

JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

MICHEL TESNIÈRE

Fréquence des noms de famille

Journal de la société statistique de Paris, tome 116 (1975), p. 24-33

http://www.numdam.org/item?id=JSFS_1975__116__24_0

© Société de statistique de Paris, 1975, 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/conditions>). 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/>

FRÉQUENCE DES NOMS DE FAMILLE

W

*« En la Toponimia i en l'Onomàstica s'emmi-
ralla tota la història i cultural d'un país »*
Pompeo FABRA

The author shows the method he used to find out the most frequent family names, to assess their frequency and to test the results he got. He finds out that a million people bear the twelve first most usual names.

Der Verfasser gibt die Technik an, die er verwendet hat um die am häufigsten vorkommen den Familiennamen festzustellen. Um ihre Häufigkeit abzuschätzen und um die erhaltenen Resultate zu prüfen. Er stellt fest, dass die zwölf Namen, die in erster Linie stehen von mehr als einer Million Personen getragen werden.

Parmi tous les noms de famille, l'un des plus beaux, sinon le plus beau, c'est certainement ... le vôtre.

Vous en connaissez l'origine, la signification, souvent l'histoire. Mais avez-vous une idée de sa fréquence, ou si vous préférez du nombre de personnes qui portent le même nom que vous ?

M. Malinvaud, qui avez un nom caractéristique du Limousin, combien y a-t-il de Malinvaud ? Et vous tous, dont « le plus beau patrimoine est un nom vénéré » comme dit Victor Hugo, savez-vous, même approximativement, combien vous avez d'homonymes ?

Telles sont les questions que nous allons aborder ensemble, en deux temps, d'abord la méthode, ensuite les résultats.

La méthode ?

Tout a commencé il y a une dizaine d'années. Avec quelques parents, nous avons décidé de réunir au Mont-Saint-Michel tous les *Tesnière* de France et de Navarre. Or, pour réserver une salle de restaurant, il était nécessaire de savoir combien de personnes viendraient effectivement. Environ 10 à 15 % du nombre total de *Tesnière*, nous avait prédit un traiteur habitué à ce genre d'exercice.

Le problème se déplaçait. Pour estimer le nombre probable de convives, il fallait maintenant évaluer combien il y avait de *Tesnière* en tout.

Comment faire ? Il n'existe aucune table, aucun dictionnaire donnant le nombre de personnes qui portent un même nom. Aussi, après bien des hésitations, nous avons dépouillé les annuaires téléphoniques pour les 38 000 communes de France.

En procédant ainsi, nous avons découvert qu'il y avait 40 *Tesnière* abonnés au téléphone. A raison d'un abonné pour 25 habitants, cela faisait 1 000 *Tesnière* environ. En prenant les 10 à 15 % annoncés par le traiteur, il devait venir 100 à 150 *Tesnière*.

Sans trop y croire, on réserva une salle pour 120 personnes. A tout hasard, l'arrière-salle fut aussi retenue. Et maintenant, combien pensez-vous qu'il vint de *Tesnière*? 173. C'était un grand succès pour la réunion et accessoirement une première confirmation que le mode de calcul était valable.

Si d'aventure vous aviez l'intention d'en faire autant, je me ferais un plaisir de vous fournir une méthode plus rapide pour évaluer votre nombre d'homonymes.

N'oubliez pas non plus, si les gens doivent arriver la veille, de prévenir les hôteliers et voici pourquoi.

Avec prudence, chaque *Tesnière* avait réservé des chambres à son nom, *Tesnière*. Au premier, on avait répondu : c'est d'accord. Au second : c'est déjà noté. Au troisième : ne vous inquiétez pas, vous êtes bien inscrits. Et ainsi de suite dans les différents hôtels du Mont-Saint-Michel. Partout, un seul nom, *Tesnière*.

Je vous laisse imaginer la confusion qui s'empara du Mont lorsque la marée des *Tesnière* déferla. Plus personne ne savait où il en était.

Nous avons vu tout à l'heure qu'il y avait environ 1 000 *Tesnière*. Les recoupements ultérieurs ayant confirmé cet ordre de grandeur, nous avons étendu la méthode à tous les noms de famille, en commençant par les plus fréquents.

L'entreprise présente trois difficultés majeures.

Premièrement, comment savoir *a priori* quels sont les noms de famille les plus fréquents parmi les centaines de milliers de noms différents?

Bien sûr, il y a *Martin*, *Durand*, *Dupont* et quelques autres. Et ensuite?

Nous nous étions fixé comme objectif de trouver les 1 000 premiers noms de famille. Mais comment les repérer? C'est un sondage dans l'annuaire téléphonique de Paris qui donna la réponse. Il fallait retenir les noms représentés par 50 abonnés et plus. Ainsi 1 100 noms, présumés les plus fréquents, furent isolés.

Deuxième difficulté. Il n'était évidemment pas question de relever dans les 38 000 communes le nombre d'abonnés portant chacun des 1 100 noms retenus. Nous nous sommes limités à 38 villes choisies surtout en fonction de considérations linguistiques régionales afin de bien couvrir l'ensemble de la France et notamment ses extrémités bretonne ou alsacienne, flamande ou catalane.

Malgré tout, cela représentait encore un travail matériel important puisqu'il y avait près de 40 000 relevés élémentaires à effectuer, et ensuite autant de multiplications par des coefficients multiplicateurs appropriés. Mais, comme dit un proverbe chinois :

La patience et le temps
Font de la feuille de mûrier
Robe de satin blanc.

Troisième difficulté, tester les résultats et par la même occasion les hypothèses de travail adoptées. Pour cela, les considérations qui nous ont guidés sont de deux ordres.

Premièrement, utiliser des sources d'information autres que les annuaires téléphoniques. Nous avons pris les tables alphabétiques de différents annuaires nationaux, par exemple les annuaires des grandes écoles.

Deuxièmement, établir une procédure de contrôle. Là, nous avons fait appel au calcul des probabilités. La fréquence des noms de famille, rapportée aux 50 millions de

Français est très petite. Pour le nom *Tesnière*, elle est de $1/50\ 000$ soit $2 \cdot 10^{-5}$. En conséquence la probabilité d'apparition d'un nom donné dans une table alphabétique sera également très petite et nous avons supposé que le phénomène était de ce fait justiciable de la loi de Poisson sur les petites probabilités.

L'expérience a d'ailleurs montré que le nombre d'homonymes observé dans les index alphabétiques se situait généralement bien à l'intérieur de l'intervalle de confiance annoncé par la théorie.

Nous voici donc en possession d'un répertoire des 1 000 noms les plus fréquents. On dispose pour chacun d'eux :

- du nombre d'homonymes;
- de la fréquence;
- du rang;
- de la répartition régionale.

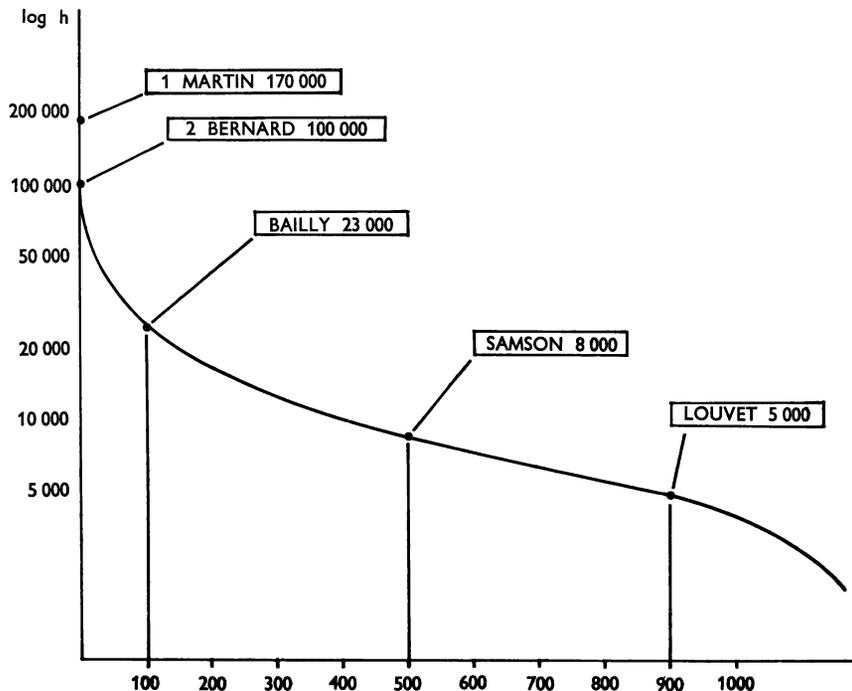
Que révèle cette documentation?

Voyons d'abord les premiers noms les plus fréquents. Ce sont :

Rang	Nom	Homonymes	Fréquence
1.	<i>Martin</i> . . .	170 000	$1/300 = 3,4 \cdot 10^{-3}$
2.	<i>Bernard</i> . . .	100 000	$1/500 = 2,0 \cdot 10^{-3}$
3.	<i>Durand</i> . . .	80 000	$1/650 = 1,6 \cdot 10^{-3}$
4.	<i>Dubois</i> . . .	79 000	$1/650 = 1,6 \cdot 10^{-3}$
5.	<i>Petit</i>	78 000	$1/650 = 1,6 \cdot 10^{-3}$

On tombe ensuite autour de 72 000 avec *Thomas*, *Robert*, *Moreau*, *Richard* et *Michel*. *Martin* et *Bernard* se détachent nettement. Même si la marge d'erreur dans les évaluations reste grande, leur classement en tête paraît indiscutable.

Le nombre d'homonymes diminue rapidement, ainsi que le montre le graphique.



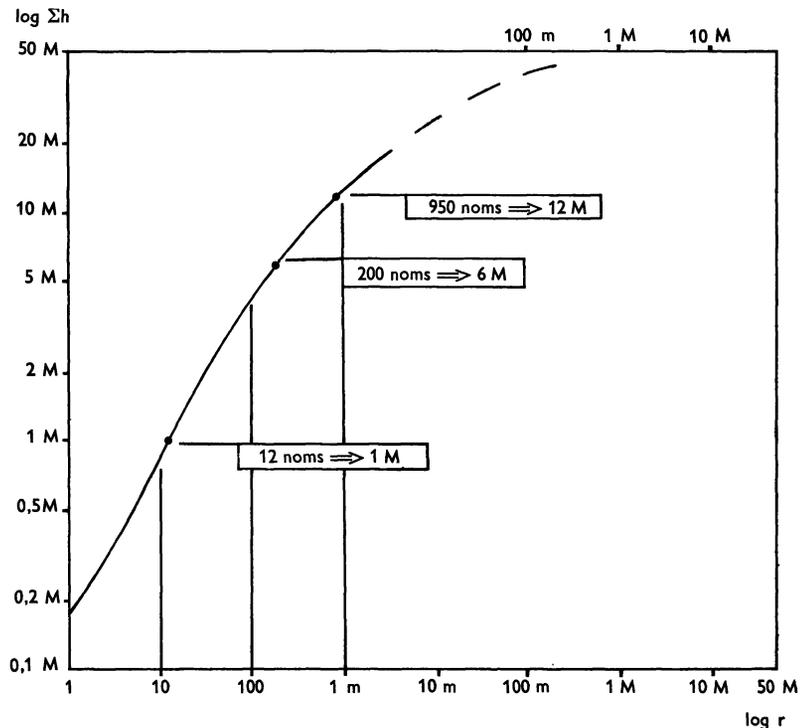
Au 100^e rang, *Bailly* 23 000 homonymes.

Au 500^e rang, *Samson* 8 000 homonymes.

Au 800^e rang, *Louvet* 5 000 homonymes.

A partir du 750^e rang, il est probable qu'il manque des noms et ceci expliquerait la forme et la courbe à son extrémité.

La courbe cumulée des homonymes est également instructive.



Les 12 premiers noms sont portés par un million de personnes. Le cap des 6 millions est franchi vers le 200^e nom, celui des 12 millions au 950^e nom.

Bref, le répertoire concerne directement un quart de la population française. Quel marché potentiel pour un livre sur les noms de famille les plus fréquents, me confiait récemment un ami!

Changeons de discipline.

On sait que les noms de famille ont commencé à se transmettre héréditairement, en général de père en fils, à partir du XIII^e siècle. La majorité des noms était définitivement fixée au XVI^e siècle de telle sorte que l'immense collection des noms actuels est un reflet assez fidèle de la langue et de la civilisation de la fin du Moyen Age.

L'examen des annuaires modernes permet ainsi de remonter d'un seul bond à six ou sept cents ans en arrière et de revivre le langage et les préoccupations de nos aïeux d'il y a une vingtaine de générations.

Voici deux séries d'exemples : les métiers et les mois de l'année. D'abord les noms de métier.

Coiffeur et *Perruquier* ne sont pas des noms de famille tandis que 38 000 personnes s'appellent *Barbier*. En effet, *perruquier* date de la Renaissance, *coiffeur* du siècle de Voltaire.

En groupant les noms qui désignent le même métier, on obtient le classement suivant :

Métier	Nombre	Noms
1. « forgeron »	200 000	<i>Fabre, Faure, Favre, Lefebvre, Lefèvre</i>
2. « boulanger »	70 000	<i>Fournier, Boulanger</i>
3. « bouvier »	70 000	<i>Bowier, Bouyer, Boyer</i>
4. « meunier »	60 000	<i>Meunier, Monnier, Lemonnier</i>
5. « colporteur »	50 000	<i>Mercier, Lemercier</i>
6. « charpentier »	45 000	<i>Carpentier, Charpentier</i>
7. « marchand »	42 000	<i>Marchand, Lemarchand</i>
8. « barbier »	38 000	<i>Barbier</i>

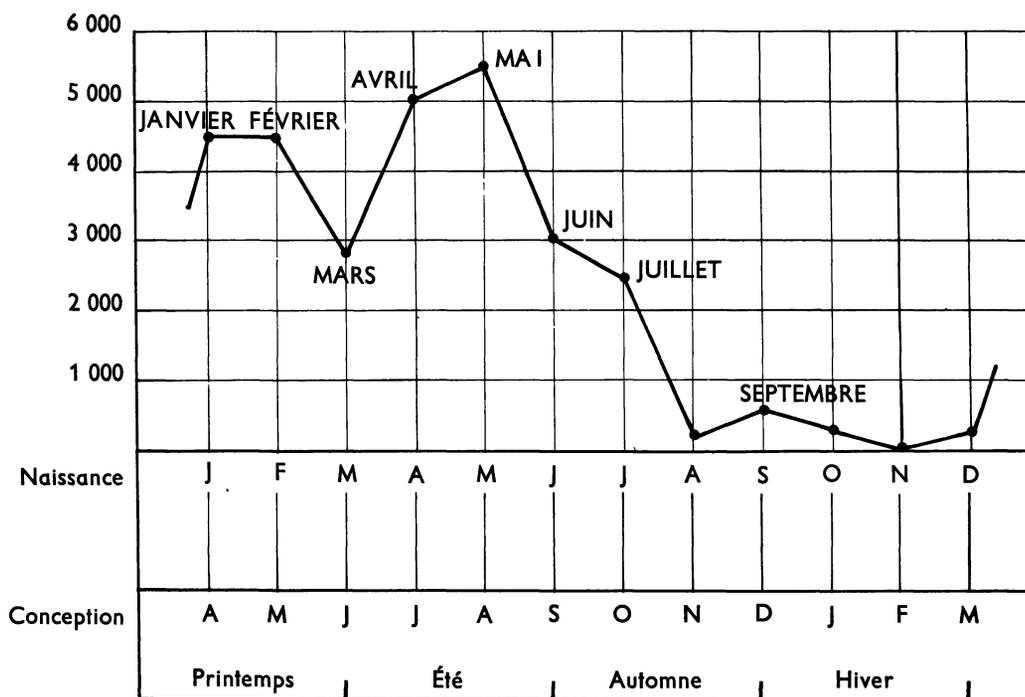
Tout un monde oublié réapparaît.

Et qui sait s'il ne serait pas possible, à partir de ces données, de reconstituer des statistiques professionnelles des *xiv^e* et *xv^e* siècles.

Venons-en maintenant aux noms de famille qui évoquent les mois de l'année, *Janvier, Février*, etc.

Selon les linguistes, il s'agit de noms donnés à des enfants trouvés, le mois étant celui de la naissance ou du baptême.

Comment se présente le nombre d'homonymes suivant le mois?



Le creux de mars peut s'expliquer par le fait que certains enfants nés à cette époque étaient tout simplement appelés *Pascal*.

La pointe pour le nom *May* correspond en partie à une autre étymologie. Autrefois, celui qui avait le privilège, le 1^{er} mai, de planter l'arbre symbolisant le réveil printanier de la nature était surnommé, comme l'arbre d'ailleurs, *May* ou *Lemay*. L'existence du nom de famille *Lemay* en est une preuve.

Bref, de janvier à juin, le niveau de la courbe oscillerait entre 4 000 et 5 000 pour devenir insignifiant à partir du mois d'août. Comment expliquer cette dissymétrie?

Faut-il remonter neuf mois en arrière et conclure que les enfants trouvés étaient conçus essentiellement au printemps et en été? La question est posée aux historiens, aux démographes et aux sociologues de la fin du Moyen Age.

Mais revenons à nos jours et plus spécialement, pour terminer, au nom *Malinvaud*.

Les statistiques montrent qu'il y a environ 1 500 *Malinvaud* à $\pm 20\%$ près. Cela fait une fréquence de $1/33\ 000$ soit $3 \cdot 10^{-5}$.

Ouvrons le carnet rose. Chaque année, l'I. N. S. E. E. enregistre 800 000 naissances environ, $8 \cdot 10^5$. En supposant que la fréquence ne varie pas avec le temps, il naît donc :

$$8 \cdot 10^5 \times 3 \cdot 10^{-5} = 24 \text{ Malinvaud par an}$$

ou encore :

un petit *Malinvaud* tous les quinze jours.

Bravo!

Michel TESNIÈRE

Ancien élève de l'École polytechnique

DISCUSSION

M. J. GARNIER. — Je signale à M. Tesnière que *Économie et Statistique* a publié il y a un an ou deux une courte étude sur la fréquence des noms patronymiques basée sur un dépouillement dans une soixantaine de département.

Réponse :

1. Étude de l'I. N. S. E. E. dans *Économie et Statistique*.

Voir commentaire ci-dessous.

2. Le nom *Garnier*.

Comme toute peine mérite récompense, pour vous remercier, voici ce que nous avons trouvé sur votre nom.

Il y a environ 43 000 *Garnier*. Cet ancien nom de baptême d'origine germanique apparaît avec une fréquence qui dépasse la moyenne dans l'Ouest, l'Est et le Sud-Est, une fréquence moyenne ailleurs sauf dans le Nord où le nom est rare.

Pour le carnet rose, il naît près de deux petits *Garnier* par jour. Bravo!

M. BENEZECH. — L'orateur nous a parlé de noms de métier et de noms de mois. Peut-il nous dire quelles sont les principales catégories de noms de famille?

Réponse :

Catégories de noms de famille.

M. *Bénézech* soulève une question importante. L'étude de l'origine du nom patronymique varie beaucoup suivant la catégorie à laquelle il appartient.

Dauzat adopte la classification en quatre groupes suivante :

1. les noms de baptême (*Bénézech* — forme languedocienne de *Benôit* —, *Durand*, *Garnier*, etc.). Pendant longtemps, l'Église laissa toute liberté pour le choix des noms de baptême. On prenait des noms à la mode. Le culte des saints s'est vulgarisé au XII^e siècle seulement. L'obligation de choisir parmi les saints, le nom de baptême date du concile de Trente en 1563;

2. les noms d'origine (*Dubois*, *Lavigne*, *Malinvaud*, *Pagézy*, *Tesnière*, etc.);

3. les noms de profession, d'état, de parenté (*Fabra*, « forgeron » en catalan, *Leclerc*, *Neveu*, etc.);

4. les surnoms (*Courtois*, *Leduc*, *Petit*, etc.).

Les diminutifs affectifs comme *Martinet* sont rattachés au nom dont ils dérivent, *Martin* dans le cas présent.

J.-C. ANTOINE demande à l'orateur s'il n'a pas essayé de recouper les résultats fournis par les annuaires par d'autres formes de sondages. Il est en effet à craindre que la population des abonnés au téléphone, ou plus généralement, inscrite dans des annuaires professionnels ou autres, ne comprenne — plus largement que le reste de la population — des familles appartenant depuis des siècles à des milieux bourgeois ou aristocratiques chez lesquels les noms n'étaient sans doute pas les mêmes qu'en milieu paysan. La population des inscrits sur les listes électorales paraît beaucoup moins « biaisée » à cet égard ; l'auteur ne devrait-il pas essayer d'effectuer des recoupements de cette manière?

...demande d'autre part s'il n'est pas concevable que certains noms aient pu servir à dénommer à l'origine des sortes de clans locaux plutôt que des individus ou des familles individuelles. Dans le cas des Antoine (ou Anthoine) en particulier, la plupart paraissent provenir de deux origines géographiques bien définies : la Lorraine et le Dauphiné. On peut encore constater actuellement que dans certains villages des environs d'Embrun, près du tiers des habitants portent le nom d'Anthoine. D'autres cas de ce genre s'observent dans différentes parties des pays de langue d'Oc. Ils suggèrent qu'il s'agissait moins de différencier les familles les unes des autres que de faire apparaître au contraire l'existence d'un bien entre certaines familles formant peut-être à l'origine une sorte de confrérie groupée autour d'un saint protecteur commun. Ceci est une hypothèse, à faire vérifier par les historiens.

Enfin, il aurait lieu — dans une étude correcte de la fréquence des noms, de prendre en considération les variations orthographiques d'un même nom, liées à des questions de dialecte. Il est arrivé souvent que des familles changent l'orthographe de leur nom en passant du patois au français, surtout au début du XIX^e siècle.

Réponse :

1. Recoupements

La question du recouplement des résultats est essentielle. Sans aucun doute, l'approche que nous avons utilisée pour la première fois de façon aussi poussée, est-elle insuffisante. Tester les résultats obtenus à partir des annuaires téléphoniques avec des chiffres issus d'autres annuaires, aussi bien choisis soient-ils, permet d'éviter des erreurs grossières sans donner toute la certitude désirable. Aussi, depuis plusieurs années, nous cherchions des

moyens de preuve plus solides. Or le document signalé par M. *Garnier* apporte une confirmation étonnante, comme on le verra plus loin.

2. Origine du nom de famille Antoine

Comme beaucoup de noms de saint, Antoine apparaît avec une fréquence nettement supérieure à la moyenne régionale dans l'Est de la France. Mais on trouve des Antoine dans toutes les régions; avec une fréquence normale dans le Nord, la Région parisienne et le Sud-Est, avec une fréquence légèrement inférieure dans l'Ouest, le Sud-Ouest et le Centre.

Au total, il y a environ 17 000 Antoine.

Maintenant comment Antoine est-il devenu nom de famille? A côté de la voie individuelle, peut-on concevoir une dénomination collective de famille autour d'un saint patron ou d'autre manière? L'hypothèse ne doit pas être écartée. Aux historiens de répondre.

3. Orthographe.

Pour étudier complètement un nom, il faut évidemment examiner toutes les variantes orthographiques. Pour notre part, nous avons travaillé en considérant que Antoine et Anthoine étaient deux noms distincts. Cette convention n'interdit pas les regroupements ultérieurs fondés sur l'étymologie, par exemple. Par ailleurs, les gens n'aiment pas, en général, qu'on écorche leur nom et je suis persuadé que les Anthoine tiennent beaucoup à leur H, car c'est une marque qui les distinguent nettement des autres Antoine. Vous-même, M. Antoine, admettriez difficilement qu'on ajoute un H inutile à votre nom. Chacun s'attache à son nom et le défend tel qu'il est orthographié.

Marie LAVIGNE. — Comment peut-on expliquer que le nom le plus répandu, parmi les noms de métiers, soit celui du forgeron, Lefèvre ou Fabre, etc.? On peut d'ailleurs aussi le constater dans d'autres langues (Smith, Schmidt, Kouznetsov). A quoi tient cette prédominance?

Réponse :

1. Pourquoi le nom du « forgeron » est-il le plus répandu des noms de métier?

L'observation de M^{me} Lavigne est très pertinente. Malheureusement, je ne suis pas en mesure de lui donner une réponse et j'ignore même si la question a déjà été traitée.

Pour les langues étrangères, il serait instructif de comparer la fréquence des Smith, Schmidt, etc., et celle des Lefèvre, Fabre, etc. Mais existe-t-il seulement des répertoires de fréquence des noms de famille anglais, allemands ou russes?

Limitons-nous à la France. Pour devenir nom de famille, la dénomination doit identifier et distinguer. Le caractère retenu peut avoir un sens (nom de métier, nom d'origine, surnom) ou n'être qu'une simple étiquette (nom de baptême, diminutif affectif).

A la fin du Moyen Age, parmi les métiers, quels étaient ceux qui pouvaient remplir ce rôle d'identification dans le village, la paroisse, la seigneurie ou la ville naissante? Les métiers spécialisés comme les métiers d'artisan, surtout s'ils se transmettaient de père en fils. Forgeron appartient sans conteste à cette catégorie.

Mais combien y avait-il de forgerons? Avec quelle densité territoriale? Même si nous avons les réponses, pourrait-on considérer que les métiers caractéristiques ont donné partout la même proportion de noms de famille? Voilà autant de questions auxquelles il faudrait répondre.



2. L'attribution d'un nom

L'attribution de certains noms comme Leprince ou Renard ne s'explique pas de façon satisfaisante et voici deux histoires qui pourraient mettre sur la voie.

A Soissons, à la fin du Moyen Age, les jeunes gens choisissaient au mois d'avril un « prince de la jeunesse ».

C'était, bien entendu, le plus beau de tous. Tout le temps de la fête, l'heureux élu couronné de fleurs était surnommé « le prince ». Il conservait ensuite le titre, souvent sa vie durant.

Pour Renard, surnom attesté dès 1258 à Dôle, faut-il chercher la clé dans la « fête du Renard » qui se déroulait sous Philippe le Bel (fin du XIII^e siècle, début du XIV^e siècle).

« Au milieu du clergé de la procession se tenait un renard habillé d'un surplis et coiffé de la tiare pontificale; des poules lui étaient présentées, sur lesquelles il se précipitait, naturellement. »

« Le roi s'amusa fort et dit que c'était l'image vraie du pape dont l'avidité égalait celle du renard. »

Celui qui avait incarné le renard avec talent était amicalement surnommé *Renard* ou *Lerenard*.

UNE PREUVE INESPÉRÉE, L'ÉTUDE DE L'I. N. S. E. E.

« Monsieur Dupont s'appelle Martin et son prénom est Jean ». Tel est le titre sybillin d'une étude publiée par l'I. N. S. E. E. dans *Économie et Statistiques*, n° 35 de juin 1972. La conclusion : Martin est le nom patronymique le plus fréquent et Jean le prénom le plus répandu.

En fait, l'Institut de la Statistique cherchait deux sortes d'informations :

— la longueur moyenne des noms de famille (réponse : 6,90 lettres avec un écart-type de 1,87);

— les noms de famille les plus fréquents.

Dans quel but? Pour savoir premièrement la taille minimum des « mémoires » à prévoir dans les ordinateurs traitant des fichiers de personnes volumineux, deuxièmement l'intérêt éventuel d'une codification des noms les plus courants.

Bien entendu, seule la seconde partie nous intéresse. En diffusant le résultat d'un sondage effectué sur 413 000 enfants nés en 1950 dans 56 départements, l'I. N. S. E. E. apporte de précieux éléments de contrôle et de vérification de nos propres travaux. Certes, les méthodes utilisées pour établir les fréquences des noms de famille sont totalement différentes. De plus, vingt ans séparent les années de référence. L'I. N. S. E. E. a travaillé sur les naissances de 1950, nous-mêmes sur les annuaires téléphoniques de 1970. L'un a opéré sur 56 départements, l'autre sur la France entière

Aussi notre surprise a été grande lorsque nous avons constaté que nous aboutissions, malgré des cheminements si divergents, à un classement et à des fréquences presque identiques.

Dans les deux études, *Martin* et *Bernard* arrivent en tête respectivement premier et second parmi les noms patronymiques les plus fréquents. Les 10 premiers noms sont les mêmes. Dans la liste des 100 noms les plus répandus publiée par l'I. N. S. E. E., 89 se retrouvent classés dans les 100 premiers noms de notre répertoire.

Poussant plus loin, nous avons calculé le coefficient de Spearman pour les 100 noms cités par l'I. N. S. E. E. et trouvé $R = 0,79$. La corrélation des classements est excellente. La probabilité que ce soit le fait du hasard est quasi-nulle (test t de Student).

Mais y avait-il la même concordance au niveau des fréquences? C'est-à-dire sur le nombre d'homonymes par rapport aux 413 000 enfants de l'étude I. N. S. E. E. ou par rapport au 50 millions de Français dans notre répertoire. Pour en juger voici les résultats obtenus (fréquence exprimée en 1/1 000) :

	<i>Tesnière</i>	<i>I. N. S. E. E.</i>	
1. <i>Martin</i> . . .	3,4	3,7	+9 %
2. <i>Bernard</i> . . .	2,0	2,0	0 %
3. <i>Durand</i> . . .	1,6	1,7	+6 %
4. <i>Dubois</i> . . .	1,6	1,7	+6 %

Pour les 100 premiers noms, le sondage de l'I. N. S. E. E. donne des chiffres supérieurs de 5 % en moyenne aux nôtres.

Ainsi, en dépit de la disparité des sources et des méthodes, il y a un accord étonnant des résultats. Le classement comme l'ordre de grandeur des fréquences des noms de famille sont analogues.

N'est-ce pas là la plus belle satisfaction que nous pouvions espérer? Nous en remercions chaleureusement l'équipe inconnue de l'I. N. S. E. E.