

RÉGIMES DÉMOGRAPHIQUES ET TERRITOIRE : les frontières en question

*Colloque international de La Rochelle
22 - 26 septembre 1998*



ASSOCIATION INTERNATIONALE DES DÉMOGRAPHES DE LANGUE FRANÇAISE

AIDELF

La distribution des patronymes par rapport à la frontière franco-belge

Anna DEGIOANNI *, Pierre DARLU *, Michel FOULON **, Michel POULAIN **

* INSERM U535, Le Kremlin-Bicêtre, France

** GEDAP, UCL, Belgique

Introduction

Le patronyme est né dans une aire linguistique dont il porte la trace en filigrane. Depuis que les patronymes se sont figés aux XVI^e et XVII^e siècles et qu'ils se transmettent par voie masculine, le stock de patronyme évolue sous l'effet de plusieurs mécanismes. Quelques patronymes se transforment à la suite d'erreurs de transmission, d'autres se raréfient et finissent par disparaître faute de porteurs tandis que la plupart des nouveaux patronymes surviennent à la suite de mouvements migratoires.

Certains patronymes correspondent à une aire linguistique très vaste. C'est principalement le cas des prénoms comme *Martin*, *Bertrand*, *Pierre*, *André*... qui couvrent toute l'aire linguistique francophone. Les patronymes indiquant une origine très commune exprimée en français tels que *Dubois*, *Duval*, *Dumont*, *Dupont* ou *Dupuis* recouvrent également toute l'aire francophone. D'autres patronymes présentent des variantes et s'expriment différemment selon l'aire dialectale d'origine. Ainsi les *Duchêne* et *Duchesnoy* se rattachent-ils à l'aire dialectale wallonne tandis que les *Duquenne* et *Duquesnoy* sont d'origine picarde. Les mêmes différences s'observent dans des noms de métiers tels que *Charlier*, *Charpentier* ou *Charron*, en pays wallon, et *Carlier*, *Carpentier* ou *Carron*, en pays picard. Enfin de nombreux patronymes sont d'origine strictement locale parce qu'ils se rattachent à un lieu-dit précis ou que les circonstances qui les ont vu naître sont uniques dans le temps et dans l'espace.

Dans une population suffisamment nombreuse, en l'absence de mouvements migratoires et de mariages entre époux provenant de lieux distincts, les patronymes d'un lieu devraient être aujourd'hui à l'image des patronymes tels qu'ils furent figés dans ce même lieu, il y a quelques siècles. Mais, au fil du temps, les migrations et les mariages ont largement contribué au brassage des populations et des patronymes. Ainsi, par exemple, les *Carlier* en territoire wallon et les *Charlier* en terre picarde sont le fait de migrations présentes ou passées. Avec plus de 3% de patronymes portant le préfixe VAN typiquement flamand, certaines communes industrielles wallonnes telles que Tilleur, à proximité de Liège, ou Châtelineau, près de Charleroi, sont les témoins de l'importance de l'immigration flamande pendant la seconde moitié du XIX^e siècle. Malgré l'impact des migrations, le renouvellement des patronymes d'un lieu donné n'est souvent que très partiel pour autant et les *Duchêne* reste majoritaires en terre wallonne tandis que les *Duquenne* priment en terre picarde. Toutefois, il n'est pas impossible que le stock des patronymes d'une aire très limitée comme une communauté villageoise de quelques centaines d'habitants soit entièrement renouvelée sur une période de trois siècles (Poulain M., 2000). Si les patronymes portés en dehors de leur aire d'origine constituent un excellent marqueur des migrations (Poulain M. et Foulon M., 1981 ; Degioanni A. et al. 1996), l'objectif de cette contribution s'intéressera plutôt à la comparaison des stocks patronymiques propre à chaque région étudiée.

Le postulat sur lequel repose cette recherche est simple. Deux zones qui échangent beaucoup de migrants auront à long terme des stocks patronymiques plus proches et plus semblables que deux zones entre lesquelles les échanges migratoires sont limités. Sous un autre angle et de façon corollaire, on constate que les échanges migratoires sont plus intenses au sein

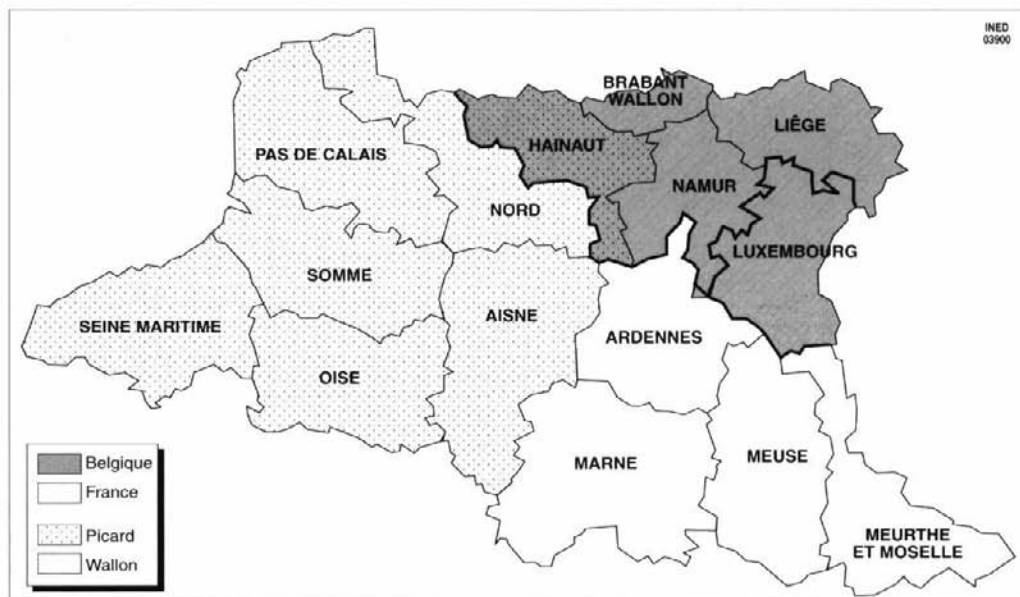
d'une même aire linguistique qu'entre deux aires distinctes¹. En remontant aux origines de la fixation des langues et des dialectes, on ne s'aventure guère en affirmant que les régions marquées par de forts échanges migratoires appartiennent le plus souvent à une même aire linguistique et qu'il existe dès lors une relation étroite entre l'étendue d'une aire linguistique et la configuration des champs migratoires.

En étudiant dans le détail la répartition spatiale des patronymes, on peut mettre en évidence des fractures ou discontinuités spatiales qui, en se référant au postulat énoncé ci-dessus, sont le signe de la présence d'une frontière qui donne lieu à un effet de barrière et de frein par rapport aux mouvements migratoires. La mise en évidence d'une discontinuité spatiale dans la répartition des patronymes entre deux zones contiguës peut s'expliquer de deux façons complémentaires :

– Soit parce qu'elles appartiennent à des aires linguistiques distinctes ou à des zones dialectales différentes, ces deux situations faisant suite le plus souvent à une prépondérance des échanges migratoires à l'intérieur de chaque aire linguistique ou dialectale dès avant l'époque de fixation de ces aires.

– Soit parce que des mouvements migratoires relativement plus récents sont dirigés plus intensément à l'intérieur de chaque zone plutôt que de l'une à l'autre. Le tracé d'une nouvelle limite administrative comme la frontière entre la France et la Belgique au lendemain de Waterloo et de l'indépendance Belge, au début du XIX^e siècle, peut, en favorisant les migrations internes à chaque pays et en freinant les migrations internationales, avoir le même effet sur les stocks de patronymes en créant progressivement une discontinuité selon le tracé de la nouvelle frontière.

FIGURE 1. L'AIRES LINGUISTIQUE FRANCOPHONE CONSIDÉRÉE : LES DÉPARTEMENTS FRANÇAIS ET LES PROVINCES BELGES QUI LA COMPOSE.



¹ Voyer les autres contributions à ce séminaire de La Rochelle.

L'objet plus précis de cette contribution est l'étude de la distribution patronymique dans le Nord de la France et en Wallonie qui correspond à la partie francophone de la Belgique. Nous retiendrons, d'une part, les 10 départements du Nord de la France, soit, par ordre alphabétique, l'Aisne, les Ardennes, la Marne, la Meurthe-et-Moselle, la Meuse, le Nord, l'Oise, le Pas-de-Calais, la Seine-Maritime et la Somme et, en Belgique, les 5 provinces wallonnes, le Brabant-Wallon, le Hainaut, Liège, le Luxembourg et Namur. En étudiant par le détail et de façon comparative les patronymes portés dans ces 15 zones appartenant toutes à l'aire linguistique francophone, on tentera d'identifier des discontinuités spatiales éventuelles et de vérifier dans quelle mesure elles s'identifient, par exemple, à la frontière nationale qui sépare la France de la Belgique ou plutôt à la frontière linguistique qui partage, de façon moins nette, le dialecte picard à l'Ouest (Aisne, Oise, Seine-Maritime, Somme, Pas-de-Calais, Nord, Hainaut) et le dialecte wallon à l'Est (Brabant-Wallon, Namur, Luxembourg, Liège, Ardennes, Marne, Meurthe-et-Moselle, Meuse) (figure 1). Il convient d'être conscient dès à présent que le choix que nous faisons de tenter de rechercher d'éventuelles frontières au sein de la même aire linguistique francophone est largement plus malaisé que d'effectuer une telle recherche entre deux aires linguistiques comme les aires francophone et néerlandophone en Belgique.

2 - Les sources de données sur les patronymes

Les sources disponibles permettant de connaître la distribution des patronymes varient entre la France et la Belgique. Pour la **Belgique**, grâce au Registre National, registre de population entièrement informatisé au niveau national, nous disposons de la liste exhaustive de tous les patronymes portés par province au 31 décembre 1987. Par ailleurs, nous connaissons également la distribution des patronymes au niveau de chacune des 589 communes belges pour autant que les patronymes concernés comptent au moins 5 porteurs. Dans l'impossibilité de traiter tous les patronymes distinctement, une part non négligeable n'étant d'ailleurs portés que par une personne, nous avons retenu pour cette analyse les patronymes les plus fréquents. Toutefois, afin de maintenir au sein de cette analyse les patronymes spécifiques de certaines régions ou communes, nous retiendrons les 5 patronymes les plus fréquents pour chacune des 262 communes wallonnes, même si ces patronymes ne figurent pas parmi les plus fréquents à l'échelle de l'ensemble de la Wallonie. Ceci nous donne, compte tenu des patronymes apparaissant dans plusieurs communes, 980 patronymes distincts dont la fréquence d'apparition varie de 3,4 pour mille pour les *Dubois* à 7 pour cent mille dans le cas des patronymes *Bure* ou *Trum*. En sélectionnant ainsi les patronymes, on ne se borne pas à considérer les plus fréquents pour l'ensemble de la Wallonie mais on inclut dans le corpus des patronymes très peu courants qui sont caractéristiques de l'une ou l'autre commune spécifique.

Pour la **France**, les données proviennent des registres patronymiques de l'INSEE. Ce registre permet de calculer, par canton, la fréquence de tous les patronymes des enfants nés avant 1940 et non décédés avant 1970². Pour les mêmes raisons que pour l'échantillon de patronymes belges, seuls ont été retenus les cinq patronymes les plus fréquents dans chacun des 496 cantons³ situés dans les 10 départements de cette étude. Les données regroupent donc 1055 patronymes, depuis *Masson* (9,9 pour mille) le plus fréquent jusqu'à *Mourière* (1 pour cent mille) le plus rare.

Les deux sources de données n'étant pas identiques, un test était indispensable afin de vérifier l'importance du biais que cela pouvait introduire dans notre analyse. Pour deux communes wallonnes, Mouscron et Ellezelles, nous avons comparé la distribution

² Il s'agit par conséquent de la population de plus de 30 ans en 1970 et répartie par lieu de naissance.

³ Suite à la fusion des communes belges en 1997 et leur accroissement de taille, ce sont les cantons français qui sont les plus appropriés pour réaliser une analyse comparative.

patronymique au 31 décembre 1987 telle qu'elle est connue à l'aide du Registre National et la distribution des patronymes des nouveau-nés des années 1890 à 1970 reconstituée à partir des tables des registres des naissances de ces deux communes. En confrontant les deux distributions par ordre décroissant, on découvre des listes semblables pour ce qui est des patronymes les plus fréquents. Certes, les fréquences d'apparition diffèrent mais les classements ne sont guère perturbés. On note toutefois que la fréquence des patronymes les plus fréquents est quelque peu supérieure sur la base des nouveau-nés par le simple fait que de nombreux patronymes portés au sein d'une population par une seule personne ont de fortes chances de ne jamais figurer parmi les patronymes des nouveau-nés. Ceci a pour effet d'accroître d'autant le poids des patronymes les plus fréquents. Toutefois, en considérant, comme nous allons le présenter, un corpus de plus de 1500 patronymes, ce biais sera largement minimisé.

En considérant simultanément les 980 patronymes wallons et les 1055 patronymes du Nord de la France, on forme un corpus de 1529 patronymes, compte tenu des patronymes communs. C'est ce corpus auquel nous aurons recours pour comparer les distributions patronymiques au sein des différentes régions considérées.

3 - Distribution spatiale de quelques patronymes

La première analyse s'attache à décrire les fréquences avec lesquelles les différents patronymes sont portés dans les 15 régions concernées. Seules les fréquences peuvent être comparées compte tenu des différences de taille entre les régions en terme de population. En faisant la moyenne des fréquences de chaque patronyme pour toutes les régions, on peut déterminer les patronymes les plus fréquents. Ici c'est *Martin* qui l'emporte avec une fréquence moyenne de 33 porteurs pour 10.000 individus. La figure 2 propose la répartition des *Martin* dans l'espace considéré et on constate que la fréquence dépasse les 50 pour 10.000 aussi bien en Meurthe-et-Moselle qu'au Luxembourg belge. Seuls le Nord et le Hainaut présentent des niveaux inférieurs à 20 pour 10.000. Viennent ensuite dans l'ordre les *Petit* (28), *Dubois* (27) et *Lambert* (26). En parcourant la liste des plus fréquents, on y note la plupart des patronymes correspondant à un prénom tel que *François* (24), *Gerard* (21), *Simon* (20), *Laurent* (20), *Thomas* (18), *Noel* (18), *Michel* (17), *Henry* (17) *Bernard* (16), *Bertrand* (15), *Mathieu* (15) ... On y retrouve par ailleurs des professions telles que *Lefevre* (29) ou *Lefebvre* (29), *Leclercq* (16) ou *Leclerc* (8), *Lemaire* (21), *Boulangier* (10), *Mercier* (10) ... Les qualificatifs sont également nombreux tels que *Petit* (28), *Legrand* (19), *Lebrun* (10), *Leroux* (9)... Enfin, certains toponymes communs viennent également en tête avec *Dupont* (20), *Dubois* (27), *Dumont* (18), *Fontaine* (14), *Dupuis* (9), *Duval* (9)... Il n'est pas étonnant que ce soit ces patronymes qui soient les mieux représentés puisqu'il sont communs à l'ensemble de l'aire étudiée et, bien au-delà, à toute l'aire linguistique francophone. C'est d'ailleurs parmi ces patronymes que l'on retrouvera ceux qui sont distribués de la façon la plus uniforme. En rapportant l'écart-type de la distribution des fréquences dans les 15 régions à la moyenne de ces fréquences, on obtient un coefficient de variation qui est de 33% pour les *Benoit* et les *Bernard*, de 35% pour les *Lenoir* et *Olivier* et 36% pour les *Fontaine* et les *Lecomte*. La figure 3 propose la distribution des *Bernard* avec un minimum de 6 pour 10.000 en Seine-Maritime et un maximum de 23 dans le Pas-de-Calais.

FIGURE 2. LA DISTRIBUTION SPATIALE DES PORTEURS DU PATRONYME MARTIN

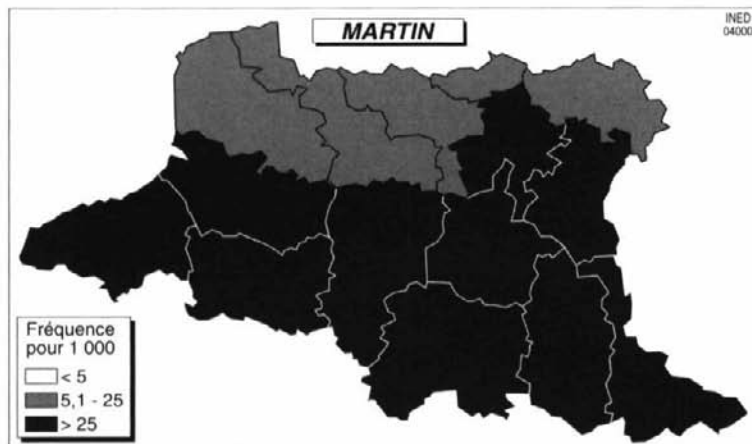
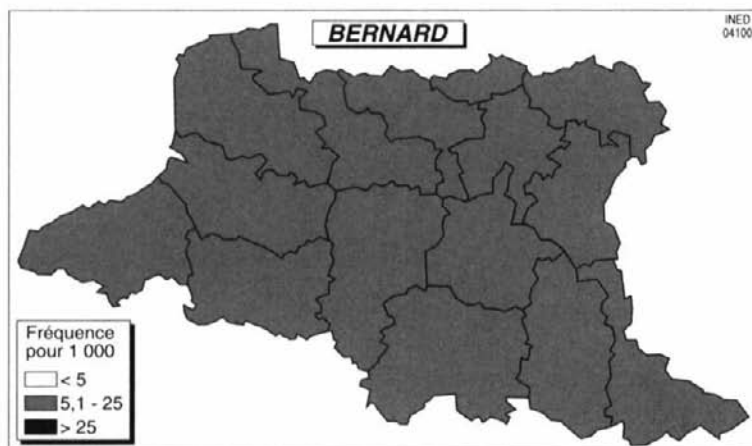
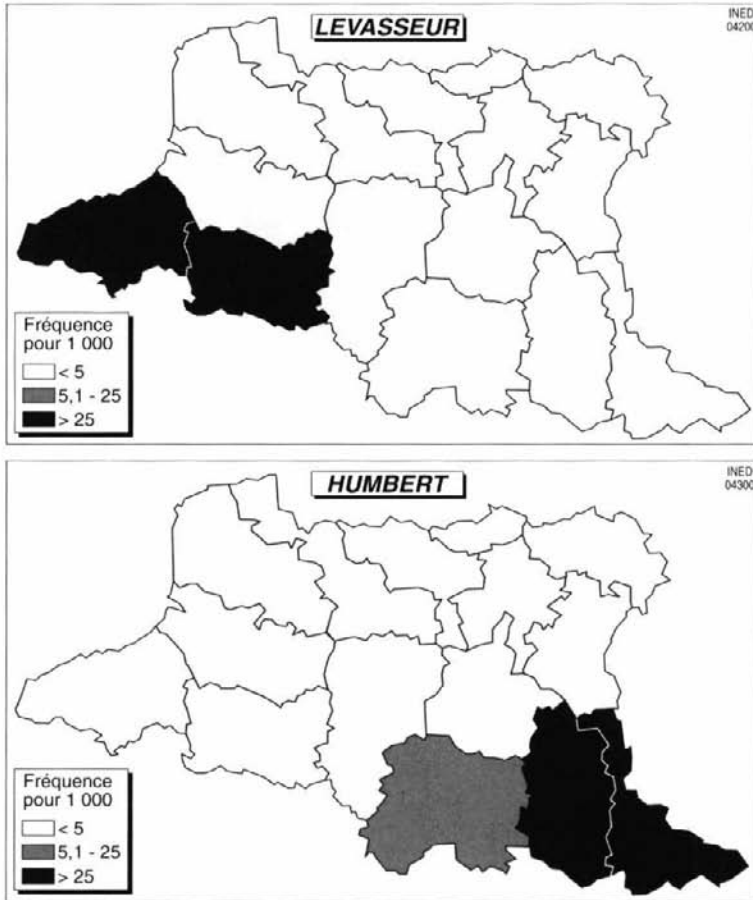


FIGURE 3. LA DISTRIBUTION SPATIALE DES PORTEURS DU PATRONYME BERNARD



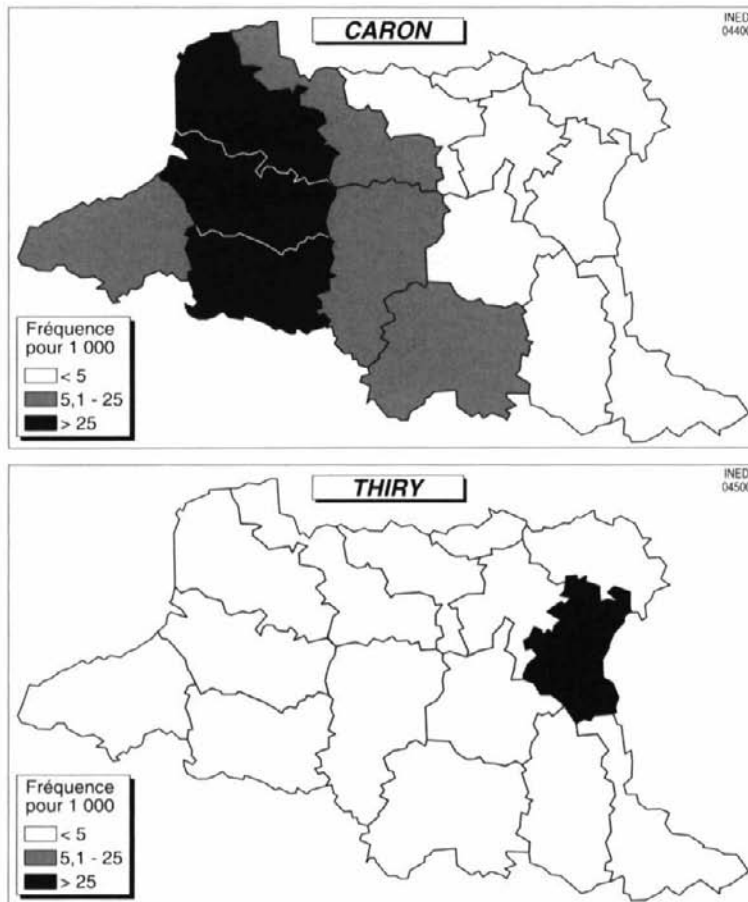
A l'inverse, ce même coefficient de variation permet de déceler les patronymes spécifiques à l'une ou l'autre région et que l'on ne retrouve guère ailleurs. C'est le cas des *Pruvost* typiques du Pas-de-Calais où ils représentent 37 pour 10.000 de la population pour seulement 4 pour 10.000 dans l'ensemble de l'aire étudiée. On notera également les *Levasseur* en Seine-Maritime (32 contre 4) et les *Humbert* en Meuse et Meurthe-et-Moselle (36 et 28 contre 6).

FIGURE 4. LA DISTRIBUTION SPATIALE COMPARÉE DES PORTEURS DES PATRONYMES LEVASSEUR ET HUMBERT



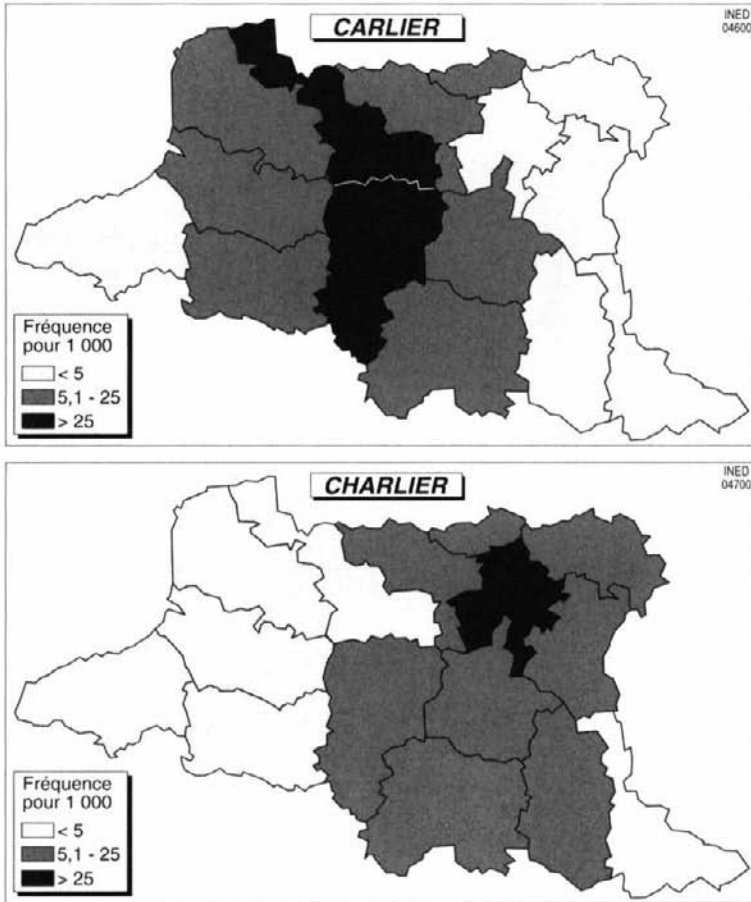
Parmi ces patronymes spécifiques, on peut décèler ceux qui marquent mieux la frontière entre la France et la Belgique. Ce sont, pour la France, *Vasseur* et *Caron* et, pour la Belgique, *Bastin*, *Servais* et *Thiry*. Les figures 5 et 6 permettent de comparer la distribution spatiale des *Caron*, d'une part, des *Thiry*, de l'autre.

FIGURE 5. LA DISTRIBUTION SPATIALE DES PORTEURS DU PATRONYME CARON ET THIRY



D'autres patronymes calquent mieux la frontière entre les aires dialectales wallonne et picarde. Ce sont bien logiquement des patronymes marqués par leur origine dialectale tels que *Carl*ier et *Char*lier, *Carpen*tier et *Charpen*tier. La figure 7 détaille la répartition entre *Charlier* et *Carl*ier sachant que ces deux patronymes présentent une fréquence moyenne de 10 porteurs pour 10.000 individus et peuvent donc être directement comparés.

FIGURE 7. LA DISTRIBUTION SPATIALE COMPARÉE DES PORTEURS DES PATRONYMES CARLIER ET CHARLIER



En conclusion de cette première approche, on constate que certains patronymes courants dans toute l'aire linguistique francophone sont tout à fait inadéquats pour mettre en évidence l'effet d'une frontière éventuelle. Par contre, des patronymes teintés de régionalisme ou spécifique à un dialecte précis seront d'excellents indicateurs pour identifier un éventuel effet de barrière causé par la présence d'une frontière.

4 - Similitude patronymique et effet de frontière

Plutôt que de s'intéresser, en particulier, à un patronyme spécifique et de comparer pour chacun leur distribution dans l'espace, on considérera ici, de façon globale, le corpus de 1529 patronymes représentatifs de l'espace étudié. En comparant leur fréquence d'apparition dans chacun des départements français et provinces belges, on tentera de mesurer une similitude patronymique entre chaque paire de régions. Pour ce faire, nous aurons recours à l'estimation de la corrélation r entre les fréquences spécifiques à chaque patronyme retenu :

$$r_{ij} = \frac{\sum_{k=1}^N p_{ik} p_{jk}}{\left(\sum_{k=1}^N p_{ik}^2 \sum_{k=1}^N p_{jk}^2 \right)^{\frac{1}{2}}}$$

où p_{ik} et p_{jk} sont les fréquences du patronyme k dans les aires i et j respectivement. Les sommes sont effectuées sur les $N=1529$ patronymes.

Cette corrélation mesure la similitude entre les patronymes qui sont portés dans deux régions différentes parmi les 15 régions retenues. Elle varie de 0,89 entre la Meuse et la Meurthe-et-Moselle à 0,30 entre la Somme et le Luxembourg belge. Parmi les plus fortes similitudes, on retrouve les couples de régions suivants : AisnelOise (0,87), OiselSomme (0,84), ArdenneslMarne (0,82), NordlPas-de-Calais (0,82), et Brabant-WallonlHainaut (0,81). Parmi les plus faibles valeurs, on note les suivantes : Pas-de-CalaislLuxembourg (0,30), Seine-Maritimeliège (0,31), Seine-MaritimelLuxembourg (0,31) et Meurthe-et-MosellelPas-de-Calais (0,31).

De prime abord, le rôle descriptif de la corrélation entre chaque paire de régions est essentiel pour rendre compte de la plus ou moins grande similitude entre les patronymes qui y sont portés. Des régions contiguës auront beaucoup de patronymes semblables alors que des régions éloignées présenteront une moindre similitude. C'est bien évidemment le frein de la distance géographique sur les migrations qui explique cette diminution de la diffusion des patronymes lorsque cette distance augmente. Afin d'analyser plus en détail cet effet de la distance, on aura recours à un modèle spatial faisant intervenir explicitement la distance entre les deux régions considérées. Pour ce faire, on transformera au préalable la corrélation r qui varie de 0 à 1 en un indice de similitude variant de 0 à l'infini. S_{ij} sera calculé par la formule :

$$S_{ij} = \frac{r_{ij}}{1 - r_{ij}}$$

Le modèle spatial que nous retiendrons fait intervenir la distance d_{ij} entre les centres des deux régions i et j à travers une fonction de Pareto où le paramètre α à estimer concrétisera l'effet négatif de la distance géographique sur la similitude des patronymes entre les deux zones. La constante k est une constante de niveau telle que la somme des similitudes estimées par ce modèle soit égale à la somme des similitudes effectivement observées. On a par conséquent le modèle décrit par la formule suivante :

$$S_{ij} = kd_{ij}^{-\alpha}$$

L'estimation des paramètres de ce modèle sur les 105 couples de régions se fera par moindres carrés pondérés de telle sorte à minimiser le χ^2 entre estimations et observations. Dans ce premier essai de modélisation, le paramètre α prend une valeur estimée de 1,23 et le niveau de description du modèle est de 43,2% (le χ^2 ayant été réduit d'autant par rapport à une situation de référence où les similitudes entre toutes les paires de régions seraient identiques).

Plutôt que de considérer la distance, on ne retiendra comme facteur descriptif que l'appartenance à un pays, la Belgique ou la France, ou à une aire dialectale, la zone picarde ou la zone wallonne. Dans ce cas le modèle spatial comprend une variable binaire A_{ij} qui vaut 0 si les régions i et j appartiennent au même pays ou à la même aire dialectale et 1 dans le cas contraire. Cette variable apparaît en tant qu'exposant du paramètre γ qui permettra de mesurer l'effet négatif de la frontière considérée sur la similitude des patronymes entre deux régions se trouvant de part et d'autre de cette frontière. Les similitudes sont réduites par un facteur γ

lorsque l'exposant A_{ij} vaut 1 et que les régions sont séparées par la frontière concernée. Ce modèle s'écrit donc comme suit :

$$S_{ij} = k\gamma^{A_{ij}}$$

En considérant la frontière franco-belge, on constate que le paramètre α prend la valeur 0.493 en réduisant le χ^2 de 14,2 % alors que pour la frontière dialectale wallo-picarde, le paramètre vaut 0.404 avec un niveau de description du modèle qui est de 22,7%.

A ce stade de la démarche de modélisation, on peut considérer que l'impact de la distance est de loin supérieur à l'effet de l'une ou l'autre des deux frontières considérées, tout en observant que le rôle de la frontière dialectale semble l'emporter sur la frontière nationale. Néanmoins, comme l'appartenance à un pays ou à une zone dialectale et la distance sont deux variables largement corrélées, il convient de les considérer simultanément dans un modèle spatial du type suivant :

$$S_{ij} = kd_{ij}^{-\alpha}\gamma^{A_{ij}}$$

Le tableau ci-dessous résume les résultats de l'application de ce modèle aussi bien en considérant la frontière nationale que la frontière dialectale.

| | Frontière nationale | Frontière dialectale |
|------------------------------------|---------------------|----------------------|
| Effet de la distance (α) | 1.12 | 1.05 |
| Effet de la frontière (γ) | 0.744 | 0.648 |
| Niveau de description | 44,7% | 46,3% |

En conclusion, la distance est sans conteste la variable qui décrit le mieux la plus ou moins forte similitude des patronymes entre deux régions. Toutefois, un effet négatif de la frontière, qu'elle soit franco-belge ou wallo-picarde vient réduire la similitude entre régions se trouvant de part et d'autre. Le niveau de description est de 44,7% pour la frontière nationale et de 46,3% pour la frontière dialectale. Ainsi, bien que cette frontière dialectale soit moins nette dans l'espace par comparaison à la frontière nationale, elle a un effet légèrement plus marqué sur la répartition spatiale des patronymes.

5 - Partitions patronymiques de l'espace franco-belge.

Une démarche complémentaire de la précédente consiste à rechercher les meilleures partitions que l'on peut obtenir parmi les 15 régions étudiées. Selon la pertinence de ces partitions, la présence des frontières entre elles seront plus ou moins significatives. Plusieurs méthodes peuvent être mises en place dans ce but.

a) Bipartition en ensembles contigus

Cette première approche consiste à procéder à l'ensemble des regroupements possibles à partir des 15 régions en deux blocs contigus en excluant les blocs comptant moins de 5 régions. Voici les résultats de cette démarche :

| Composition des deux ensembles de régions contiguës (dans l'ordre croissant de la similitude) | r |
|---|----------|
| Frontière Wallo-Picarde (Nord + Pas-de-Calais + Somme + Hainaut + Oise + Seine-Maritime + Aisne <i>autres régions</i>) | 0.622 |
| Nord + Pas-de-Calais + Somme + Hainaut + Oise + Seine-Maritime + Aisne + Brabant-Wallon <i>Luxembourg + Liège + Namur + Ardennes + Marne + Meuse + Meurthe-et-Moselle</i> | 0.630 |
| Nord + Pas-de-Calais + Somme + Oise + Seine-Maritime + Aisne <i>Luxembourg + Liège + Namur + Brabant-Wallon + Hainaut + Ardennes + Marne + Meuse + Meurthe-et-Moselle</i> | 0,638 |
| Nord + Pas-de-Calais + Somme + Oise + Seine-Maritime <i>Luxembourg + Liège + Namur + Brabant-Wallon + Hainaut + Ardennes + Marne + Meuse + Meurthe-et-Moselle + Aisne</i> | 0,650 |
| Nord + Pas-de-Calais + Somme + Hainaut + Oise + Seine-Maritime + Aisne + Brabant-Wallon + Namur <i>Luxembourg + Liège + Ardennes + Marne + Meuse + Meurthe-et-Moselle</i> | 0.666 |
| Nord + Pas-de-Calais + Somme + Hainaut + Oise + Seine-Maritime + Aisne + Brabant-Wallon + Namur + Liège <i>Luxembourg + Ardennes + Marne + Meuse + Meurthe-et-Moselle</i> | 0.670 |
| Frontière Franco-Belge (Nord + Pas-de-Calais + Somme + Seine-Maritime + Oise + Aisne + Marne + Ardennes + Meuse + Meurthe-et-Moselle <i>provinces belges</i>) | 0.705 |
| Pas-de-Calais + Somme + Oise + Seine-Maritime + Aisne <i>Luxembourg + Liège + Namur + Brabant-Wallon + Hainaut + Ardennes + Marne + Meuse + Meurthe-et-Moselle + Nord</i> | 0,711 |
| Pas-de-Calais + Somme + Seine-Maritime + Oise + Aisne + Marne + Ardennes + Meuse + Meurthe-et-Moselle <i>provinces belges + Nord</i> | 0.777 |
| Somme + Seine-Maritime + Oise + Aisne + Marne + Ardennes + Meuse + Meurthe-et-Moselle <i>provinces belges + Nord + Pas-de-Calais</i> | 0.789 |

b) Nuées dynamiques

Une deuxième méthode est fondée sur le principe des nuées dynamiques. Il s'agit de trouver les partitions au sein des 15 régions qui minimisent les distances intra-groupes et maximisent les différences entre groupes. La distance utilisée est ici la distance patronymique euclidienne suivante :

$$\delta_{ij}^2 = \sum_{k=1}^N (p_{ik} - p_{jk})^2$$

Étant donnée le grand nombre de partitions possibles dans chacune de ces situations, les noyaux initiaux autour desquels s'effectuent les regroupements de régions ont été choisis de façon à ce qu'ils soient les plus distincts géographiquement (Seine-Maritime et Meurthe-et-Moselle pour deux partitions ; Seine-Maritime, Liège et Meurthe-et-Moselle pour trois partitions).

De manière à contrôler la robustesse de chacune des partitions obtenues, et pour vérifier que le résultat n'est pas dû simplement au hasard, une centaine de tirages aléatoires de 25% des patronymes ont été effectués dans chacun des cas. Le tableau suivant indique le pourcentage de partitions qui sont retrouvées identiques (formes fortes) sur l'ensemble des 100 tirages aléatoires.

| Deux partitions | |
|---|-----|
| Aisne, Ardennes, Marne, Meuse, Meurthe-et-Moselle, Nord, Oise, Pas-de-Calais, Seine-Maritime, Somme | 22% |
| Brabant-Wallon, Hainaut, Liège, Luxembourg, Namur | 21% |
| Aisne, Nord, Oise, Pas-de-Calais, Seine-Maritime, Somme, Hainaut | 19% |
| Ardennes, Marne, Meuse, Meurthe-et-Moselle, Brabant-Wallon, Liège, Luxembourg, Namur | 17% |
| Trois partitions | |
| Aisne, Nord, Oise, Pas-de-Calais, Seine-Maritime, Somme | 58% |
| Ardennes, Marne, Meuse, Meurthe-et-Moselle, Luxembourg | 53% |
| Hainaut, Liège, Brabant-Wallon, Namur | 52% |
| Ardennes, Marne, Meuse, Meurthe-et-Moselle | 20% |
| Aisne, Oise, Pas-de-Calais, Seine-Maritime, Somme | 18% |

Les deux bipartitions retrouvées les plus fréquemment distinguent les provinces belges des départements français (22% et 21% respectivement). Cependant, deux autres partitions, presque aussi fréquentes, séparent les régions picardes des régions wallonnes (19 et 17%). Lorsque l'option est choisie d'effectuer trois partitions au lieu de deux, les partitions se retrouvent avec une fréquence nettement plus élevée que dans le cas de deux partitions, montrant ainsi que ce type de partition rend mieux compte de la diversité patronymique. Les partitions les plus fréquentes regroupent alors les départements français de dialecte picard (58%), les départements français de dialecte wallon auxquels s'ajoute le Luxembourg (53%), et les quatre provinces belges (sans le Luxembourg) (52%).

c) Analyses de variance

L'analyse précédente nous a montré que la partition la plus significative se faisait entre trois groupes. Pour mieux délimiter ces groupes, indépendamment du choix du noyau initial, il a été procédé à des partitions aléatoires des 15 régions en trois groupes, chacun de ces groupes incluant au moins deux régions. Le critère retenu pour choisir la meilleure partition est donné par le rapport F entre l'estimation de la variance des distances patronymiques entre partitions et l'estimation de la variance des distances entre régions à l'intérieur d'une même partition.

La meilleure partition pour laquelle les contrastes entre groupes sont les plus significatifs ($F=27.4$; $df=1$; $P<0.000$) est celle regroupant la totalité des régions belges, les départements au dialecte picard de l'Ouest (Seine-Maritime, Somme, Nord, Pas-de-Calais, Oise, Aisne) et les départements au dialecte wallon de l'Est (Ardennes, Meuse, Meurthe-et-Moselle, Marne).

6 - Représentation arborescente des distances patronymiques

A partir de la matrice de distances patronymiques calculés entre les régions deux à deux, il est possible d'obtenir une représentation qui relie entre elles les régions les plus proches, sous la forme d'un arbre dichotomique. La méthode utilisée est celle du Neighbor-Joining appliquée à la matrice des distances calculées selon la formule de Nei :

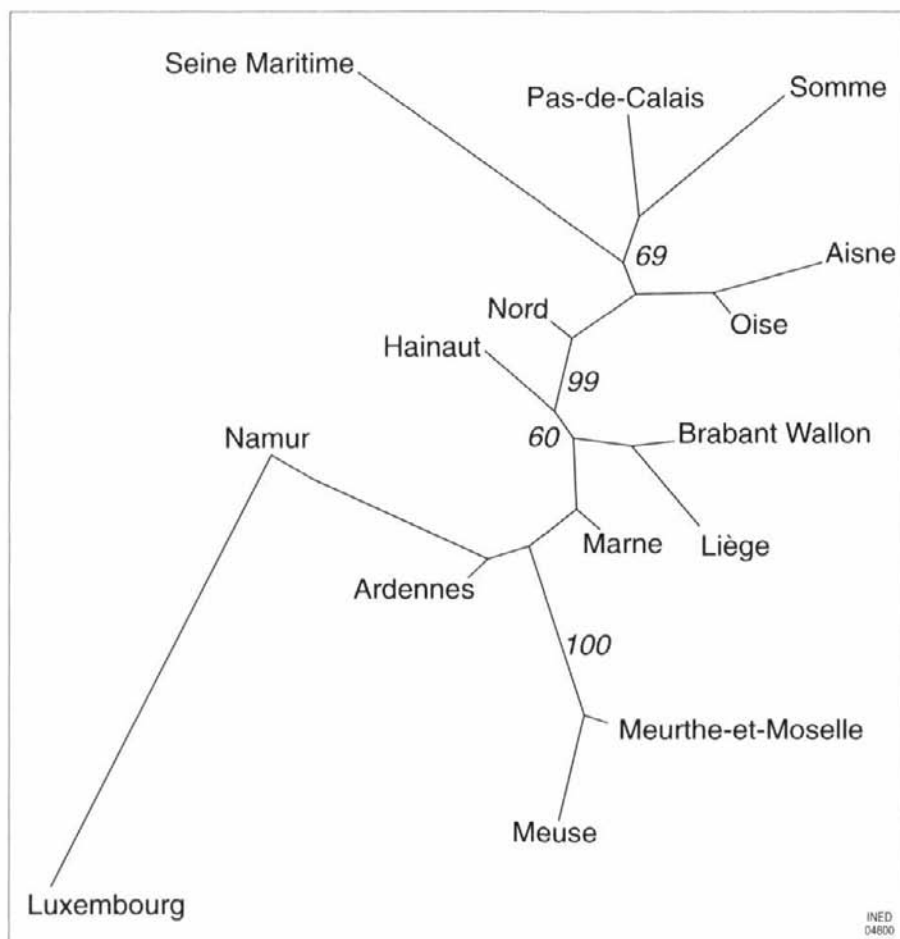
$$\delta_{ij} = -1n(r)$$

où r est la corrélation définie plus haut.

Pour évaluer la robustesse des regroupements, la reconstruction a été effectuée 100 fois. Chacune de ces 100 reconstructions est fondée sur un ensemble de patronymes tirés au hasard (avec remise) parmi l'ensemble de tous les patronymes. La figure 8 montre l'arbre obtenu. Les longueurs des branches sont proportionnelles à la distance patronymique estimée et les valeurs

sur les branches internes représentent le nombre de fois, sur les 100 ré-échantillonnages, où l'on retrouve la partition créée par cette branche. Ainsi, la valeur de 69% indique que le groupement (Pas-de-Calais, Somme) a été retrouvé 69 fois sur les 100 tirages aléatoires. Seuls les pourcentages supérieurs à 60 ont été indiqués.

FIGURE 8. LA REPRÉSENTATION ARBORESCENTE DES DISTANCES PATRONYMIQUES ENTRE LES 15 RÉGIONS



Cette figure montre que les seules dichotomies réellement robustes sont celles :

- séparant la Meuse et la Meurthe-et-Moselle des autres régions ;
- séparant les départements de l'Ouest de la France (Somme, Nord, Seine-Maritime, Pas-de-Calais, Aisne, Oise) de toutes les autres régions. L'inclusion du Hainaut à ce groupe, qui délimite la partition picard/wallon, se fait avec une valeur de bootstrap bien plus faible (60%). Le Hainaut est donc une région de transition entre deux groupes. Le Luxembourg reste la région la plus distante des autres régions bien qu'elle se branche logiquement avec la région de Namur et des Ardennes.

7 - Conclusion

La longue histoire commune au Nord de la France et aux provinces wallonnes de Belgique, associée au fait que ces régions partagent une même langue, se traduit par une large similitude de leurs patronymes, particulièrement des plus fréquents, comme par exemple les *Martin*. Malgré cette large similitude, il apparaît cependant des différences entre départements ou régions en fonction de leur éloignement géographique. Entre la Seine-Maritime et le Luxembourg, s'échelonnent toutes les aires, françaises ou belges, comme le montre la représentation de la matrice de distance patronymique en arbre des plus proches voisins et comme le montre également l'application du modèle ajustant les similitudes patronymiques aux proximités géographiques. Cependant, aux deux extrêmes, une différenciation plus prononcée peut s'observer. Elle est probablement à mettre sur le compte de l'influence des régions voisines non étudiées, les régions neerlandophones ou germanophones pour le Luxembourg, les régions normandes pour la Seine-Maritime.

Cette continuité n'est, en fait, qu'apparente. En effet, on a pu y déceler des distorsions et détecter des frontières, en particulier une frontière dialectale (picard-wallon) et une frontière nationale (France-Belgique). On peut penser que la première a résisté aux échanges migratoires, alors que la seconde, inversement, a été créée là où elle n'existait pas auparavant. La réalité de ces frontières a pu être mise en évidence par différentes méthodes, par bipartition d'ensembles contigus, par recherche des partitions géographiques maximisant les différences patronymiques entre aires géographiques tout en minimisant les distances à l'intérieur de ces aires. Le modèle d'analyse spatiale montre d'ailleurs que la frontière nationale est légèrement plus faible que la frontière dialectale.

La diversité des approches conduit parfois à des résultats dont les contradictions sont intéressantes à analyser. En particulier on observe que

- les départements des Ardennes, de la Meuse, de la Meurthe-et-Moselle et de la Marne constituent parfois un groupe unique parfois se rattachent aux provinces belges de dialecte wallon,
- le Hainaut, pourtant de dialecte picard, se situe en position charnière entre les régions picardes et les régions belges wallonnes,
- le Luxembourg tantôt fait partie du groupe français des Ardennes, de la Marne, de la Meuse et de la Meurthe-et-Moselle et tantôt du groupe belge.

En dehors de ces ambiguïtés, certains groupes restent fortement constitués, comme la Somme, l'Oise, l'Aisne, le Pas-de-Calais, comme la Meuse, la Meurthe-et-Moselle, les Ardennes ou comme Liège et Namur.

BIBLIOGRAPHIE

- DEGIOANNI A., LISA A., ZEI G. et DARLU P. (1996), Patronymes italiens et migration italienne en France entre 1891 et 1940. *Population*, VI, 1996, PP 1153-1180.
- NEI M., Genetic distance between populations, *Amer. Nat.* (1972), 106, pp. 283-292.
- POULAIN M. et FOULON M. L'immigration flamande en Wallonie : évaluation à l'aide d'un indicateur anthroponymique (1981), *Revue belge d'histoire contemporaine*, Vol 1-2, pp. 205-244.
- POULAIN M. (2000), Le renouvellement de la population au sein d'une communauté villageoise : l'apport respectif des patronymes et des généalogies, *Mélanges en l'honneur d'Alfred Perrenoud*, à paraître.