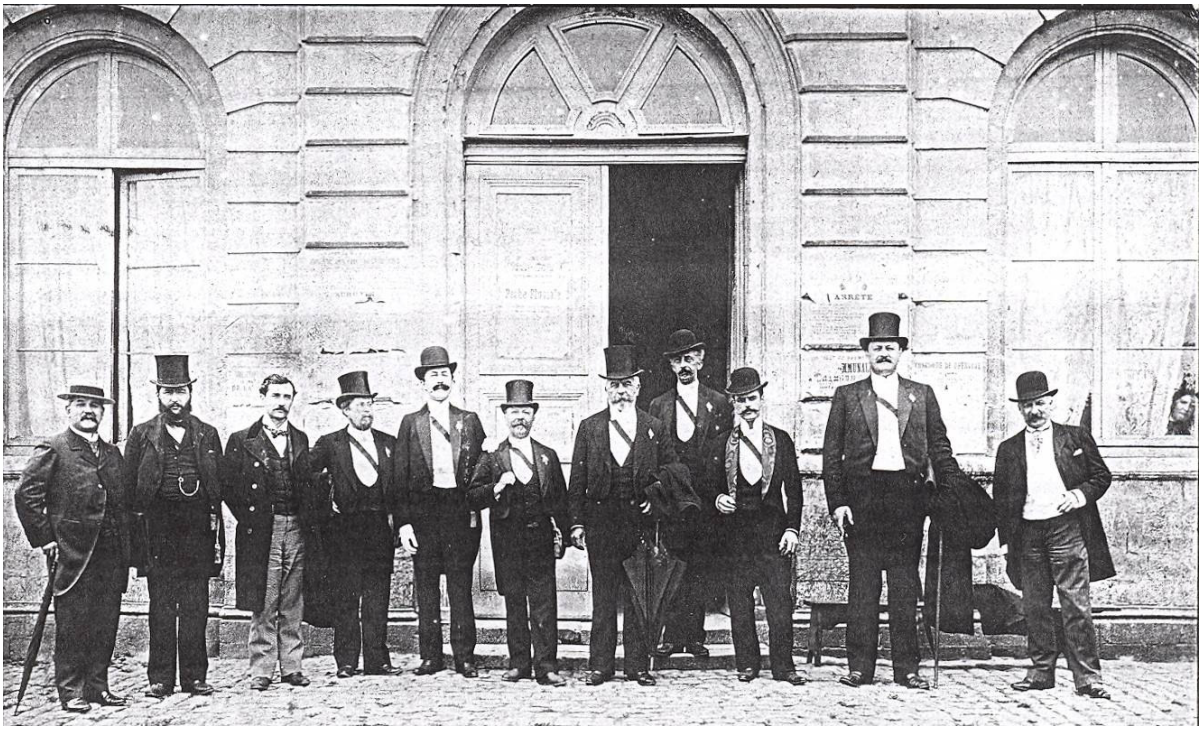
Les discussions avec mon cousin Jacques Estoup et les documents qu'il m'a apporté m'ont permis d'en savoir plus, et aussi les recherches généalogiques de mon frère Jean-Paul avec Geneviève Estoup.
Je vous retrace donc sa vie et son œuvre en première partie d'exposé.
Puis je vous parlerai plus précisément de la loi de Zipf et de quelques-uns des innombrables travaux dans son sillage.

La jeunesse de J.B. Estoup : *cap de piteu !**
* Tête de mule, en béarnais.

JB Estoup est né en 1868 d'une famille établie depuis longtemps dans le Comminges et la vallée de Luchon, marchands ou artisans, puis instituteurs, commis-voyageurs en librairie (le colportage de livres était un métier traditionnel dans la région), aubergistes, à Gaud tout d'abord, puis à Luchon au moment de la vogue du thermalisme dans la première moitié du 19e siècle, avec des fortunes diverses – ruine ou réussite sociale.
Il étudie les classiques au petit séminaire de Gourdan-Polignan (près de Saint-Gaudens), d'où il se fait exclure temporairement avec son frère pour irrespect religieux.

Bachelier es-lettres à Toulouse, il s'engage dans l'artillerie à Pau où, toujours indiscipliné, il a la révélation de la sténographie et la pratique passionnément avec un camarade sous-officier, jusqu'à payer un de ses hommes pour lui dicter des discours parlementaires ! Malgré des permissions pour concours sténographiques chichement accordées et quelques échecs, il persiste – une photo le montre sténographiant le discours d'inauguration de la statue de la Vallée du Lys, à Luchon.

Il quitte l'armée en 1896 et « monte à Paris », où il réussit aussitôt le concours de sténographe à la Chambre.



Le sténographe (« en cheveux »), la concierge (à l'extrème droite), et MM. les Députés des Pyrénées.

On le trouve en 1899 au 2e procès Dreyfus, pour le compte du Figaro, avec son ami René Havette, historien de la sténographie et ami malgré leurs différence de méthodes sténographiques, Havette pratiquant la Prévost-Delaunay et Estoup la Duployé simplifiée. La fille de René Havette, Andrée, épousera le fils de Jean-Baptiste, Jean-Henri.

Il participe activement aux concours et congrès de l'Institut Sténographique de France, et devient secrétaire général de l'Union des Sociétés Sténographiques de France.

Il se marie en 1897 avec sa cousine Henriette, malgré l'avis de son père qui tente un moment de lui demander le remboursement de ses études, et qui parle de mésalliance (pour une lointaine histoire de suicide dans la famille…) en dépit de la situation sociale plus enviable des proches de sa fiancée qui possédaient un café et deux villas de location, dont une sur l'Allée des Bains, au cœur du Luchon chic.
Ici une photo d'elle, sur plaque autochrome, par Albert Kahn, avec qui le couple fut ami, et qui leur envoya des photos de Chine.

Les premiers enfants d'une lignée de sept ne tardent pas à naître. Une carte postale montre Marguerite, Henri, Annette devant la Villa des Ormes, sous la garde d'un garçon de café de leur grand-père.
Les quatre ainés, selon l'habitude bourgeoise d'alors, seront placés en nourrice dans la montagne luchonnaise, chez des cultivateurs de Benqué-Dessus – ma mère aura toujours plus de tendresse pour *Mémé de Benqué* que pour sa vraie mère. A trois ans, elle ne parlait pas un mot de français, elle parlait la belle langue gasconne (« le patois ») qui subsiste aujourd'hui comme 3e langue officielle du Val d'Aran, après le catalan et l'espagnol.

La maturité : un progressiste (et féministe !) à la Belle Epoque.

Il y avait sans doute un brin de jalousie : ma mère disait que Père (on ne disait pas Papa-Maman chez les gens bien) vouait à Mère une admiration sans borne.
De fait il insuffla la passion de la sténographie à sa femme et à quatre de ses filles, outre son fils. Il ne ménagea pas son soutien pour leur ouvrir les portes de la profession de sténographe de discours réservée à cette époque aux hommes.
Ainsi Marguerite, pianiste virtuose et championne sténographe à 22 ans avec 190 mots à la minute, ne parvint pas à forcer la porte du Sénat (elle ne pouvait fournir son livret militaire en règle !), et trouva la reconnaissance en 1924 à la Cour Internationale de Justice de La Haye, puis à l'ONU à New-York.
De son côté ma grand-mère Henriette milita contre toute discrimination, et si elle ne put jamais forcer la porte de la Chambre des députés, elle fut sténographe de deux conseils généraux et à la Société des Nations.
Le couple noua des contacts internationaux, en particulier lors du congrès international de sténographie organisé à l'occasion de l'exposition universelle de 1900. Ils comptèrent comme amis proches les sténographes allemand R. Fuchs et turc S. Hudaverdoglu. Ces congrès étaient de véritables fêtes, occasions de concours et de luxueuses photos de groupe.

Un esprit rationnel et inventif

De Duployé à Estoup

L'abbé Emile Duployé conçut vers 1860 une écriture phonétique *pour instruire les illettrés*, une écriture populaire, simplifiée et sans ambitions de rapidité d'écriture.
Cette généreuse utopie n'eut pas le succès escompté, mais fut reprise par des sténographes, qui lui ajoutèrent bientôt des procédés d'accélération (*métagraphie*).
Un Cours Parlementaire fut publié en 1895. JB Estoup y contribue à partir de 1897 : progression plus logique et coordination améliorée entre les parties de l'ouvrage enrichissent l'édition 1898.
Mais ce cours ne le satisfait toujours pas ; il y critique une « masse de trucs » de métier hétéroclites .
Il conçoit alors les principes qui devaient mener vers un tout cohérent, et qu'il ne cessera de promouvoir avec opiniâtreté sa vie durant :

- Jeter ce qui ne correspond pas à des *règles rationnelles*
- Enseigner directement la métagraphie sans passer par l'*intégrale* – quelques règles, plutôt qu'une masse de trucs !
- Primauté aux *données d'expérience*
  - Fréquences des sons, des liaisons, des mots
  - Mesure du nombre de levées de plume et changements de direction à la minute
- Il invente le concept et crée des *gammes sténographiques* (de 50 mots/minute à 140) ; son corpus, où varient les thèmes abordés – politique, économie, commerce, sciences et techniques, droit…- atteindra 112 000 mots.

Une analyse scientifique

De ses comptages et d'expériences diverses, il ressort :

- Qu'un mot français sténographié comporte en moyenne 3,5 changements de direction de trait.
- Que la limite physiologique est de 800 changements de direction par minute.
- D'où une limite pratique de 230 mots par minute, très supérieure aux 120 à 170 mots par minute des orateurs « normaux »

Il s'oppose donc à une abréviation exagérée, et propose d'une part d'abréger les mots fréquents, tout en rendant plus lisibles les mots peu fréquents.

En effet la « traduction » est LE problème des sténographes, car l'écriture phonétique est ambiguë, et la métagraphie l'est encore plus ! Le sténographe qui traduit doit avoir le contexte « frais à l'esprit », mais aussi une bonne culture générale, et des capacités littéraires pour traduire en langue écrite correcte.

Une frénésie de comptages

Les comptages de N-grammes de caractères sont vieux comme la cryptographie. Mais c'est une frénésie de comptage de phonèmes qui s'empare des sténographes dans la deuxième moitié du 19e siècle, en particulier :

- Société Française de Sténographie (1896) : dépouillement de 33 000 mots.
- Friedrich W. Kaeding et une vingtaine de collègues de l'Université de Dresde dépouillent en 1898 un corpus de 11 millions de mots allemands ! Hélas, ce travail se révélera inutilisable dans l'optique ultérieure de Zipf…

Des comptages de *mots* apparaissent :

- Reverent J. Knowles (Londres, 1904) : 100 000 mots
- R.C. Eldridge (New York, 1911) : 35 000 mots

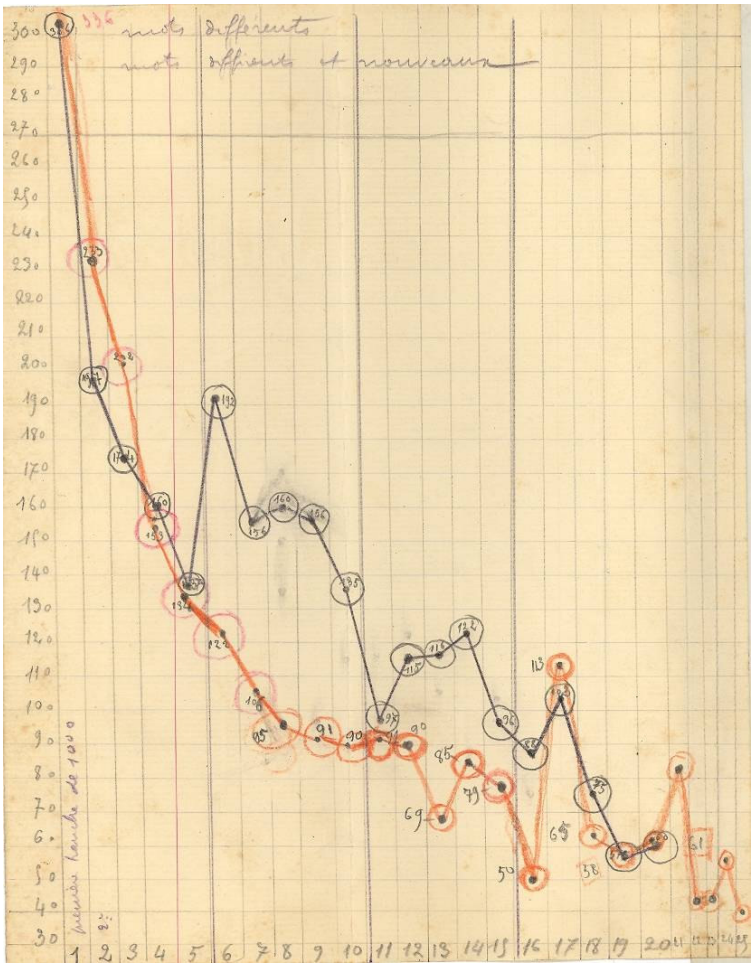
Deux études de comptage des mots du français

JB Estoup entreprend alors (avant 1912) deux études sur les mots du français. Son corpus est celui de ses gammes sténographiques, déjà calibrées par tranches d'environ 1000 mots (une à deux pages).

1- Etude d'accroissement lexical

Il s'agit de répondre à la question « Combien le discours comprend-il de mots différents ? »
Sur un corpus de 30 000 occurrences de mots :

- 20 000 sont confiés à Auguste Touzeau, professeur de sténographie (histogramme noir).
- 10 000 (au départ, 14 000 ?) sont coordonnés par JB Estoup



Après dépouillement, il observe sur le graphique fac-similé ci-dessus une relation en gros hyperbolique entre le nombre de mots nouveaux et différents d'une part, et l'effectif cumulé des mots d'autre part. Il en extrapole une limite pratique d'environ 3000 mots différents pour un corpus de 60 000 mots, vocabulaire restreint de l'orateur moyen auquel le sténographe de discours a réellement à faire. L'élève sténographe devra les apprendre et s'entraîner jusqu'à les rendre automatiques.

2- Etude des répétitions

Elle répond à la question « Combien de fois sont répétés les mots les plus fréquents ? », de façon à guider, à informer objectivement les initiatives de création ou d'amélioration des signes.
Elle se base sur un corpus de 20 000 mots :

- JB Estoup extrait une liste de listes de mots répétés n fois, par ordre décroissant de répétitions ; sont regroupés certains homophones courants (*et, est, ai*), des articles (*le, la, les*), les formes conjuguées des verbes, les formes singulier et pluriel des noms, les formes masculines et féminines des adjectifs.
- En-deçà de 7 répétitions, les mots ne sont plus détaillés.
- Pas de graphique (mais plus tard Zipf traduira directement de façon graphique : nombre de mots répétés × nombre de répétitions, en coordonnées log-log).

	Nombre des répétitions		Nombre des répétitions
Le, la, les	1949	peut, peu	58
de, du, des	1712	dire	57
que, qui	766	comme	56
à, au	743	bien	53
et, est, ai	741	industrie	50
un, une	413	français	48
ce, se	393	certain	44
il	282	grand	44
ne	279	avec	41
en	258	sans	40
ces, ses, cette	232	venir (et temps)	39
être (et temps)	200	quel	39
dans	181	commerce	38
pas	179	nouveau	37
nous	167	très	37
plus	162	vouloir (et temps)	37
pour	157	droit	36
par	155	autre	36
on	155	produit	36
son, sont	141	environ	35
je	137	falloir (et temps)	35
faire (et temps)	136	voir (et temps)	35
tout, toute	119	donner (et temps)	34
avoir (et temps)	118	savoir (et temps)	33
nos	115	nation	32
vous	108	développer	31
ou, où	100	économique	31
leur	94	messieurs	30
si	93	sa	30
elle	92	France	29
même	84	quelque	29
sur	82	trouver (et temps)	29
mais	82	intérêt	29
y	70	aller (et temps)	28
pouvoir (et temps)	66	lui	27
celui, celle, ceux	62	nombre	27
dont, donc	62	point	26
devoir (et temps)	60	public	26
travailler (et temps)	60	année	25

Ces deux études furent publiées en 1916, semble-t-il, dans le fascicule théorique qui accompagnait désormais chaque édition de ses deux livres de gammes sténographiques, la 4e en l'occurrence ; nous n'avons dans la famille que la 7e édition. La 4e semble conservée, avec son fascicule d'annexes, à la bibliothèque municipale de New-York, où B. Mandelbrot a pu la consulter. La 3e en 1912 comporte une allusion à la limite des 3000 mots courants, ce qui ferait remonter les comptages à une période antérieure à 1912 [1].

moins	25	chose	23
demander (et temps)	25	Etat	23
tant, temps	24	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, formuler, 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, démocratie, 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.

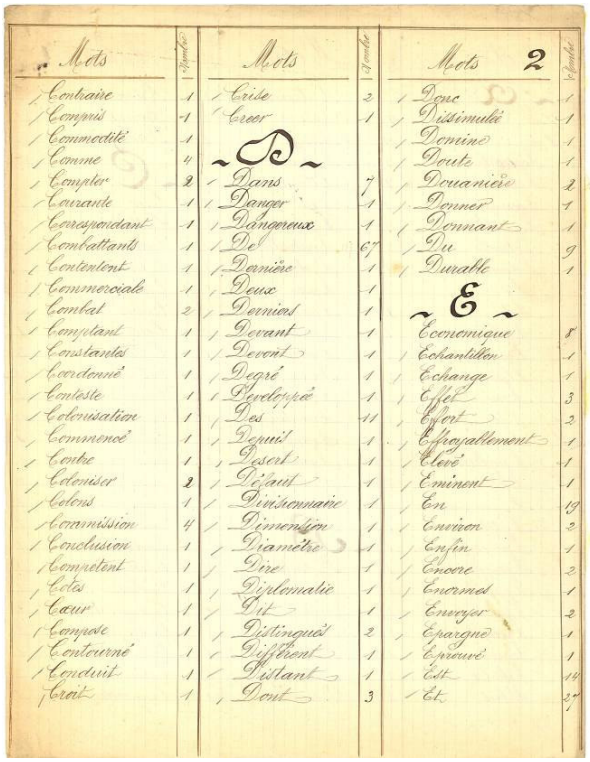
Sont répétés 7 fois :
action, adresse, admettre, affaire, agir, Allemagne, Allemand, Amérique, apparaître, apprenti, assez, atelier, avenir, cependant, chemin, civiliser, colon, combien, constituer, déjà, dessin, divers, doute, échange, entier, établir, extérieur, fois, fromage, garder, houille, impôt, 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.

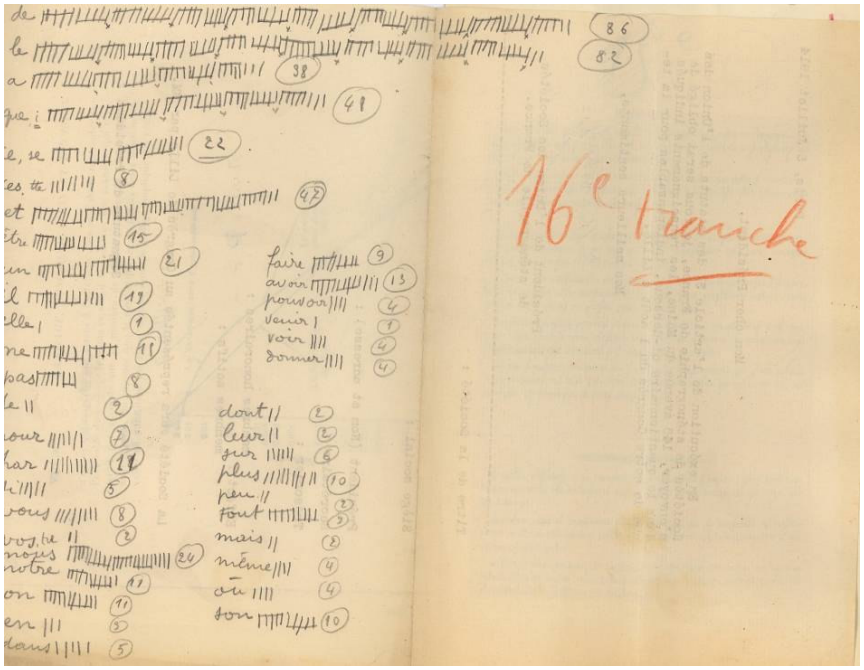
Puis viennent :

62 mots répétés 6 fois (1) ;
82 mots répétés 5 fois ;
131 mots répétés 4 fois ;
194 mots répétés 3 fois ;
329 mots répétés 2 fois ;
Enfin 922 mots ne se présentent qu'une fois.

(1) La liste n'a plus d'intérêt.

Il subsiste une bonne partie des dépouillements partiels par tranches. C'est un travail réparti, comme en témoignent la diversité des écritures et quelques noms : André Fauconnier, Max Müller (sténographe anglais ?), …

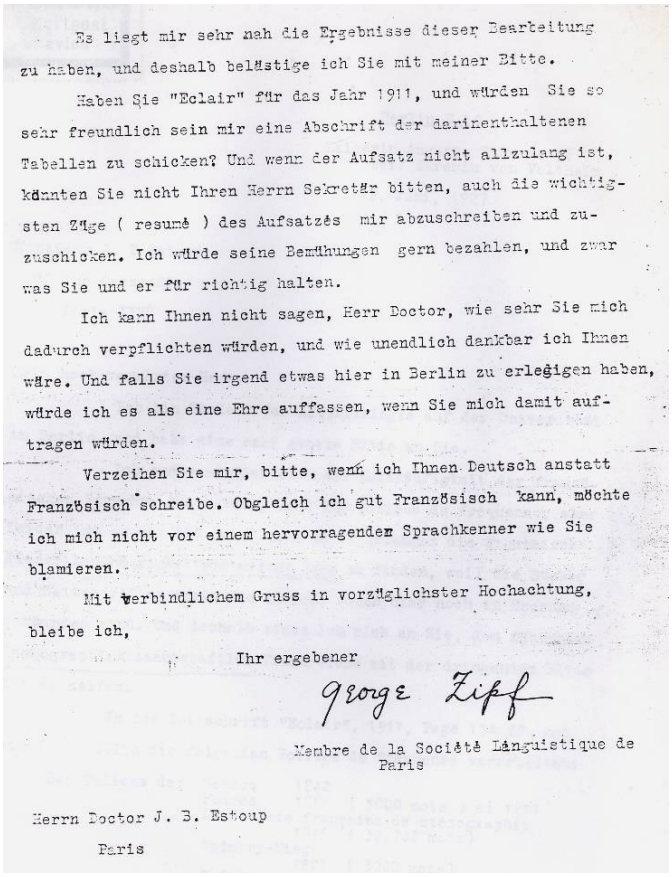




JB Estoup, formé à la littérature et aux classiques, ne s'est jamais intéressé à la théorie en soi ; mais il a appliqué avec obstination au seul perfectionnement de la technique (ou de l'art ?) sténographique une démarche rationnelle d'observation – d'où son initiative de comptages - et d'expérimentation (en testant au besoin ses innovations sur ses élèves et ses propres enfants…) ; son action a tendu constamment vers l'épuration progressive d'un système entaché de beaucoup d'arbitraire au départ, et vers son enseignement par un apprentissage progressif, mais à un seul degré, les mêmes règles s'appliquant au fil de la difficulté croissante des mots et de la vitesse à acquérir (par exemple, celles de transcription des diphtongues). Cette démarche lui a valu le succès – ses gammes sténographiques auront été tirées au total à une centaine de milliers d'exemplaires – mais aussi une dérive croissante par rapport à d'autres collègues plus conservateurs.

Estoup et Zipf

George Kingsley Zipf, né en 1902 aux Etats-Unis et mort en 1950, mena une thèse de philologie en Allemagne de 1924 à 1929 à Bonn, puis Berlin ; il la soutint à Harvard en 1929. Dans sa lettre à JB Estoup (18 juin 1927), il lui demande avec une extrême politesse de lui faire envoyer une copie des tableaux de l'article du sténographe JB Illio dans le journal L'Eclair [sténo-dactylographique ?] de 1911 concernant une liste de « polices de fréquence » (en français dans le texte), en allemand *Lauthaüfigkeit*, plutôt fréquences de phonèmes. [Mince espoir de vérification qu'il ne s'agissait pas plutôt des *fréquences de mots* : à la BNF ce périodique est classé « HU », hors d'usage… Il paraît vraisemblable qu'il s'agisse de phonèmes, sur lesquels portait sa thèse].



Es liegt mir sehr nah die Ergebnisse dieser Bearbeitung zu haben, und deshalb belästige ich Sie mit meiner Bitte.

Haben Sie "Eclair" für das Jahr 1911, und würden Sie so sehr freundlich sein mir eine Abschrift der darinenthaltenen Tabellen zu schicken? Und wenn der Aufsatz nicht allzulang ist, könnten Sie nicht Ihren Herrn Sekretär bitten, auch die wichtigsten Züge (resumé) des Aufsatzes mir abzuschreiben und zuzuschicken. Ich würde seine Bemühungen gern bezahlen, und zwar was Sie und er für richtig halten.

Ich kann Ihnen nicht sagen, Herr Doctor, wie sehr Sie mich dadurch verpflichten würden, und wie unendlich dankbar ich Ihnen wäre. Und falls Sie irgend etwas hier in Berlin zu erledigen haben, würde ich es als eine Ehre auffassen, wenn Sie mich damit auftragen würden.

Verzeihen Sie mir, bitte, wenn ich Ihnen Deutsch anstatt Französisch schreibe. Obgleich ich gut Französisch kann, möchte ich mich nicht vor einem hervorragenden Sprachkenner wie Sie blamieren.

Mit verbindlichem Gruss in vorzüglichster Hochachtung, bleibe ich,

Ihr ergebener

George Zipf

Membre de la Société Linguistique de Paris

Herrn Doctor J. B. Estoup

Paris

Première formulation de la loi de Zipf

- 1929 : thèse de philologie comparée, *Relative Frequency as a Determinant of Phonetic Change*, où JB Estoup fait partie des personnes remerciées.
- 1932, ouvrage : *Selected studies on the principle of relative frequency in language* [4]
  - cite JB Estoup
  - traduit directement en graphique log-log le formalisme « liste de listes » de l'étude Estoup des fréquences de mots :
    - Abscisse (a) : effectifs des classes de fréquences de mots (mots revenant 1 fois, 2 fois, …)
    - Ordonnée (b) : fréquences de ces mots
  - … en l'appliquant à d'autres données : latin de Plaute, anglais (comptages Eldridge), mots et unigrammes chinois.
  - constate la relation « universelle » pour 95% des mots : $\boldsymbol{ab^2=constante}$, formule « exactement identique à celle de la gravitation »…
  - …mais « triche » pour les mots d'effectifs 1 (les hapax, disent les linguistes), qui devraient avoir une fréquence fractionnaire pour obéir à sa loi, en ne les représentant pas !

Deuxième formulation de la loi de Zipf

- 1935 : *The psycho-biology of language* [5]
- Nouvelle formulation en utilisant les rangs des mots classés par effectifs décroissants

  - Abscisse : rangs des classes de fréquences de mots (mot le plus fréquent, mot le 2e plus fréquent, …)
  - Ordonnée (b) : fréquences de ces mots
- constate la nouvelle relation « universelle »

fréquence × rang = constante

- …qui inclut cette fois les mots d'effectifs 1.

On sait depuis [6]**…**

- Que ces 2 formulations sont équivalentes
  - La 1ère peut être exprimée comme une loi de densité de probabilité : $P(j) \sim j^{-b}$, autrement dit, probabilité qu'un mot soit présent j fois dans le corpus
  - La 2e comme comme une loi de densité de probabilité qu'un mot ait le rang i par ordre de fréquences décroissantes : $F(i) \sim i^{-a}$
  - Avec, comme pont entre les deux, la relation : $b = 1 + 1/a$

Après Estoup et Zipf…

Les constats et formalisations progressives d'Estoup et Zipf s'insèrent dans un courant général de découvertes de ces « lois de puissance » dans de très nombreux domaines des sciences de l'homme et de la vie (économie, réseaux sociaux, génomique, …), tout comme des sciences de la matière (longueur des fleuves, météorologie…), de la fin du 19e siècle à nos jours – cf. la synthèse de Marc Barbut [7] ; ils ont suscité beaucoup de tentatives de modélisation explicative. On peut citer dans le domaine de la langue :

- Benoît Mandelbrot, en 1960 (via l'entropie de Shannon) [8].
- Harald Baayen, en 2001 (via les distributions LNRE = *Large Number of Rare Events*) [9].

Epilogue : un passionné dans une époque de passions

Les années de guerre 1914-1918 furent douloureuses : séances de nuit interminables à la Chambre, difficultés de la vie dans Paris en guerre et décès de sa fille Françoise âgée de dix ans, renoncement à ses amitiés internationales, comme R. Fuchs en Allemagne.

Il crée cependant en 1917 le bulletin La Vérité Sténographique, support de ses convictions et de ses enseignements, qui lui survivra jusqu'en 1992 ; il continue plus que jamais ses multiples activités de chercheur indépendant (publication de son étude des fréquences en 1916), professeur, éditeur, et prosélyte de sa « Métagraphie directe » duployenne.

1918 vît le début d'éclatement des duployens entre ceux qui refusaient de reprendre contact avec les sténographes « ennemis de la France », et ceux, comme JB Estoup et ses amis, qui le souhaitaient ardemment. En 1924 ceux-ci quittent l'Institut Sténographique de France et fondent l'Institut International de Métagraphie Duployé. Il prend sa retraite de sténographe parlementaire en 1929, mais continue de plus belle ses activités.

Après la deuxième guerre mondiale et les premières ombres poprtées sur l'avenir de la sténographie par l'enregistrement magnétique naissant, il devenait urgent que les divers courants de la sténographie Duployé présentent un front uni pour perdurer dans l'enseignement public français : ce fut le grand projet de « Codification » achevé en janvier 1949 par Henri, fils de Jean-Baptiste, qui ajoutait aux talents de son père – théoricien, praticien et animateur – celui de conciliateur. Jean-Baptiste, qui n'était pas homme de compromis, mourut en avril 1950.

Bibliographie

1 - M. Petruszewycz – L'histoire de la loi d'Estoup-Zipf : documents. Mathématiques et Sciences Humaines N°44, 1973, pp. 41-56.
2 - L. Lebart et A. Salem - Statistique textuelle, Paris, Dunod, 1994
3 - B. Mandelbrot (1968) - Les constantes chiffrées du discours, Le Langage,. Encyclopédie de la Pléiade, vol XXV, Gallimard, Paris
4 – G.K. Zipf (1932): Selected Studies of the Principle of Relative Frequency in Language. Cambridge (Mass.).
5 – G.K. Zipf (1935): The Psycho-Biology of Language. Cambridge (Mass.).
6 - S.D. Haitun, Stationary scientometric distributions. Part II. Non-Gaussian nature of scientific activities. – Scientometrics, 4 (1982) N°2, pp.89-104.
7 - M. Barbut, « Homme moyen ou homme extrême ? De Vilfredo Pareto (1895) à Paul Lévy (1935) en passant par Maurice Fréchet et quelques autres », Journal de la Société Française de Statistique, vol. 144, n° 1-2, 2003.
8 - B. Mandelbrot - On the theory of word frequencies and on related markovian models of discourse. Structure of Language and its Mathematical Aspects (New York, 1960). Ed Roman Jakobson (Symposia in Applied Mathematics XII). Providence, R.I.: American Mathematical Society, 190-219.
9 - R. H. Baayen, Word Frequency Distributions, Kluwer Academic Publishers, Dordrecht, 2001.