

# ORIGINE GÉOGRAPHIQUE DE LA POPULATION URBAINE

**Sandrine ASSALIN** : U.M.R. ESPACE, Université d'Avignon

Email : [sandrine.assalin@univ-avignon.fr](mailto:sandrine.assalin@univ-avignon.fr)

---

*RÉSUMÉ.* Depuis environ deux siècles, la population des campagnes a migré vers les villes. Cette concentration démographique s'est faite selon un modèle de diffusion à partir des campagnes, en provoquant un brassage des populations dans les espaces urbains. Les villes ont eu une attraction différenciée en fonction de leur localisation et de leur taille. L'étude portera sur les communes centres des villes moyennes. L'attraction peut se mesurer à travers la diffusion des patronymes rares, qui sont utilisés comme des traceurs permettant de suivre les populations sur une longue période. Cette présentation portera sur la mise au point d'une méthode expérimentale de détection des flux de population à travers la diffusion de patronymes rares et par l'utilisation de la métrique du  $\chi^2$ .

*ABSTRACT.* For approximately two centuries, the population of the country side has migrated towards the cities. This demographic concentration was done according to a model of diffusion starting from the country side by causing a mixing of the populations in urban spaces. The cities had an attraction differentiated according to their localization and from their size. The study will relate to common the centers of the average cities. Attraction can be measured through the diffusion of the rare patronyms, which are used like tracers making it possible to follow the populations over a long period. This presentation will relate to the debugging of experimental method of detection of flows of population through the diffusion of rare patronyms and by the use of metric of the  $\chi^2$ .

*MOTS CLEFS* : Attraction urbaine, exode rural, migrations, patronymes,  $\chi^2$ .

*KEY WORDS* : Urban attraction, rural migration, migrations, patronyms,  $\chi^2$ .

---

Si l'on connaît très bien le phénomène de peuplement des villes qui a eu lieu pendant l'exode rural au détriment des campagnes, on connaît moins bien les trajectoires suivies par les migrants. D'où sont originaires les populations que l'on retrouve aujourd'hui dans les communes centres des villes moyennes ? peut-on définir des aires d'attraction urbaines ? Ce sont des questions auxquelles il est difficile de répondre faute d'informations suffisantes sur le sujet. Les sources les plus fréquemment utilisées ne permettent pas de travailler à la fois sur la longue période, la petite échelle et sur de grandes masses de population. Afin de déterminer l'origine des migrants dans les villes moyennes il a fallu recourir à une nouvelle source d'information : l'information patronymique. Par leur mode de transmission les patronymes constituent une information relativement stable dans le temps, ils peuvent être utilisés comme des traceurs permettant de suivre les populations sur la longue période. C'est donc à partir de patronymes rares et de leur localisation trouvée dans les annuaires téléphoniques que l'information a été bâtie. La recherche est effectuée sur les villes moyennes qui ont été relativement attractives durant l'exode rural. Cette attraction modérée par rapport à celle exercée par les grandes villes devrait permettre de retrouver le berceau d'origine des migrants. La méthode est expérimentale, elle sera appliquée à la ville du Mans, et se déroule en trois temps : Tout d'abord il faut bâtir un échantillon représentatif de la population en place, puis trouver une méthode permettant de détecter les origines des migrants, et enfin valider de façon statistique les résultats obtenus.

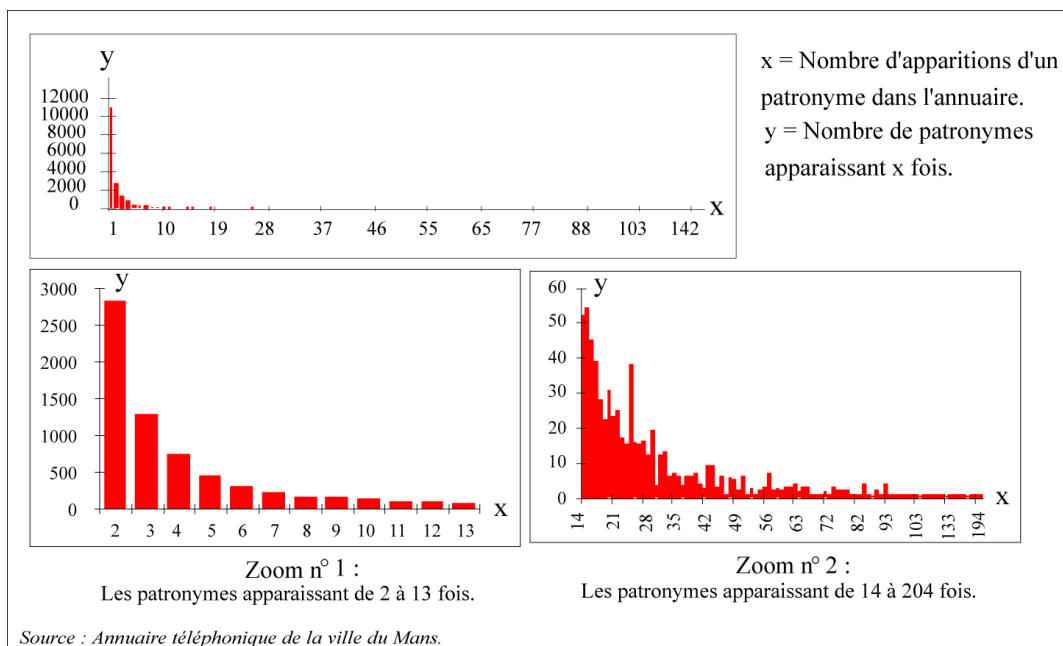
### 1. La construction d'un échantillon de population urbaine

Afin d'étudier les migrations il est indispensable de sélectionner un petit nombre de patronymes représentatifs : d'échantillonner. Etudier l'ensemble des patronymes présents dans une ville est impossible, essentiellement pour des raisons pratiques. L'échantillon est donc fixé à 50 patronymes, qui ne sont pas choisis au hasard : ils doivent être représentatifs de la population totale du point de vue des migrations. L'annuaire téléphonique donne deux renseignements :

- le nombre de patronymes présents dans une ville,
- la fréquence d'apparition de chacun d'eux.

La fréquence d'apparition d'un patronyme semble étroitement liée à son lieu d'origine. En effet, on peut émettre l'hypothèse que plus un patronyme revient fréquemment dans l'annuaire téléphonique d'une ville, plus il est probable qu'il est originaire d'un département proche, au contraire plus il est rare et plus il est probable que son origine est lointaine. L'échantillon doit donc respecter la structure patronymique trouvée dans les annuaires téléphoniques (Figure 1). Ces patronymes ont été sélectionnés dans les communes centres, afin de faciliter le travail de comptage. Dans les annuaires téléphoniques les postes téléphoniques attribués à chaque patronyme sont présentés par commune, sélectionner des patronymes dans l'ensemble de l'agglomération aurait été trop fastidieux.

Figure 1 : Exemple de la structure patronymique trouvée dans la ville du Mans



L'échantillon doit être calibré de manière à se caler sur la structure patronymique. Si dans l'annuaire téléphonique on compte 50% de patronymes n'apparaissant qu'une seule fois alors l'échantillon devra lui aussi contenir 50% de patronymes ayant une fréquence d'apparition unique.

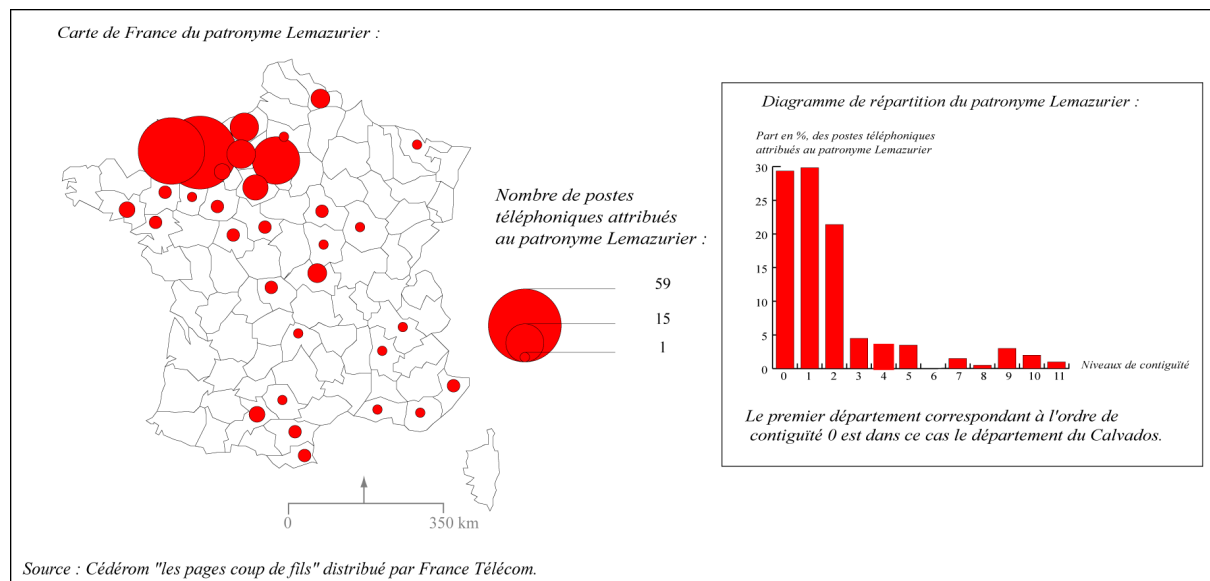
## 2. Détermination de l'origine des migrants

L'aire d'attraction urbaine d'une ville correspond ici aux départements d'où est originaire sa population. Pour délimiter les aires d'attraction urbaines, il faut arriver à déterminer l'origine des migrants et donc le département d'origine de chaque patronyme échantillonné. Les annuaires téléphoniques donnent la répartition des patronymes en France, il est donc possible d'établir les cartes de France de chacun des patronymes échantillonnés. L'hypothèse est alors que le département qui contient le plus fort pourcentage d'un patronyme, que l'on appellera « premier département », est le berceau d'origine de ce patronyme. La démarche va consister à infirmer ou au contraire confirmer cette hypothèse, par confrontation entre les répartitions réelles de chaque patronyme et des répartitions théoriques.

### 2.1. Les diagrammes de répartition réelle

La démarche consiste dans un premier temps à établir les diagrammes de répartition patronymique réelle selon les ordres de contiguïtés au premier département. Ces diagrammes sont tirés des cartes de France de chacun des patronymes élaborées grâce aux annuaires téléphoniques. Afin de comparer les différents diagrammes entre eux les valeurs sont des pourcentages (Figure 2). Une fois les 50 diagrammes de répartition réelle établis, ils sont confrontés à des diagrammes de répartition théorique.

Figure 2 : Répartition du patronyme Lemazurier présent dans l'échantillon du Mans



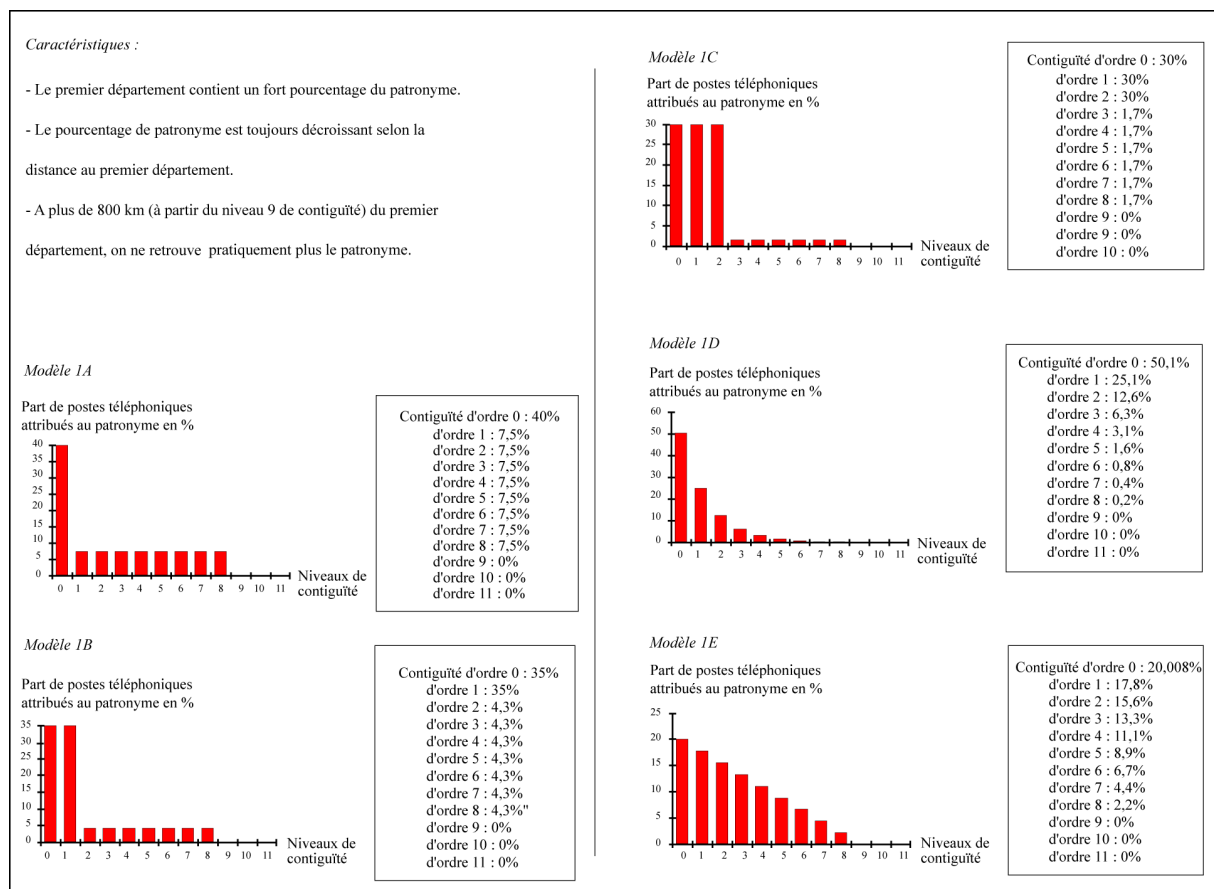
### 2.2. Construction des diagrammes de répartition théorique

L'ensemble des diagrammes de répartition théorique ont été construits à partir de l'observation de répartition de patronymes dont on connaissait les origines. Il existe trois possibilités :

- Soit le patronyme est bien originaire du premier département et on le retrouve presque exclusivement dans le premier département et dans les départements de contiguïtés d'ordre 1 ou 2 à ce premier département. Son diagramme de répartition réelle correspond à une répartition qui laisse penser que le premier département est bien l'origine du patronyme. Dans ce cas il existe une forte corrélation entre un des diagrammes théoriques présentés en figure 3 et la répartition réelle du patronyme.

- Soit le patronyme n'est pas originaire du premier département ou bien a une origine multiple. Dans ce cas, il n'existe plus une hyperfréquence unique dans le premier département et ses alentours, on retrouve le patronyme dans une autre région et son diagramme de répartition réelle se rapproche d'un des diagrammes théoriques présentés par la figure 4. Bien souvent on retrouve ce type de configuration lorsque le premier département est un département dans lequel existe une ville importante. La ville a été si attractive qu'elle concentre l'essentiel du patronyme mais on retrouve une hyperfréquence plus lointaine car elle n'est pas pour autant le berceau d'origine du patronyme.
- Soit la répartition réelle d'un patronyme n'a aucune des caractéristiques des différents diagrammes théoriques. Souvent ce type de situation se retrouve lorsqu'un patronyme s'est tellement diffusé que l'on ne peut plus arriver à interpréter sa répartition et dans ce cas son origine reste indéterminée. C'est essentiellement pour cette raison que l'information utilisée est limitée aux patronymes rares, les seuls qui ont encore des hyperfréquences d'apparition dans certains lieux et qui permettent donc d'obtenir des informations sur leurs origines ainsi que sur les trajectoires migratoires de leurs porteurs.

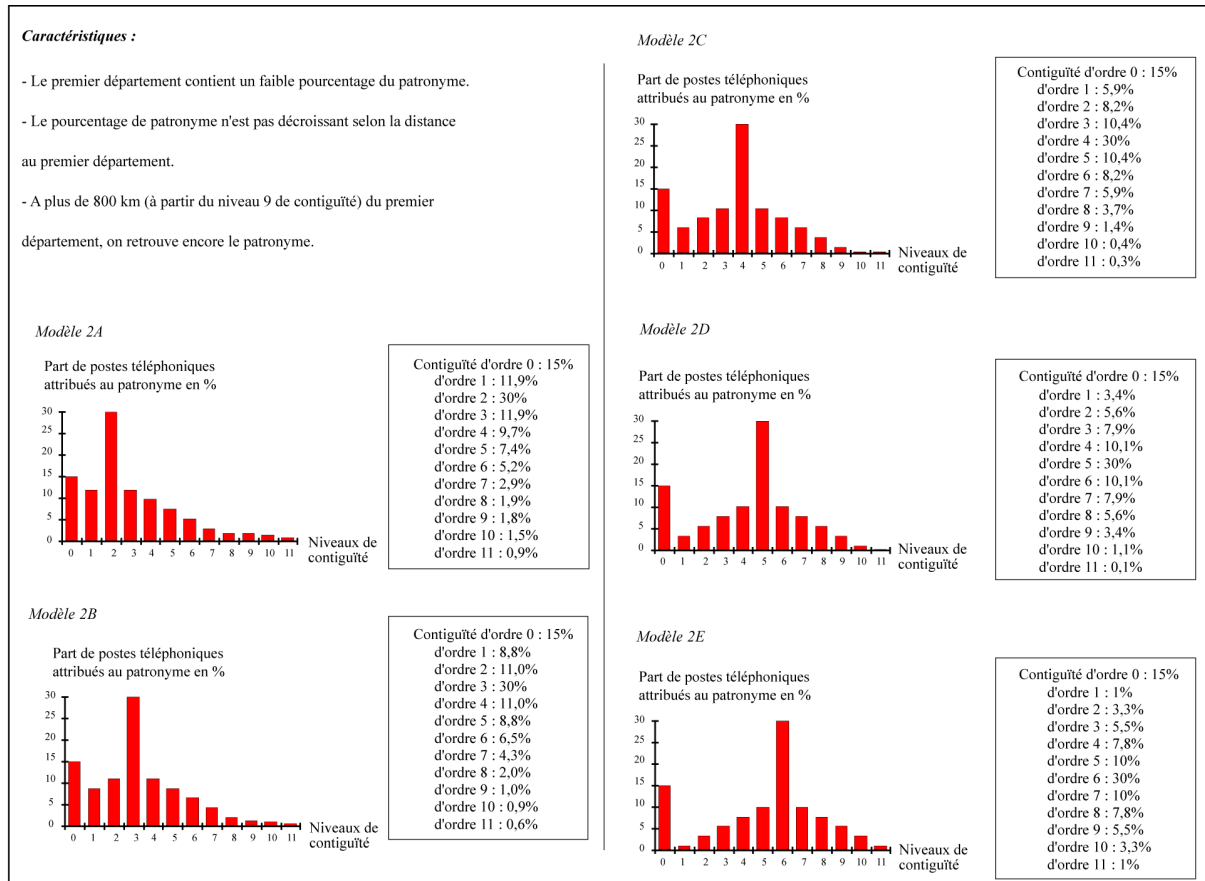
**Figure 3 : Les diagrammes théoriques qui confirment l'hypothèse selon laquelle le premier département est le département-origine**



- Soit le patronyme n'est pas originaire du premier département ou bien a une origine multiple. Dans ce cas, il n'existe plus une hyperfréquence unique dans le premier département et ses alentours, on retrouve le patronyme dans une autre région et son diagramme de répartition réelle se rapproche d'un des diagrammes théoriques présentés par la figure 4. Bien souvent on retrouve ce type de configuration lorsque le premier département est un département dans lequel on retrouve une ville importante. La ville a été si attractive qu'elle concentre l'essentiel du patronyme mais on retrouve une hyperfréquence plus lointaine car elle n'est pas pour autant le berceau d'origine du patronyme.
- Soit la répartition réelle d'un patronyme n'a aucune des caractéristiques des différents diagrammes théoriques. Souvent ce type de situation se retrouve lorsqu'un patronyme s'est tellement diffusé que l'on ne peut plus arriver à interpréter sa répartition et dans ce cas son origine reste indéterminée. C'est essentiellement pour cette raison que l'information utilisée est limitée aux patronymes rares, les seuls

qui ont encore des hyperfréquences d'apparition dans certains lieux et qui permettent donc d'obtenir des informations sur leurs origines ainsi que sur les trajectoires migratoires de leurs porteurs.

**Figure 4 : Les diagrammes théoriques qui infirment l'hypothèse selon laquelle le premier département est le département-origine**

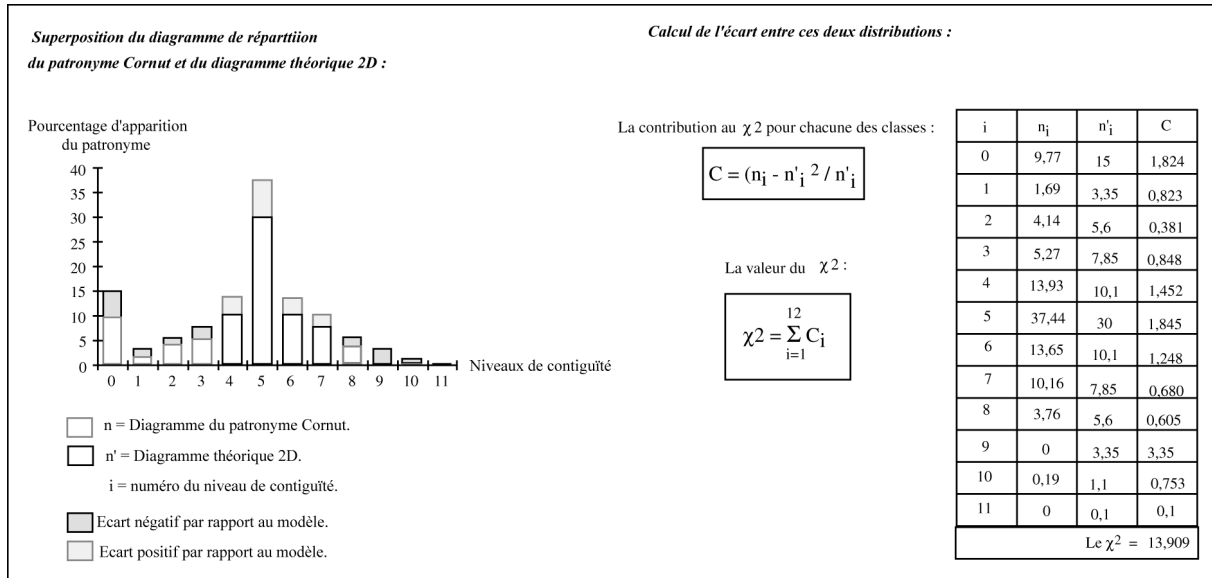


### 2.3. Confrontation des répartitions réelles et théoriques par la métrique du $\chi^2$

Pour chacun des patronymes la répartition réelle va être confrontée aux modèles théoriques. C'est la métrique du  $\chi^2$  qui est utilisée pour comparer les distributions (Figure 5). Avec la métrique du  $\chi^2$ , les écarts sur les faibles valeurs prennent donc toute leur importance et contribuent largement à la valeur totale de  $\chi^2$  obtenue. Cette métrique s'adapte beaucoup mieux à la problématique que la métrique euclidienne, puisque le moindre pic d'hyperfréquence non attendu est alors détecté.

C'est la valeur minimale de  $\chi^2$  pour l'ensemble des confrontations qui sera retenue. Par exemple pour un patronyme on compare distribution réelle avec l'ensemble des répartitions théoriques. Si on trouve une valeur de  $\chi^2$  minimum entre sa distribution réelle et une distribution théorique de la figure 3 alors on dira que le premier département est bien le département origine du patronyme. Si au contraire la valeur minimum est trouvée par comparaison à une des répartitions de la figure 4 alors on dira que le premier département n'est pas le département d'origine. Si pour l'ensemble des comparaisons on ne trouve pas de  $\chi^2$  inférieur à 50 alors on dira que l'origine du patronyme est indéterminée.

**Figure 5 : Confrontation modèle théorique et réalité**



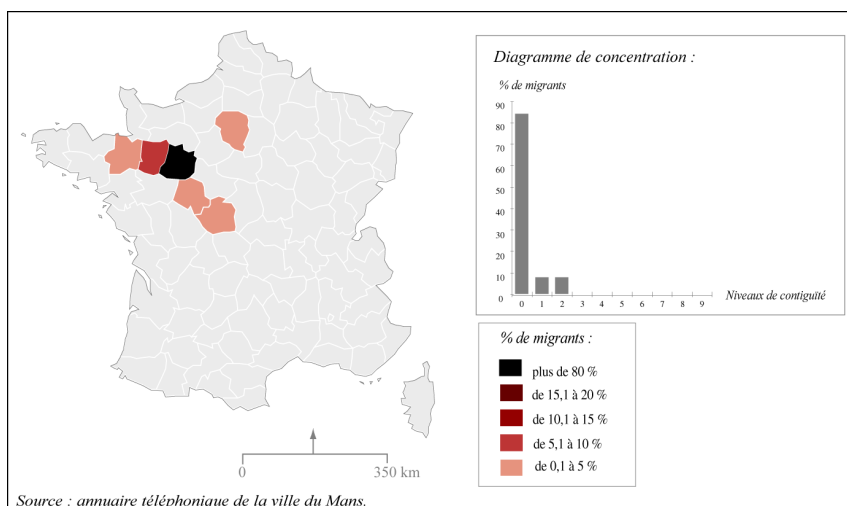
### 3. Validation statistique des résultats

Les résultats obtenus peuvent être validés de façon statistique. La validation des hypothèses émises au départ permet de tester la robustesse de la méthode en la réutilisant sur des échantillons différents de patronyme.

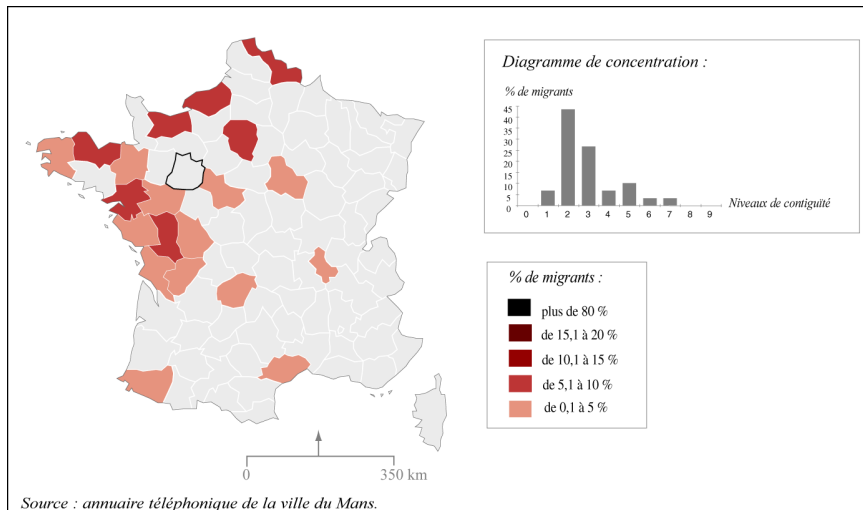
#### 3.1. Validation de l'hypothèse selon laquelle la fréquence d'apparition d'un patronyme dans une ville est liée à la distance à son origine

La première hypothèse émise était que la fréquence d'apparition d'un patronyme dans l'annuaire était liée à la distance à son origine. Afin de valider cette hypothèse, deux échantillons différents ont été sélectionnés. Tout d'abord un échantillon composé uniquement de patronymes fréquents, puis un échantillon composé de patronymes rares (Figure 6 et Figure 7).

**Figure 6 : Echantillon de patronymes apparaissant au moins dix fois dans l'annuaire du Mans**



**Figure 7 : Echantillon de patronymes n'apparaissant qu'une seule fois dans l'annuaire du Mans**



Sur ces deux cartes il apparaît clairement que plus les patronymes sont fréquents dans un annuaire et plus leur origine est proche ; au contraire plus ils sont rares plus leur origine est lointaine. La figure 6 montre que les patronymes apparaissant plus de dix fois dans l'annuaire du Mans sont à plus de 80% des patronymes originaires du département de la Sarthe et les autres ont une origine éloignée au maximum de deux contiguïtés de ce même département. Les fréquences d'apparition semblent donc bien liées à la distance à l'origine.

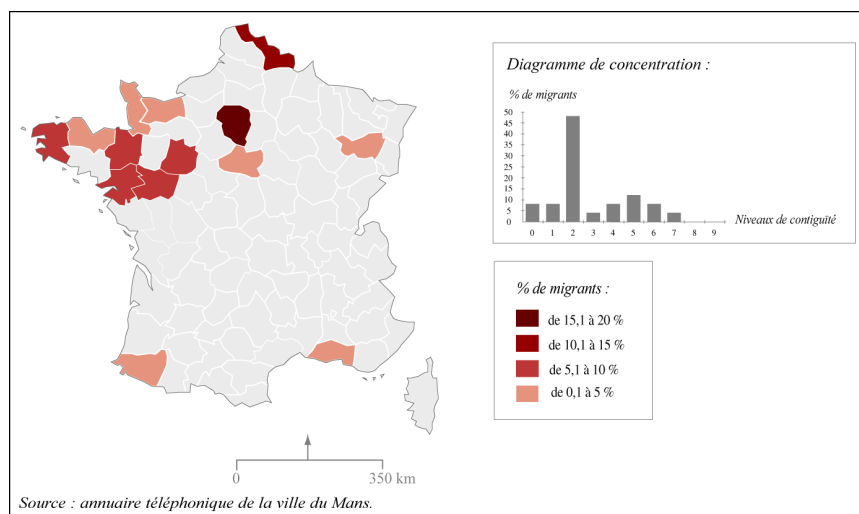
### 3.2. Validation de la méthode d'échantillonnage

La méthode est-elle robuste ? Pour tester cela trois échantillons ont été sélectionnés :

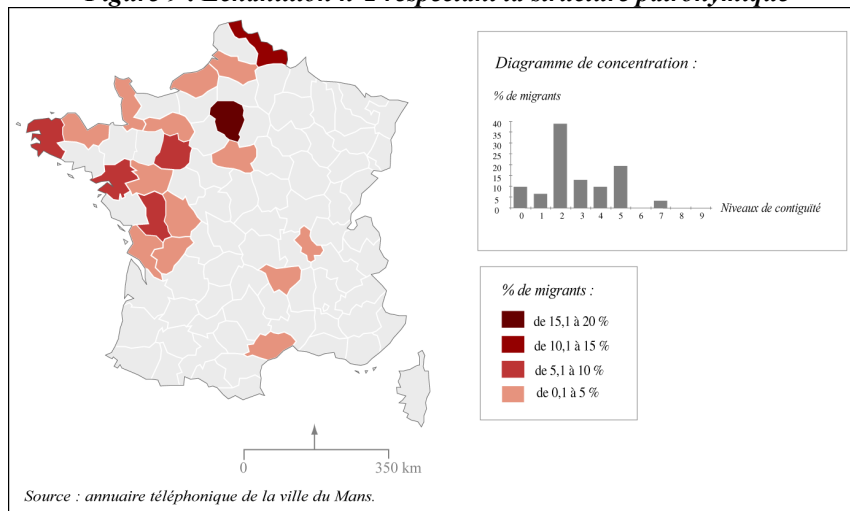
- un échantillon respectant la structure patronymique trouvée dans la ville du Mans (Figure 8).
- un deuxième échantillon respectant toujours la structure patronymique mais composé de patronymes différents (Figure 9).
- un échantillon tiré au hasard (Figure 10).

Ces échantillons devraient permettre d'une part de savoir si en prenant des échantillons de patronymes qui respectent la structure patronymique on obtient des résultats plus conformes à la réalité qu'en prenant un échantillon tiré au hasard ; d'autre part de tester la robustesse de la méthode en vérifiant que deux échantillons sélectionnés de manière identique mais composés de patronymes différents donnent des résultats sensiblement identiques.

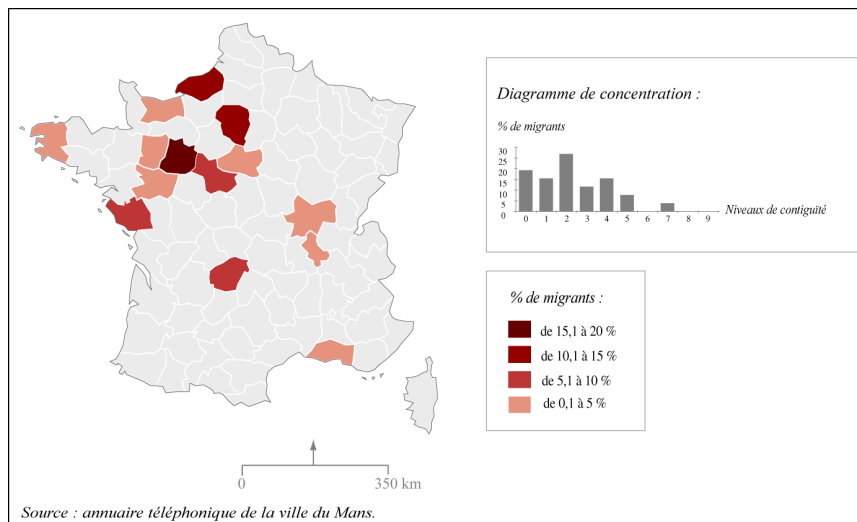
**Figure 8 : Echantillon n°1 respectant la structure patronymique**



**Figure 9 : Echantillon n°2 respectant la structure patronymique**



**Figure 10 : Echantillon tiré au hasard**



Les cartes établies montrent deux enseignements. Tout d’abord les deux échantillons qui respectent la structure patronymique se ressemblent plus que celui tiré au hasard. Ceci signifie que la structure patronymique trouvée dans l’annuaire est liée aux origines des migrants et qu’il n’y a pas indépendance entre fréquence d’apparition et origine. Mais on constate aussi que cette ressemblance est assez relative et que l’on ne peut pas obtenir de résultats probants à partir de cet échantillon. En effet, les résultats sont sensiblement identiques pour les départements contenant un grand nombre de migrants, départements relativement proches, mais très disparates lorsque l’on travaille sur les foyers de migrations plus éloignés.

Cette méthode va donc permettre de travailler sur les grands foyers d’origine des migrants et non sur les foyers secondaires, pour cela la sélection de l’échantillon devra être modifiée. Seuls les patronymes fréquents dans les annuaires devront être sélectionnés afin d’obtenir des résultats significatifs.



## Conclusion

Cette méthode mise au point de façon expérimentale, va pouvoir être utilisée afin d'étudier les grands foyers d'origine des migrants des villes moyennes françaises. Plus largement elle devrait permettre de trouver des pistes de recherche sur la place de ces villes moyennes dans le processus de diffusion des populations durant l'exode rural, en tenant compte du fait que le choix de travailler sur les villes moyennes induit un biais, puisqu'elles peuvent être à la fois réceptrices et émettrices de population. Pour cela une étude comparative devra être menée sur plusieurs villes choisies en fonction de leur position dans l'espace et les résultats interprétés en fonction de ces positions relatives.

## Bibliographie

- BARRAI I., FORMICA G., BARALE R., BERETTA M., (1989). Isonimy and migration distance. *Annals of human genetics*, n°53, pp. 249-262.
- DARLU P., DEGIOANNI A., GIANNA Z., (1996). Patronymes italiens et migrations italienne en France entre 1891 et 1940. *Population*, n°6, p.1153-1180.
- DARLU P., DEGIOANNI A., RUFFIE J., (1997). Quelques statistiques sur la distribution des patronymes en France. *Population* n°3, p.607-634.
- DARLU P., RUFFIE J., (1992). L'immigration dans les départements français étudiée par la méthode des patronymes. *Population*, n°3, p.719-734.
- GRIMMEAU J-P., (1992). Les migrations et la frontière linguistique en Belgique. *Espace, Populations et Sociétés*, n°2, p.253-258.
- POULAIN M., FOULON., (1998). Frontières linguistiques, migrations et distribution spatiale des noms de famille en Belgique. *L'espace géographique*, n°1, p.53-62.