

# DÉFINITION, IDENTIFICATION ET LOCALISATION DE GÉOHAPAX DANS LE CORPUS DES PATRONYMES FRANÇAIS : COMPARAISON DE SOURCES ANCIENNES ET RÉCENTES

PASCAL CHAREILLE

MCF, UMR7323 Centre d'Études Supérieures de la Renaissance,  
CNRS-Université de Tours, France

PIERRE DARLU

DR, UMR7206 Eco-anthropologie et ethnobiologie, MNHN-  
CNRS-Université Denis Diderot, France

## **Definition, identification and localisation of *geohapax* in the corpus of French patronyms: A comparison of old and recent sources**

**Abstract:** Surnames have been widely used to describe the spatial structure of populations and the migrations between them. In France, as elsewhere, the distribution of documented surnames in the twentieth century reveals a strong micro-regional specificity and long-standing presence of some surnames, maintained despite the inevitable depletion of the initial stock of names and its regular renewal by migration. *Geohapax* are at present the subject of systematic researches by demographers and historians but also by geneticists who see them as possible vectors of genetic peculiarities.

For ancient periods and in some places, corpuses of surnames are available allowing to reveal the geographic distribution of surnames in the late Middle Ages, and to question the continuity, in the long term, of the initial stocks, and to track down the *geohapax* by comparing old and recent corpuses. In this paper, we illustrate and discuss, from various historical and contemporary data, the process of extinction of surnames with time and the various conditions allowing to select *geohapax*.

**Keywords:** surnames, *geohapax*, corpus analysis.

## **Une problématique**

Le nom de famille est réputé permettre de retracer l'histoire des personnes et des groupes et d'aider à en retrouver les origines géographiques. C'est ce qui est souvent affirmé. Pourtant, la démonstration rigoureuse de l'efficacité du nom pour atteindre ces objectifs n'a que rarement été fondée sur des confrontations avec des documents historiques. On sait que les noms peuvent disparaître au fil des générations, et donc l'histoire s'interrompre ; on sait également que la spécificité géographique originelle peut être brouillée par les mouvements migratoires et les origines devenir imprécises. L'objet du présent texte est de fournir des éléments de discussion sur cette problématique déjà ancienne.

Dès le milieu du XIXe siècle, la question de l'extinction des noms est posée par le mathématicien Irénée Jules Bienaymé (1845) et reprise ensuite par Francis Galton (Watson et Galton 1874)<sup>1</sup>. La motivation idéologique de ce dernier était de vérifier, à partir de l'exemple des noms de famille, si la crainte de voir les familles des classes supérieures disparaître peu à peu au profit des familles des classes inférieures démographiquement plus prolifiques était fondée. De son côté, Georges H. Darwin (1875) s'intéressa aux patronymes pour vérifier si les mariages isonymes, possiblement consanguins, avaient un effet délétère sur la constitution physique et mentale de leur descendance.

À partir du milieu des années 1960, les études patronymiques ont connu un nouvel essor<sup>2</sup>. En effet, la transmission des noms par voie patrilinéaire fournit aux généticiens des populations un modèle qui simule parfaitement la transmission d'un gène du chromosome Y. Sujet à mutations, ce pseudo-gène a en outre l'avantage de présenter un polymorphisme considérable<sup>3</sup>. En France comme ailleurs, les distributions spatiales des patronymes révèlent de fortes spécificités micro-régionales, ce qui fait des corpus anthroponymiques un matériau privilégié pour examiner la structuration spatiale des populations, évaluer la réalité des frontières géographique, historique, politique, culturelle ou linguistique, et éclairer, en creux, la question des migrations.

Les préoccupations des généticiens rejoignent celle des démographes. Pour les uns comme pour les autres, la question des origines est devenue essentielle, qu'il s'agisse de rechercher le premier porteur d'une mutation génétique aux conséquences possiblement dramatique (cf. par ex. les recherches faites sur la maladie de Rendu-Osler dans la région du Saguenay-Lac Saint-Jean au Canada (Heyer 1991)), ou d'établir l'origine géographique d'un patronyme pour tenter d'éclairer, dans la longue durée, l'histoire d'un groupe d'individus apparentés partageant le même patronyme. De plus, une personne qui porterait un nom géographiquement bien localisé serait susceptible de fournir une information génétique sur une période révolue, celle de l'origine même du nom (Manni et al. 2005).

Les possibles dérivées d'une "quête des origines" revendiquée par certains programmes d'étude génétique des populations<sup>4</sup> ont été soulignées (Bourgain et Darlu 2013) et l'objectif n'est pas ici de revenir sur ce débat. Pour autant, les présupposés implicites qui fondent l'approche des généticiens quand ils (re)cherchent des noms vecteurs de possibles singularités génétiques méritent d'être examinés.

## Des corpus

Pour la France du XXe siècle, le « fichier des noms patronymiques » (1891–1990), produit par l'INSEE et informatisé dans les années 70, fournit la matière classiquement

<sup>1</sup> La question, complexe, de l'appauvrissement a été reprise ultérieurement par Lotka (1920, 1931) et, plus récemment, par Brouard (1989) et Dupâquier (1992).

<sup>2</sup> À l'origine de cet essor, les travaux de Crow et Mange (1965) sur la question de la relation quantitative entre isonymie et *inbreeding* (sur cette question, cf. également Crow, 1980).

<sup>3</sup> Pour la plupart des patronymes, les variantes graphiques, phonétiques et linguistiques sont nombreuses, en lien avec les problèmes de transcription, de transmission, ou de traduction. Le patronyme est donc sujet à des mutations transmissibles. Il est en outre sélectivement neutre.

<sup>4</sup> Ainsi le projet "Genographic", qui entend retracer l'épopée migratoire de l'espèce humaine. Sur ce projet, cf. Darlu (2008).

utilisée pour examiner les questions relatives à la structuration spatiale des populations et son évolution au cours du temps, ou bien à l'appauvrissement et au renouvellement des stocks anthroponymiques interprétés en termes de flux migratoires. Il rassemble les noms, localisés à la commune, recueillis dans les actes de naissance. Le corpus est scindé en 4 périodes chronologiques de 25 ans (1891–1915, 1916–1940, 1941–1965 et 1966–1990) et couvre notamment la France métropolitaine<sup>5</sup>.

On ne dispose pas, pour les périodes anciennes, de corpus informatisés équivalents à ce fichier. Il est toutefois possible de s'appuyer sur le dépouillement de registres paroissiaux et d'état-civil, parfois également sur celui de documents fonciers et/ou fiscaux antérieurs qui ne permettent certes pas de prétendre à l'exhaustivité – ils ne recensent jamais tous les individus d'une population et se limitent en général aux propriétaires ou aux « chefs de feux » –, mais qui autorisent des études locales fouillées inscrites dans la très longue durée. C'est le cas notamment en Savoie, dans les Cévennes et dans le Dijonnais où nous disposons de corpus conséquents<sup>6</sup>.

La constitution et l'exploitation de ces stocks anthroponymiques anciens restent délicates, notamment pour les corpus médiévaux : *hug[ue]s boy laigue* désigné dans un document dijonnais de 1376 est-il ce *hug[ue]s boilleaux* recensé un an plus tard dans la même rue ? Sûrement, mais que dire de son patronyme, à supposer que son surnom puisse effectivement être considéré comme tel ? Les exemples de ce type foisonnent et il n'est pas toujours facile de trancher. La transcription des noms, à une époque où l'usage de l'écrit n'était pas encore généralisé et l'orthographe fluctuante, est largement phonétique. La différenciation des *Fabre, Favre, Febvre, Fèvre, Lefebvre, Lefèvre, Lefébure*, etc., ou celle des *Gauthier, Gautier, Galtier, Vaultier*, etc., acquise dans les répertoires patronymiques contemporains, n'est pas nécessairement pertinente pour le Moyen Âge. Les critères de différenciation de possibles variantes sont donc multiples (orthographiques, linguistiques, phonétiques, etc.) et décider de recenser ensemble ou séparément telle ou telle forme – c'est l'objet de la lemmatisation – est évidemment une étape cruciale car lourde de conséquences au moment de l'analyse<sup>7</sup>.

<sup>5</sup> Bien que ce fichier, informatisé dans les années 70, soit le plus exhaustif disponible à ce jour, il demeure incomplet : certaines communes manquent ou sont mal référencées ; les porteurs nés en Alsace ou en Lorraine pendant les périodes où ces régions étaient allemandes ne sont pas comptabilisés et les modifications de patronymes intervenues lorsque ces régions sont redevenues françaises ne sont pas précisées ; enfin et surtout, les personnes nées avant 1946 et mortes avant 1972 n'y figurent pas. Pour la première période P1 (=1891–1915), le taux de couverture des naissances est probablement inférieur de 25% à ce qu'il est en P2 (1916–1940), ce qui pose évidemment un certain nombre de problèmes d'exploitation. En outre, très ponctuellement, le fichier s'avère fautif avec des (re)transcriptions de noms assurément incorrectes.

<sup>6</sup> Pour une présentation détaillée de ces corpus, que nous avons déjà eu l'occasion d'exploiter, cf. pour la Savoie : Darlu et al. (2011) ; pour les Cévennes : Darlu et al. (1996) ; pour le Dijonnais : Beck et Chareille (exploitation en cours).

<sup>7</sup> Retenir toutes les variantes conduit à accroître la variabilité (certains noms très proches correspondent assurément à de mêmes lignées familiales), mais regrouper trop systématiquement conduit à réduire artificiellement la diversité patronymique (en associant des noms correspondant à des lignées familiales possiblement non apparentées). En pratique, il est souvent intéressant d'envisager différents niveaux de lemmatisation (regroupement des seules graphies et/ou variantes

Pour la Savoie, les données utilisées ici proviennent du dépouillement d'un cadastre établi dans les années 1728–1738, et de celui d'actes de baptême et de naissance d'un ensemble de 27 communes, couvrant une période qui s'étale du début du XVIII<sup>e</sup> siècle au milieu du XX<sup>e</sup> siècle, découpée pour l'occasion en 4 phases chronologiques (S1=1710–1729 ; S2=1810–1829 ; S3=1891–1915 ; S4=1916–1940).

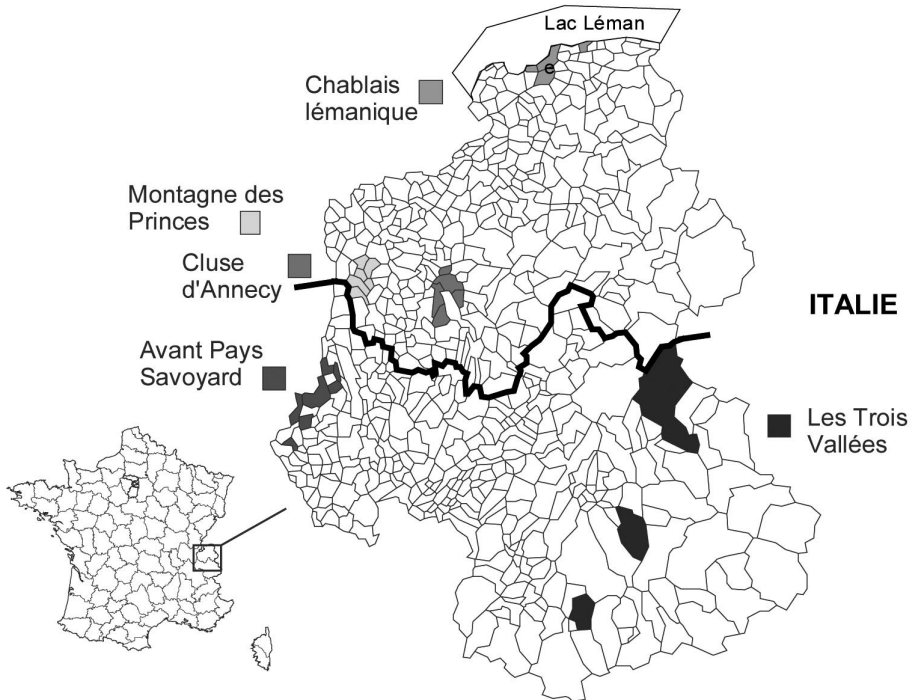


Figure 1. Localisation des cinq micro-régions en Pays savoyard (d'après Darlu et al. 2011).

Pour les Cévennes, les données concernent 23 cantons situés dans l'espace géographique traditionnellement cévenol, mais répartis sur 4 départements différents, l'Aveyron, le Gard, l'Hérault et la Lozère. Les noms de famille sont ceux relevés dans le registre de naissances des tables décennales pour la période C1=1843–1862, tandis que pour les deux périodes suivantes (C2=1891–1915 ; C3=1916–1940) les données sont extraites du fichier de l'INSEE.

mineures pour une "lemmatisation douce" ; regroupement, sous une même forme souche, de toutes les formes possiblement apparentées pour une "lemmatisation forte").

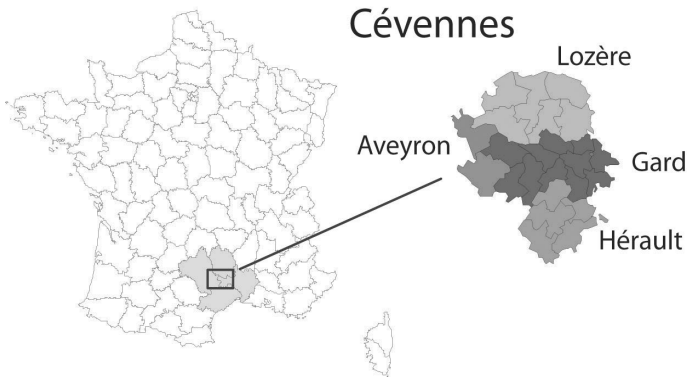


Figure 2. Localisation des 23 cantons des Cévennes répartis dans 4 départements différents.

Pour le Dijonnais, le corpus résulte du dépouillement exhaustif des *cherches des feux* du baillage de Dijon pour les années 1376, 1424 et 1470 et de la *visitation des feux* de 1610<sup>8</sup>. Bien connues des historiens du peuplement, ces *cherches*, destinées à asseoir l'assiette des aides générales accordées au duc par ses États, couvrent l'ensemble du duché et donnent les listes nominales des chefs des foyers réels imposables recensés par localité. Ce sont au total 35500 occurrences anthroponymiques qui ont été collectées, distribuées dans 288 localités.



Figure 3. Localisation des localités du Dijonnais qui ont permis d'élaborer le corpus anthroponymique – figure provisoire)

<sup>8</sup> Respectivement Archives Départementales de la Côte d'Or (=ADCO) B 11571 & B 11572 pour la *cherche* de 1376, ADCO B 11582 pour celle de 1424, ADCO B 11590 pour celle de 1470, et ADCO C 4733 pour la *visitation* de 1610.

### 1. L'extinction des noms de famille

#### a. Approche théorique

• La probabilité d'extinction d'un nom

Henri William Watson et Francis Galton (1875) ont établi les formules démontrant comment les noms de familles disparaissent, retrouvant, probablement sans les connaître, les mêmes formulations que celles de Bienyamé (1845). Considérons le cas du nom porté par un père. La probabilité  $x_1$  de voir disparaître son nom à la génération suivante est égale à  $p_0$  s'il n'a pas de descendance. Soit  $x_1 = p_0$

Si ce père n'a qu'un seul fils à qui transmettre son nom, ce dernier aura, à son tour, une probabilité,  $x_2$ , de ne pas transmettre son nom à sa descendance avec une probabilité égale à  $p_0 x_1^0 = p_0$  s'il n'a pas de fils, une probabilité  $p_1 x_1^1$  s'il n'a qu'un fils, une probabilité  $p_2 x_1^2$  s'il en a deux, ..., une probabilité  $p_s x_1^s$  s'il en a s,  $p_i$  étant la probabilité pour un père d'avoir  $i$  fils. D'où la formule de récurrence suivante :

$$x_2 = p_0 x_1^0 + p_1 x_1^1 + p_2 x_1^2 + \dots + p_s x_1^s + \dots + p_q x_1^q$$

$$x_n = p_0 x_{n-1}^0 + p_1 x_{n-1}^1 + p_2 x_{n-1}^2 + \dots + p_s x_{n-1}^s + \dots + p_q x_{n-1}^q$$

La vitesse d'extinction d'un nom dépend donc de la distribution du nombre de fils dans les familles. L'hypothèse retenue par Galton était que la probabilité de ne pas avoir de fils, d'en avoir un ou d'en avoir deux était la même ( $p_0 = p_1 = p_2 = 1/3$ ), ce qui est évidemment bien simplifié et l'a conduit à considérer que l'extinction des noms était inéluctable. En fait, la disparition, en probabilité, n'est pas certaine puisqu'elle dépend des données démographiques, comme le montre la figure 4 où la formule précédente est appliquée avec les données de Lotka (1931), de Brouard (1989) et de Toulemon (2001). De fait, dans ces situations démographiquement réelles, le nom de famille ne disparaît pas, même après de nombreuses générations.

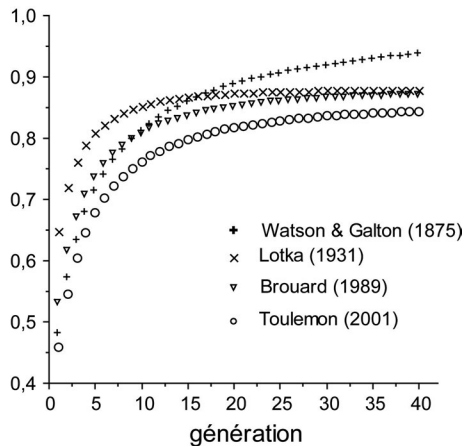


Figure 4. Probabilité d'extinction d'un nom de famille en fonction de conditions démographiques initiales différentes. Les conditions imposées par Watson et Galton sont :  $p_0 = p_1 = p_2 = 1/3$ . Pour les autres conditions démographiques, voir Brouard (1989) et Lotka (1931).

• *L'évolution de la fréquence d'un nom*

Les exemples de la figure 4 montrent le destin, en probabilité, d'un nom porté par une seule personne. En réalité, l'avenir d'un nom de famille dépend également de sa fréquence dans la population. Lorsque sa fréquence est faible, ses fluctuations aléatoires au fil des générations, selon le nombre de fils à qui il est transmis, peuvent amener à une extinction. En revanche, si la fréquence est élevée, ces fluctuations aléatoires seront faibles et l'extinction fort peu probable. Ce phénomène de variation est totalement comparable à ce que les généticiens ont largement décrit sous le terme de dérive génétique. La figure 5 montre deux situations, celle du nom *Echeverry* dont la fréquence est faible, mais dans une population importante (celle des Pyrénées Atlantiques), et celle du nom *Huboux* dont la fréquence est élevée, mais dans une population faible d'une petite commune de l'Ain. Ces modèles restent théoriques, car la réalité est certainement plus complexe. Ils ont donc besoin d'être confrontés à des documents historiques qui permettent de vérifier leur validité.

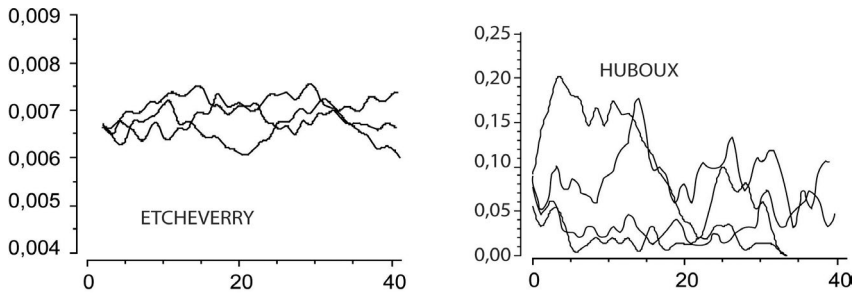


Figure 5. Évolutions aléatoires des noms *Echeverry* (faible fréquence dans une large population) et *Huboux* (forte fréquence dans une petite population), sur 40 générations. Chaque courbe correspond à une simulation.

**b. Observations**

Qu'en est-il en effet de la confrontation entre théorie et réalité ? Quelle est la véritable érosion des patronymes ? Les exemples suivant pris dans divers documents à des échelles historiques et géographiques diversifiées permettent de s'en faire une idée.

• *corpus INSEE*

L'érosion du nombre de noms à l'échelle départementale en France peut s'apprécier en comparant le nombre de noms qui sont attestés dans le corpus INSEE pendant la période P1 (1891–1915), mais qui ne se retrouvent pas dans le même département à la période suivante (1916–1940). Si l'on ne tient pas compte des hapax (noms attestés par une seule naissance en P1 dans un département donné) la diminution du nombre de nom est, sur une moyenne de 90 départements, de 18,8%, avec la plus faible diminution dans le département du Nord et des Côtes-d'Armor (13%), et les plus fortes dans les départements d'Alsace et Moselle (plus de 27%). Comme cette statistique est proposée par département, il n'est pas impossible que des noms disparaissent d'un département à une certaine période (P2) pour se retrouver dans un autre département à la période suivante.



Néanmoins, compte tenu du faible taux de migrations observé en France en ces périodes, on peut considérer que ces pourcentages donnent une idée approximative d'une érosion des noms qui s'avère importante.

• *Savoie*

Le corpus savoyard livre au total 9414 patronymes distincts sur l'ensemble des périodes étudiées, mais 609 seulement sont continûment attestés entre 1710 et 1940. Localement, c'est-à-dire en observant la situation au sein de chacune des 5 micro-régions identifiées, on constate une déperdition importante entre ces deux périodes, entre 61% et 75%, une variation qui s'explique à la fois par la taille réduite de la population rurale de Montagne des Princes, où l'exode rural fut important au XIXe siècle, et par la taille plus élevée des populations dans les régions plus stables et plus urbaines du Chablais lémanique et de la Cluse d'Annecy (avec les villes d'Evian et Annecy)

**Tableau 1. Proportion de patronymes perdus entre S1 (1710–1729) et S4 (1916–1940).**

micro-région	%
Montagne des Princes	70.7
Chablais lémanique	58.7
Cluse d'Annecy	60.9
Avant Pays savoyard	75.1
Les Trois vallées	69.5

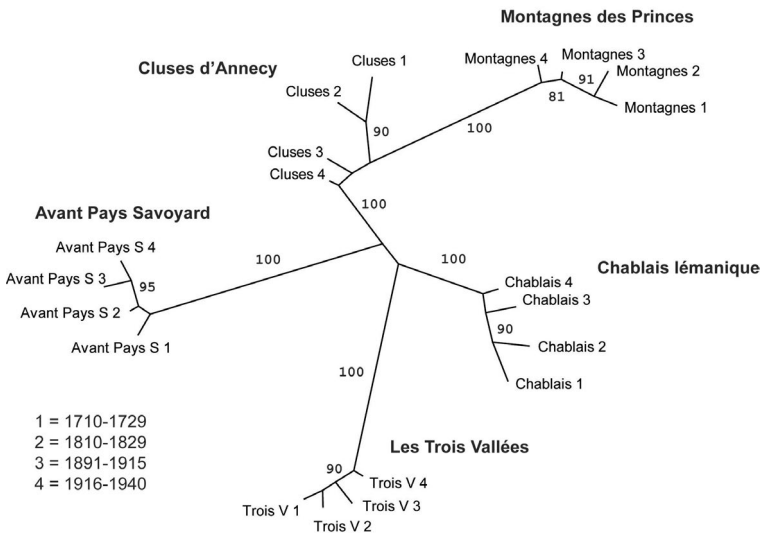


Figure 6. Représentation des relations patronymiques entre micro-régions de Savoie, en fonction des périodes d'observation (d'après Darlu et al. 2011). La période la plus récente est celle montrant le moins de différenciation entre régions (à l'exception de l'Avant Pays Savoyard en proximité avec la région lyonnaise dont elle reçoit de nouveaux noms portés par des migrants).



Malgré cette déperdition de patronymes, la signature patronymique de chaque micro-région se maintient avec le temps. Si l'on représente les distances patronymiques entre ces micro-régions et entre différentes périodes sous forme d'un arbre, il apparaît que les différences patronymiques entre elles se maintiennent solidement au cours du temps. Un stock de patronymes bien représentés et bien différenciés entre régions explique cette stabilité chronologique. Les noms qui disparaissent sont donc préférentiellement ceux ayant des fréquences relativement faibles si bien que leur disparition n'affecte que peu la structure (Darlu et al. 2011).

#### • Cévennes

Les Cévennes fournissent une autre évaluation de l'extinction des noms de famille dans un espace rural où les populations sont de taille réduite et à une époque où l'exode rural est important. Le tableau 2 montre que les communes des Cévennes appartenant administrativement à l'Aveyron sont celles qui perdent le moins de patronymes entre 1843–1862 et 1916–1940, alors que les pertes notées dans la partie de la Lozère et du Gard sont plus marquées. Le fait que ces cantons aient davantage été marqués par les migrations vers la région de Nîmes et Alès, pourvoyeuse d'emplois, est l'explication la plus vraisemblable. Cependant, à la différence de la Savoie, la différenciation patronymique entre ces cantons se modifie avec le temps, comme le montre la figure 7 (d'après Darlu et al. 1996).

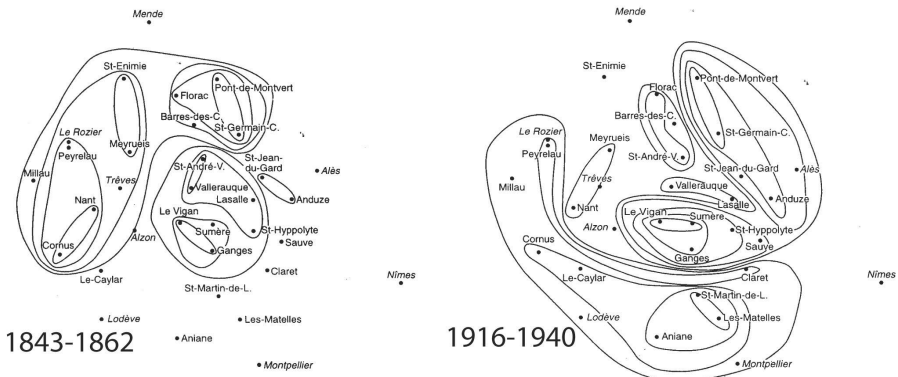


Figure 7. Regroupements hiérarchiques des 23 localités des Cévennes en fonction de la ressemblance entre leurs profils patronymiques en 1843–1862 et 1916–1940 (d'après Darlu et al. 1996).

Les ensembles déterminent, de manière emboîtée, les cantons dont l'association est statistiquement validée par *bootstrap*. La comparaison entre les deux périodes indique comment les profils patronymiques se sont modifiés au point de bouleverser les affinités entre cantons. Ce phénomène s'explique par le changement dans les attractions qu'exercent les grandes métropoles encerclant les Cévennes, depuis le milieu du XIXe siècle où les Causses et le nord du Mont Aigoual entretenaient des relations étroites, jusqu'au XXe siècle, avec la perte progressive des influences de Mende au nord et de Millau à l'est, au profit de Montpellier au sud et Alès et Nîmes à l'est, en raison également d'une ouverture des communications au cœur des Cévennes vers l'est.

**Tableau 2. Proportion de patronymes perdus entre C1 et C3.**

département	%
Aveyron	48.0
Gard	59.9
Hérault	52.7
Lozère	72.4

- *Dijonnais*

Sur les 288 localités distinctes signalées dans le fonds dijonnais, 113 seulement sont documentées en 1376, 1424, 1470 et 1610. Mais 199 font l'objet d'une recension à 3 ou 4 reprises consécutives, correspondant à 166 communes du maillage administratif contemporain. Ces 166 communes, toutes localisées dans l'actuel département de la Côte d'Or, signalent au total 33007 individus ; et un peu plus de 32000 d'entre eux portent un surnom ou "patronyme" (ce qui représente 90% environ des 35500 occurrences du corpus anthroponymique initial).

**Tableau 3. Nombre d'individus et nombre de surnoms en Dijonnais pour chaque "période".**

date	nb individus	nb surnoms
1376 = D1	6491	3808
1424 = D2	6658	3237
1470 = D3	9654	3770
1610 = D4	9239	3354
ensemble de la période	32042	10059

Le corpus dijonnais livre au total (sur l'ensemble de la période) environ 10000 "patronymes" distincts (entre 9000 et 11000 selon le niveau de lemmatisation retenu), mais ce stock patronymique semble connaître une histoire chronologique mouvementée. 60% des ces noms sont en effet des hapax du corpus : ils ne sont portés que par un seul individu et ne sont donc attestés qu'à une seule date. Cette proportion, élevée, est presque deux fois plus forte que celle observée dans le fichier INSEE (Degioanni et al. 1997). Le mode de constitution du corpus dijonnais (seules les communes continûment documentées ont été sélectionnées) et la nature des sources exploitées (absence d'exhaustivité) expliquent sans doute en partie cette situation.

Si l'on exclut ces hapax, dont le passage dans l'histoire anthroponymique des communes concernées est assurément bref, le profil chronologique des quelques 3739 patronymes restants est erratique :

**Tableau 4. Profil chronologique des 3739 "patronymes" dijonnais portés par au moins deux individus.**

profil chronologique	nb surnoms	nb individus	% surnoms	% individus
D1	255	624	6,8%	2,4%

D1D2	277	899	7,4%	3,5%
D1D2D3	301	2166	8,1%	8,4%
D1D2D3D4	503	12421	13,5%	48,3%
D1D2D4	47	299	1,3%	1,2%
D1D3	158	495	4,2%	1,9%
D1D3D4	94	711	2,5%	2,8%
D1D4	84	275	2,2%	1,1%
D2	186	432	5,0%	1,7%
D2D3	380	1349	10,2%	5,2%
D2D3D4	235	2139	6,3%	8,3%
D2D4	73	242	2,0%	0,9%
D3	392	976	10,5%	3,8%
D3D4	275	1327	7,4%	5,2%
D4	479	1367	12,8%	5,3%
Total	3739	25722	100,0%	100%

Seuls 503 noms – donc moins de 15% du stock constitué – traversent l’ensemble de la période (près de 20% cependant si l’on considère ceux qui sont présents en D1 et en D4 et qui connaissent possiblement une éclipse entre D1 et D4). Mais sur les 100 noms les plus fréquents de ce corpus, 4 seulement ne sont pas parmi ceux que l’on retrouve à chaque date (2 sont manquants en D1 et 2 en D2). Sans surprise, donc, les noms fréquents sont d’abord ceux qui s’inscrivent dans la durée. Et réciproquement, ceux qui sont continuellement attestés sont parmi les plus fréquents. Sur les 503 noms de ce type (présents en D1, D2, D3 et D4), 396 (soit près de 80%) comptent plus de 10 porteurs sur l’ensemble de la période (alors que les noms portés au moins 10 fois ne représentent que 6% du corpus).

## 2. Les géohapax

L’identification des patronymes localement enracinés fait actuellement l’objet de recherches systématiques. Pour les historiens et les démographes, la question s’insère dans un débat plus large quant à la réalité et à la force de la sédentarité et de la mobilité des populations rurales du passé<sup>9</sup>. Pour les généticiens, ces patronymes sont de possibles vecteurs de singularités génétiques. Leur repérage systématique pourrait permettre de pister les porteurs actuels de ces noms et de rechercher dans leur ADN de possibles singularités que la dérive génétique n’aurait pas fait disparaître.

Pour les patronymes français les plus fréquents, l’origine géographique multiple (ou ubiquité) est la norme<sup>10</sup>, mais ce caractère est plus ou moins prononcé selon les patronymes.

La carte de répartition des *Gauthier* témoigne d’une dispersion géographique forte et il est difficile d’espérer repérer un « foyer possiblement originel » pour ce nom.

<sup>9</sup> Sur ce débat, cf. notamment Croix (1999) ; Dupâquier (2002) ; Poussou (2002).

<sup>10</sup> Le constat vaut également pour d’autres espaces géographiques.

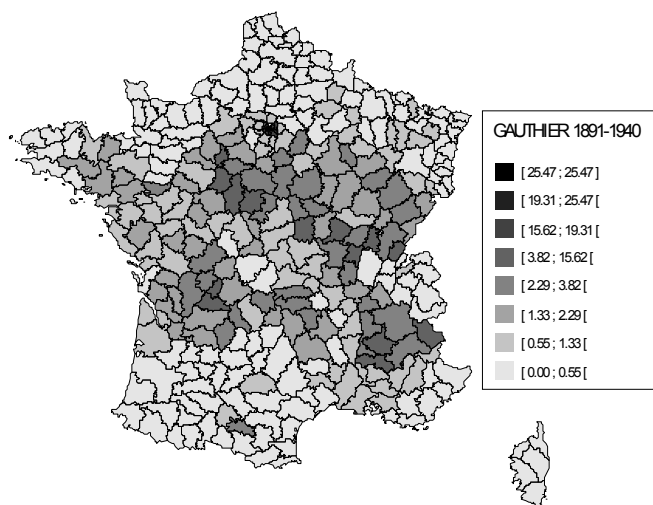


Figure 8. Carte de répartition (échelle cantonale) du patronyme *Gauthier* (corpus INSEE, 1891–1940).

D’autres patronymes, en revanche, raisonnablement fréquents dans le corpus INSEE, semblent plus nettement localisés. Ainsi *Marie*, *Gautier*, *Hue*, *Galtier* et *Fabre*.

Pour *Galtier*, variante de *Gauthier*, l’origine languedocienne ne semble guère faire de doute. Pour *Gautier*, autre variante du même nom, les choses semblent moins limpides. On peut imaginer divers scénarii : une origine unique au Nord-Ouest avec une migration au Sud-Est – l’inverse semble moins probable mais ne peut être exclu –, ou deux foyers originels distincts, sans qu’il soit possible, à l’examen de ces seules cartes, de trancher. Pour *Fabre* (nom dont on a signalé précédemment quelques variantes dans le corpus patronymique contemporain), l’arc méditerranéen se dessine nettement. Pour *Marie* et pour *Hue*, l’origine normande est bien établie, mais avec une dispersion spatiale différentielle. À l’exception possible de *Gautier*, il s’agit donc bien là de patronymes que le sens commun définirait volontiers comme « régionaux ».

Ces noms sont-ils pour autant des patronymes localement enracinés ? Et cet enracinement apparent est-il récent ou au contraire ancien ? Pour répondre à ces questions, nous proposons d’introduire le concept de *géohapax*, moins ambiguë que celui de patronyme monophylétique qui suppose une origine (génétique) commune (qui ne peut évidemment pas être établie sur la base des seuls patronymes), et de considérer comme tel tout nom qui présente à la fois une fréquence significative et qui se trouve attesté dans une et une seule entité spatiale. Deux observations : 1. le concept de “fréquence significative” suppose la définition d’un seuil qu’il convient de discuter ; 2. imposer la présence “dans une et une seule entité spatiale” est une contrainte forte, qui a l’avantage néanmoins d’être indépendante – dans sa définition – du découpage spatial retenu.

On observera qu’aucun des noms précédents, à quelque échelle d’analyse du maillage administratif français que ce soit (région, canton, commune), n’entre dans cette catégorie.

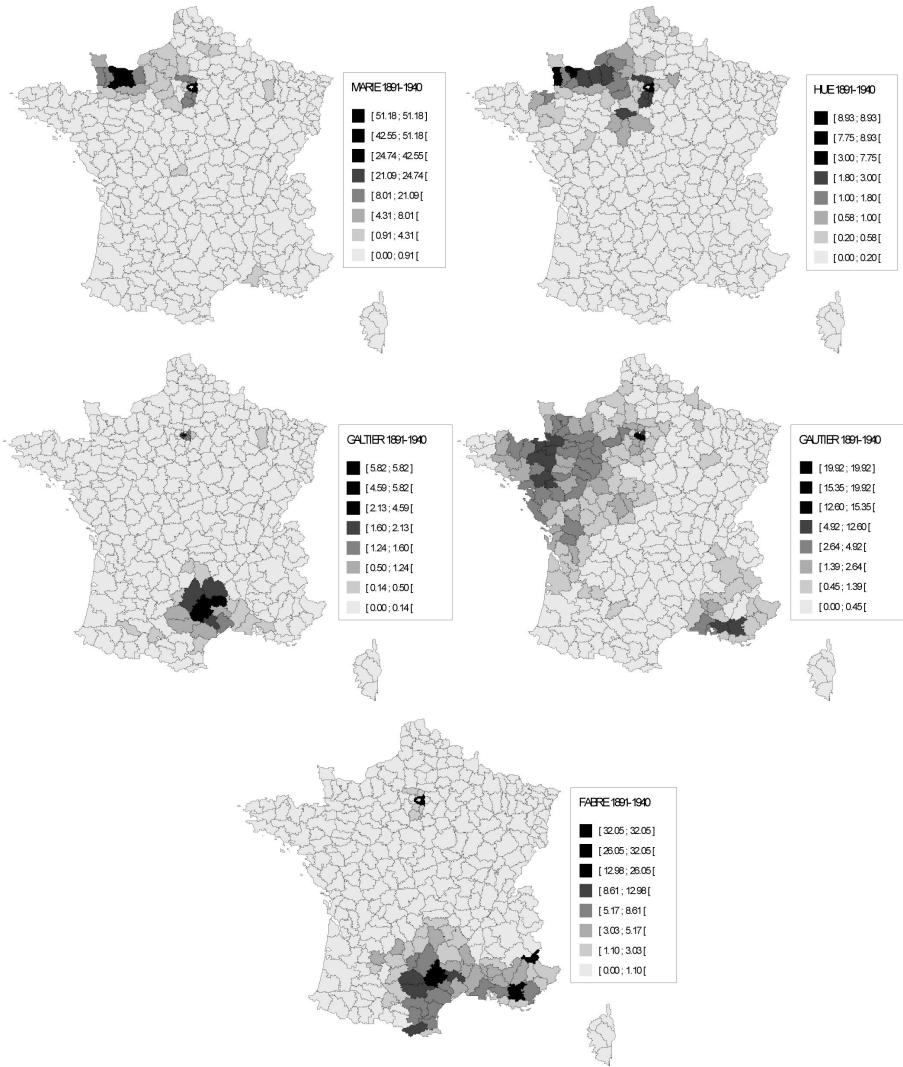


Figure 9. Carte de répartition (échelle cantonale) des patronymes *Marie*, *Hue*, *Galtier*, *Gautier* et *Fabre* (corpus INSEE, 1891–1940).

Cette définition de *géohapax* est proposée par analogie avec celle donnée à un patronyme « hapax » qui est un patronyme porté par une seule personne. La simplicité de cette définition n'est qu'apparente. En effet, le nombre des *géohapax* va dépendre d'au moins deux paramètres. Le premier est l'étendue de l'aire géographique considérée. Entre une petite commune et un vaste département ou une région entière, le nombre et les caractéristiques des *géohapax* peuvent varier considérablement. Le deuxième paramètre est le nombre de patronymes (le seuil) à partir duquel on peut estimer être en présence d'un

*géohapax*. Disons d’abord que les hapax sont nécessairement des *géohapax*, puisque, étant uniques, ils ne sont rattachés qu’à un seul espace géographique. Mais il est possible d’aller plus loin et de considérer comme *géohapax* les patronymes qui sont représentés par 2, 10, 100, 200, ou  $n$  personnes à la condition qu’elles soient toutes présentes exclusivement dans une seule et même aire géographique. Selon le seuil choisi pour  $n$ , faible ou fort, le nombre de *géohapax* sera, inversement, élevé ou faible.

• *Savoie*

Le corpus des données patronymiques de la Savoie permet d’illustrer la complexité de ce qu’est un *géohapax*. L’idée est la suivante : parmi les 609 patronymes persistant dans le corpus savoyard sur toutes les périodes, de 1710 à 1940, combien sont attestés dans un seul département, combien dans 2, 3 etc. La figure 10 montre que près de 14% des patronymes peuvent être considérés comme des *géohapax* présents dans un et un seul département (La Savoie, 73, ou la Haute-Savoie, 74). C’est le cas des *Decorzent, Decisier ... Bunoz* par exemple ; elle montre aussi que près de 40% de ces patronymes ne se retrouvent que dans 6 départements au plus. Les patronymes les plus communs comme *Durand, Bertrand, Marin*, etc. ne peuvent être considérés comme *géohapax*, dans la mesure où ils sont attestés dans les 90 départements de la France métropolitaine.

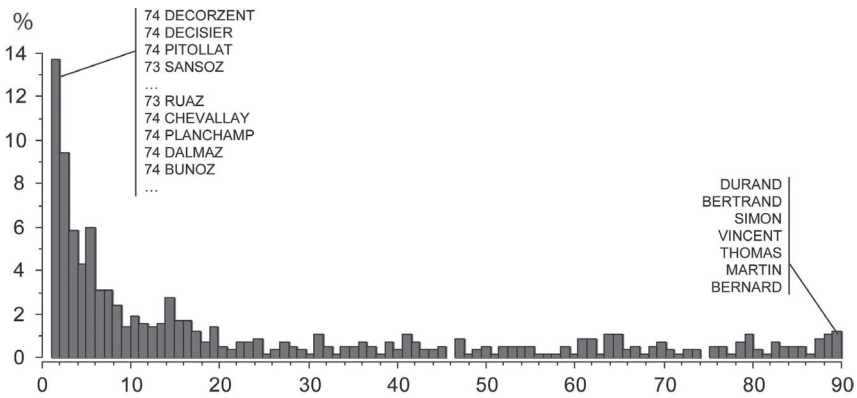


Figure 10. Pourcentage de patronymes, parmi les 609 patronymes attestés dans le corpus de Savoie entre 1710 et 1940, qui sont présents dans un seul département ou simultanément dans deux, trois, ..., 90 départements. On peut considérer qu’il a 14% de *géohapax* départementaux parmi l’ensemble de ces patronymes.

• *Dijonnais*

À une autre échelle, celle des communes du Dijonnais médiéval, les *géohapax* sont plus nombreux encore puisque plus de 75% (des quelques 10000 “patronymes” du corpus) entre dans cette catégorie.

Il est vrai cependant que si l’on exclut les véritables hapax du corpus, cette proportion tombe à 35% (et celle des noms attestés dans 2 communes passe, elle, à près de 30%). L’allure globale de la distribution est nettement affectée (les hapax conduisant de fait à une sur-représentation des *géohapax*).

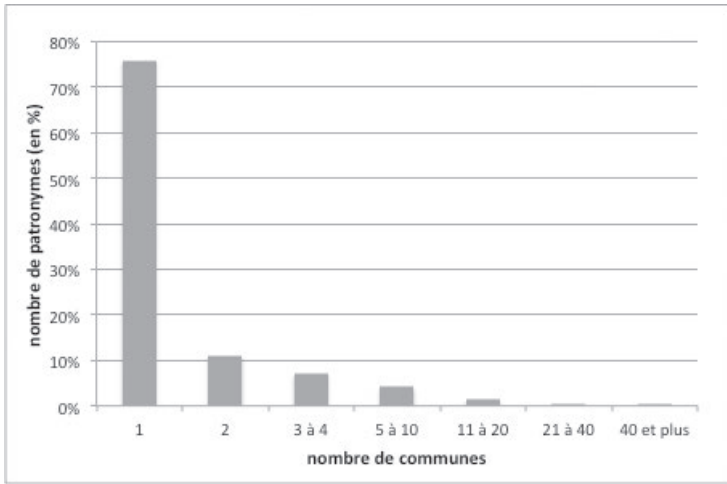


Figure 11. Pourcentage des "patronymes", parmi les 10000 du corpus médiéval dijonnais, attestés dans 1, 2, 3 à 4, 5 à 10, etc. communes.

• *corpus INSEE*

Le corpus patronymique de l'INSEE, malgré ses imperfections (voir note 5) permet d'illustrer le concept de *géohapax*. Ainsi, si on retient seulement un sous-corpus formé des 100 patronymes les plus fréquents dans chacun des 90 départements, soit 3368 patronymes, on remarque qu'un très fort pourcentage de ces patronymes ne sont localisés que dans un seul département (68%), 14% dans 2 départements et un nombre minime dans plus de 10 départements. Cette statistique montre la spécificité régionale des noms, même quand le critère de sélection cherche à retenir les noms les plus fréquents.

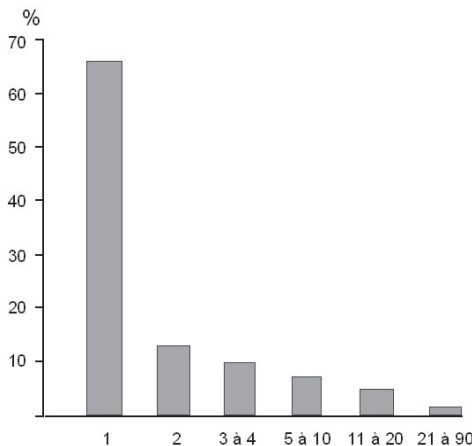


Figure 12. Pourcentage de patronymes, parmi les 3368 patronymes correspondant aux 100 plus fréquents dans chacun des 90 départements qui sont également attestés dans 1, 2, 3 à 4, 5 à 10 ... départements (d'après Degioanni et al. 1997).



Si la sélection concerne les patronymes qui correspondent, sur la période 1891–1940, à au moins 200 naissances, un seuil estimé sur l’ensemble des départements, la distribution du nombre de départements où ils sont attestés avec au moins 10 naissances mérite d’être représentée. Elle présente un maximum à 6 (départements) pour décroître ensuite rapidement.

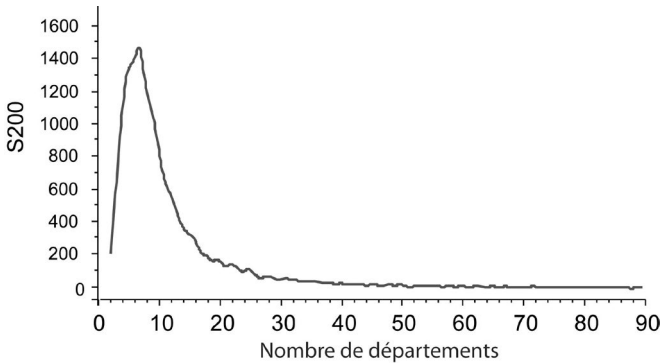


Figure 13. Nombre de patronymes correspondant à au moins 200 naissances entre 1891 et 1940 présents simultanément dans Nb départements (correspondant à au moins 10 naissances).

Comme on l’a dit précédemment, le nombre de *géohapax* est dépendant d’un paramètre de seuil, le nombre de porteurs du patronyme, qui agit comme un filtre plus ou moins strigent. Ainsi, dans le corpus de l’INSEE (période 1891–1940), peut-on retenir un patronyme comme étant un *géohapax* s’il n’y a qu’une seule naissance enregistré sous ce nom ( $n=1$ ) ou s’il y en a  $n=10, 50, 100, 200\dots$ . Comme on la vut, si  $n=1$ , les *géohapax* sont également des hapax. En revanche plus la valeur de  $n$  est élevée, moins le nombre de *géohapax* le sera, comme le montre la figure 14.

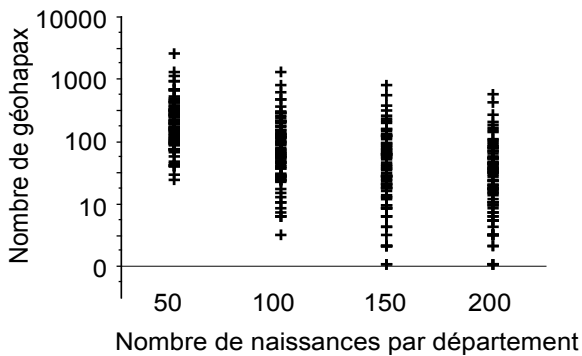


Figure 14. Nombre de *géohapax* (échelle logarithmique) par département selon la valeur de seuil retenue pour le critère du nombre minimum de naissances portant le même patronyme (de 50 à 200). Chaque croix représente la valeur d’un département.

En prenant le critère de  $n=100$ , on s'aperçoit que la distribution du nombre de *géohapax*, ou leur fréquence (figure 15), varie de manière importante d'un département à l'autre dans l'espace de la France métropolitaine. Les valeurs les plus élevées se situent essentiellement dans les zones à particularisme linguistique, comme la Bretagne, le Nord, l'Alsace, le Roussillon ou le Pays Basque. Pour ces régions, les *géohapax* peuvent être considérés comme de véritables marqueurs d'origine.

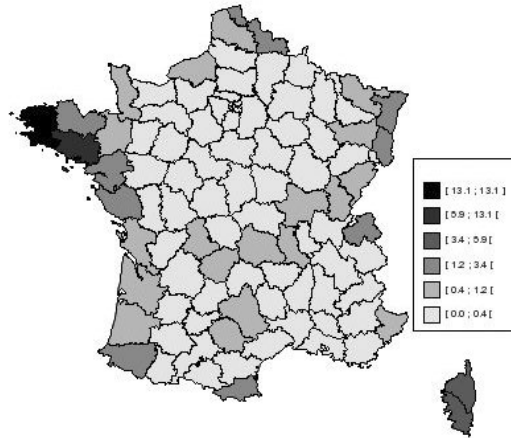


Figure 15. Distribution départementale du nombre relatif des *géohapax* définis par au moins 100 naissances (corpus INSEE, 1891–1940).

• *pérennité dans le temps des géohapax dijonnais*

Une question demeure : que sont devenus les noms continûment attestés en Dijonnais de la fin du XIV<sup>e</sup> s. au début du XVII<sup>e</sup> s. ? En les recherchant dans le fichier de l'INSEE, on constate qu'entre le quart et le tiers – selon le niveau de lemmatisation retenu – d'entre eux sont encore présents au début du XX<sup>e</sup> (en P1 et/ou P2) dans le département de la Côte d'Or. Et si l'on considère l'ensemble du fichier INSEE, donc en étendant la recherche à l'ensemble de la France métropolitaine, ce sont un peu plus de 50% des surnoms médiévaux dijonnais qui ont survécu.

Sans surprise, et conformément à la théorie, cette survie est plus forte encore pour les patronymes qui étaient les plus fréquents à l'époque médiévale : les 25 premiers du palmarès (sur l'ensemble de la période 1376–1610) sont tous attestés en Côte d'Or entre 1891 et 1940. Et les deux palmarès sont étonnamment proches.

Tableau 5. Palmarès comparé des 25 surnoms les plus fréquents en Dijonnais (1376–1610) et en Côte d'Or (1916–1940). En gras, les surnoms communs aux deux palmarès.

rang	corpus 1376–1610	corpus INSEE (P2)
1	<b>Rouyer</b>	<b>Martin</b>
2	<b>Simon</b>	<b>Bernard</b>
3	<b>Fevre</b>	<b>Garnier</b>

4	Mugnier	<b>Petit</b>
5	Barbier	<b>Moreau</b>
6	<b>Martin</b>	Perrin
7	Francois	<b>Thomas</b>
8	Maire	Roux
9	<b>Moreau</b>	<b>Roy</b>
10	Tixier	Renard
11	<b>Robert</b>	<b>Royer</b>
12	<b>Bernard</b>	<b>Perrot</b>
13	<b>Roy</b>	Masson
14	<b>Petit</b>	<b>Fournier</b>
15	<b>Fournier</b>	Richard
16	<b>Perrot</b>	<b>Robert</b>
17	<b>Garnier</b>	Noirot
18	<b>Clerc</b>	<b>Girard</b>
19	<b>Girard</b>	Deschamps
20	<b>Thomas</b>	Seguin
21	Gauthier	<b>Clerc</b>
22	Guiot	Rousseau
23	Monnot	<b>Simon</b>
24	Chapuis	Renaud
25	Monin	Dubois

Il resterait à examiner la question de la pérennité dans le temps des *géohapax*. Mais on se heurte ici à un problème difficilement surmontable : comment savoir en effet si les *géohapax* de l'actuel département de la Côte d'Or – ou d'une partie de ce département – étaient déjà des *géohapax* du Dijonnais médiéval sans disposer d'un corpus exhaustif pour cette période ? Le changement d'échelle d'analyse pose ici des problèmes complexes. On peut signaler cependant que sur les 20 *géohapax* (identifiables comme tels dans le fichier INSEE au seuil  $n=50$ ) de l'actuel département de la Côte d'Or, la moitié – seulement est-on tenté de dire – sont attestés dans le corpus du Dijonnais médiéval (et le quart le sont avec une fréquence importante).

## Conclusion

L'ensemble des résultats présentés ici souligne l'importance de plusieurs facteurs dans l'évolution des stocks patronymiques. Parmi les plus marquants se trouve la profondeur historique des analyses. L'érosion patronymique avec le temps est bien une réalité. Il ne s'agit pas là d'une simple affirmation qui viendrait corroborer la théorie, ni d'une nouveauté – l'affirmation est ancienne –, mais d'un point qui méritait d'être rigoureusement observé. La preuve a pu ici en être apportée grâce aux différents corpus de données dont nous disposons, certains remontant au XIV<sup>e</sup> siècle. Certes, ces données anciennes ne sont pas exhaustives et ne recouvrent généralement que des parties limitées de territoires, alors que les données récentes peuvent, elles, prétendre à l'exhaustivité. De plus, la nature des

sources patronymiques peut varier d'une période à la suivante, ce qui rend plus délicates les comparaisons entre données anciennes et modernes. Cependant, nos résultats soulignent plusieurs points incontestables et valables quelles que soient les données étudiées. La perte patronymique est bien réelle, mais elle ne semble pourtant pas se réaliser d'une façon aussi mathématique que le prévoit la théorie. Cette érosion progressive du stock des patronymes originels est sensible à toutes les échelles du temps et de l'espace, mais son ampleur varie d'une région à l'autre, selon qu'elles se sont repliées sur elles-mêmes ou au contraire ouvertes aux migrations. Si ce point semble clair, il en ressort également que quelques noms traversent les générations. Non seulement ils perdurent, mais semblent de surcroît enracinés dans une aire géographique précise. C'est pour répondre à cette observation que le concept de *géohapax* a été proposé. Son maniement reste délicat : la difficulté vient de ce qu'il dépend étroitement de l'ampleur de l'espace géographique considéré et du critère de stingence retenu et à partir duquel on estime qu'un nom n'est plus un hapax, mais bien un nom ayant une réelle insertion démographique dans cette espace. L'intérêt de ces *géohapax* est qu'ils sont portés par des personnes dont les origines sont, en toute probabilité, clairement identifiées. Cette caractéristique intéresse non seulement les historiens, mais aussi les généticiens qui considèrent que les spécificités génétiques des porteurs actuels de ces noms sont celles qui prévalaient également dans les populations dont elles sont originaires.

### Remerciements

Nos remerciements vont à tous ceux qui ont contribué à rendre ce travail possible, en particulier Patrice Beck, Guy Brunet et Anna Degioanni.

### Bibliographie

- Bienaymé, I., Jules. 1845. De la loi de multiplication et de la durée des familles. *Société philomatique de Paris* 5: 37–39.
- Bourgain, C., et P. Darlu. 2013. *ADN Superstar ou superflic ? Les citoyens face à une molécule envahissante*. Paris: Seuil (Science Ouverte).
- Brouard, N. 1989. L'extinction des noms de familles en France: une approche. *Dossiers et Recherches*. INED. 27.
- Croix, A. 1999. L'ouverture des villages sur l'extérieur fut un fait éclatant de l'ancienne France. Position de thèse. *Histoire et Sociétés rurales* 11(1): 109–146.
- Crow, J. F. 1980. The Estimation of Inbreeding from Isonymy. *Human Biology* 52(1): 1–14.
- Crow, J. F., and A. P. Mange. 1965. Measurement of inbreeding from the frequency of marriages between persons of the same surname. *Eugen Q* 12(4): 199–203.
- Darlu, P. 2008. Comment "National Geographic" vend le rêve des origines. *Le Monde diplomatique*, juin 2008: 20–21.
- Darlu, P., G. Brunet, et al. 2011. *Spatial and temporal analyses of surname distributions to estimate mobility and changes in historical demography: the example of Savoy (France) from the XVIIIth to XXth century*, Springer.
- Darlu, P., A. Degioanni, et G. Zei. 1996. Patronymes et migrations dans les populations humaines. Dans *Analyse spatiale des données biodémographiques*, J.-P. Bocquet-Appel, D. Courgeau, et D. Pumain (éd.), 225–253, Paris (France): John Libbey-INED.
- Darwin, G. 1875. Marriages between first cousins and their effects. *Journal of the Statistical Society* (London) 38(II): 153–184.

- Degioanni, A., P. Darlu, et J. Ruffié. 1997. Quelques statistiques sur la distribution des patronymes en France. *Population*, 52e année, n° 3: 607–634.
- Dupâquier, J. 1992. Nos patronymes vont-ils disparaître ? Dans *La Société française au XIXe siècle*. J. Dupâquier and D. Kessler (éd.), 461–488. Paris: Hachette.
- Dupâquier, J. 2002. Sédentarité et mobilité dans l'ancienne société rurale. Enracinement et ouverture: faut-il vraiment choisir ?. *Histoire et Sociétés Rurales* 18(2): 121–135.
- Fichiers des noms de famille, 1891–1990* (édition 1999) [fichier électronique], INSEE [producteur], Centre Maurice Halbwachs (CMH) [diffuseur].
- Heyer, É. 1991. Étude démogénétique d'une population humaine. Cas de la maladie de Rendu-Osler, thèse de doctorat, Lyon: Université Claude Bernard-Lyon I.
- Lotka, A. J. 1920. Analytical note on certain rhythmic relations in organic systems. *Proceedings of the National Academy of Sciences* 6: 410–415.
- Lotka, A. J. 1931. The extinction of Families. *Journal of the Washington Academy of Sciences* XXI: 377–380, 453–459.
- Manni, F., B. Toupance, A. Sabbagh, and E. Heyer. 2005. New method for surname studies of ancient patrilineal population structures, and possible application to improvement of Y-chromosome sampling. *American Journal of Physical Anthropology* 126: 214–228.
- Poussou, J.-P. 2002. L'enracinement est le caractère dominant de la société rurale française d'autrefois. *Histoire, Économie et Sociétés* 21(1): 97–108.
- Toulemon, L. 2001. Combien d'enfants, combien de frères et sœurs depuis cent ans? *Population et Sociétés* 374: 1–4.
- Watson, H. W., and F. Galton. 1875. On the probability of the extinction of families. *The Journal of the Anthropological Institut of Great Britain and Ireland* 4: 138–144.