

## Prénoms des descendants d'immigrés en France : Essai de reproduction d'un article scientifique

Résumé court (1 page).....	2
Résumé détaillé (5 pages).....	3
Introduction .....	8
1. Présentation de l'article reproduit .....	9
2. Des résultats largement médiatisés .....	11
3. Comme un doute... ..	12
4. Une méthodologie inaccessible .....	15
5. Méthodologie de la reproduction « au plus près » : la base de données .....	24
6. Méthodologie de la reproduction « au plus près » : la nomenclature des prénoms par origine.....	25
7. Essai de reproduction du premier résultat spectaculaire : « Nicolas » .....	26
8. Essai de reproduction du second résultat spectaculaire : le « 23 % » .....	46
9. Essai de reproduction des autres résultats .....	52
10. Une seule option : la rétractation .....	60
11. Erreurs de bonne foi ou fraude scientifique ? .....	65
Conclusion .....	73
Références .....	75
Annexe 1A : Copie de l'article reproduit .....	86
Annexe 1B : L'enquête Trajectoires et origines (TeO) .....	91
Annexe 2 : Les reprises médiatiques de l'article .....	92
Annexe 4A : Les échanges de messages électroniques avec les auteurs et l'INED .....	94
Annexe 4B : Les programmes informatiques communiqués par les auteurs .....	124
Annexe 5A : La base de données.....	147
Annexe 5B : Le programme informatique de la reproduction « au plus près » .....	183
Annexe 6A : Une histoire comparée des prénoms en Europe et au Maghreb .....	267
Annexe 6B : La nomenclature des prénoms par origine.....	285
Annexe 7 : Le « Nicolas » du tableau page 3 .....	327
Annexe 8 : Le « 23 % » de la figure 2 page 3 .....	355

## Résumé court (1 page)

# Prénoms des descendants d'immigrés en France : Essai de reproduction d'un article scientifique

*Ce résumé court est destiné aux chercheurs, aux journalistes et au grand public.*

Un article scientifique publié en avril 2019 dans la revue *Population et sociétés* indique qu'en France métropolitaine en 2008, les prénoms les plus fréquents chez les petits-fils d'immigrés du Maghreb seraient « Yanis » et « Nicolas ». Au total, selon l'étude, seulement 23 % des petits-enfants d'immigrés du Maghreb porteraient un prénom « arabo-musulman », un ordre de grandeur assez proche des 16 % de petits-enfants d'immigrés d'Europe du Sud qui porteraient un prénom « latin ». Considérant les prénoms des descendants d'immigrés comme « un marqueur culturel » et « une mesure de l'assimilation », les auteurs notent que « Les prénoms que reçoivent les petits-enfants [d'immigrés du Maghreb] sont, en 2008, proches de ceux que la population majoritaire donne à ses enfants », et que « La trajectoire suivie par les originaires du Maghreb mène au même point d'arrivée que celle suivie par les Européens du Sud, mais de manière différée. »

Problème : ces résultats, publiés par le bulletin d'information de l'Institut national d'études démographiques (INED) et largement médiatisés, sont inexacts. Vérification faite, en métropole en 2008 les prénoms les plus fréquents chez les petits-fils d'immigrés du Maghreb ne sont pas « Yanis » et « Nicolas », mais « Karim » et « Nassim ». Et d'après la nomenclature des prénoms de l'INED, ce ne sont pas 23 % mais 49 % des petits-enfants d'immigrés du Maghreb qui portent un prénom « arabo-musulman », contre 8 % des petits-enfants d'immigrés d'Europe du Sud qui portent un prénom « latin ». L'article étant inexact, il doit — conformément aux usages de la communauté scientifique — être rétracté en bonne et due forme. Ces erreurs sont troublantes car pendant 20 mois, d'avril 2019 à décembre 2020, les auteurs et l'INED ont refusé de communiquer leur méthodologie et ils ont empêché la vérification de leurs résultats.

### Définitions

Immigré : personne née étrangère à l'étranger.

Petit-enfant d'immigrés du Maghreb : personne ayant au moins 1 grand-parent immigré du Maroc, d'Algérie ou de Tunisie.

## Résumé détaillé (5 pages)

# Prénoms des descendants d'immigrés en France : Essai de reproduction d'un article scientifique

*Ce résumé détaillé est destiné aux chercheurs, aux journalistes et au grand public.*

## Chapitre 1. Présentation de l'article reproduit

En avril 2019, le bulletin d'information scientifique de l'Institut national d'études démographiques (INED) a publié un article intitulé : « [Quels prénoms les immigrés donnent-ils à leurs enfants en France ?](#) » (Coulmont et Simon 2019). S'inscrivant dans un courant de recherches en sciences sociales qui étudie le choix du prénom comme « un marqueur culturel » et « une mesure de l'assimilation », l'article part de la question suivante : « Les prénoms donnés par les immigrés à leurs enfants, et ceux que reçoivent à la génération suivante leurs petits-enfants, sont-ils puisés dans le registre des prénoms des pays d'origine, ou sont-ils au contraire similaires à ceux donnés par la population majoritaire [c'est-à-dire la population sans ascendant immigré ni natif des DOM] ? »

Fondé sur l'exploitation statistique de l'enquête *Trajectoires et origines* (TeO) menée en France métropolitaine en 2008, cet article présente notamment deux résultats spectaculaires :

- Tableau page 3 : les prénoms les plus fréquents chez les petits-fils d'immigrés du Maghreb sont « Yanis » et « Nicolas » ;
- Figure 2, page 3 : si 94 % des immigrés du Maghreb (1<sup>ère</sup> génération) et 84 % des enfants d'immigrés du Maghreb (2<sup>e</sup> génération) portent un prénom « arabo-musulman », ce n'est le cas que de 23 % des petits-enfants d'immigrés du Maghreb (3<sup>e</sup> génération), un ordre de grandeur assez proche des 16 % de petits-enfants d'immigrés d'Europe du Sud qui portent un prénom « latin ».

## Chapitre 2. Des résultats largement médiatisés

En avril 2019, les résultats de cette étude sont abondamment repris par la presse nationale, du journal *Le Monde* à *20 Minutes* et de *Libération* à *Valeurs actuelles*, mais aussi dans la presse régionale et étrangère et dans de nombreux médias, radios, chaînes télévisées et sites Internet d'information. La présentation dominante de l'article, fidèle à son contenu, est assez bien illustrée par celle qu'en a donné le journal *Le Monde*, dans son article intitulé « Yanis, Sarah ou Inès : quand l'intégration se lit dans les prénoms des petits-enfants d'immigrés » :

« Leurs grands-parents s'appelaient Farid, Ahmed et Rachid, Fatima, Fatiha et Khadija. Les petits-enfants s'appellent désormais Yanis, Nicolas et Mehdi, Sarah, Inès et Lina. Ce sont les prénoms les plus portés par les Français descendants d'immigrés originaires du Maghreb. Contrairement à certaines idées reçues, « *les petits-enfants de la première génération s'appelant Abdelkader ou Karim sont très largement minoritaires* », explique Baptiste Coulmont. »

## Chapitre 3. Comme un doute...

Ces résultats quelque peu contre-intuitifs sont-ils vrais ? Pour le savoir, quatre jours après la publication, j'ai contacté l'un des auteurs pour lui demander des précisions sur la méthodologie utilisée. Suite à nos échanges, je me suis rendu compte que l'encadré méthodologique de l'article (page 2) accumule les inexactitudes et les omissions, ce qui m'a conduit à chercher à vérifier ces résultats.

#### **Chapitre 4. Une méthodologie inaccessible**

Pour reproduire ces résultats, je dispose des mêmes données que les auteurs, l'enquête TeO étant accessible aux chercheurs qui en font la demande. Ne me manque que leur méthodologie, c'est-à-dire leur nomenclature des prénoms par origine ainsi que leur programme informatique. J'ai donc demandé, à partir d'avril 2019, ces éléments méthodologiques, non soumis au secret statistique et que nombre de revues scientifiques publient en ligne (en annexe des articles). Pendant 20 mois, malgré mes demandes répétées, j'ai été confronté à un refus systématique de communication — et ce, de la part des auteurs mais aussi des représentants de l'INED : le rédacteur en chef de la revue, la directrice de l'INED et le référent à l'intégrité scientifique de l'INED (Annexe 4A, p. 94). Si les auteurs ont finalement communiqué leur nomenclature des prénoms — non sans lui avoir préalablement substitué une nomenclature fallacieuse, qui incluait parmi les prénoms originaires des « États de langue arabe » des prénoms grecs, turcs, espagnols et portugais —, ils ne m'ont communiqué qu'un programme informatique incomplet (Annexe 4B, p. 124), qui ne permet pas de reproduire leur Figure 3 sur la proportion de prénoms « arabo-musulmans » chez les petits-enfants d'immigrés du Maghreb selon la religiosité du parent. Ce faisant, les auteurs et l'INED ont non seulement empêché la reproduction à l'identique de leurs analyses : ils ont aussi violé les règles de l'intégrité scientifique et la « [Charte nationale de déontologie des métiers de la recherche](#) ».

#### **Chapitres 5 et 6. Méthodologie de la reproduction « au plus près »**

Pour savoir si les résultats de l'article sont exacts, je reproduis le tableau et les trois figures sur la même base de données que les auteurs : la version nominative de l'enquête TeO. L'échantillon privilégié, le seul sans doublons, est représentatif des personnes de 18-50 / 18-60 ans en ménage ordinaire et de leurs enfants de 0-17 ans, habitant en France métropolitaine en 2008 (N = 42 208). Cet échantillon (pondéré) permet de reproduire divers mouvements de la mode des prénoms en France depuis les années 1950, ce qui indique qu'il est de bonne qualité, tout comme la variable « prénom » de l'enquête. J'utilise aussi la même nomenclature des prénoms par origine que les auteurs, une fois expurgée de leurs erreurs de recodage (comme leur inclusion des prénoms « arméniens » parmi les prénoms « arabo-musulmans »). Cette nomenclature de l'INED concorde globalement bien avec une autre nomenclature conçue indépendamment et elle est globalement conforme aux connaissances établies en onomastique.

#### **Chapitre 7. Essai de reproduction du premier résultat spectaculaire : « Nicolas »**

Coulmont et Simon (2019, p. 3) indiquent qu'en France métropolitaine en 2008, les prénoms les plus fréquents chez les petits-fils d'immigrés du Maghreb seraient « Yanis » et « Nicolas ». Est-ce vrai ? L'échantillon maximal TeO des petits-fils d'immigrés du Maghreb ne contient que 23 « Yanis » et 11 « Nicolas » — des effectifs trop faibles pour permettre d'établir un palmarès des prénoms comme celui qui a été publié. Mais sur l'échantillon retenu pondéré, « Nicolas » est effectivement l'un des prénoms les plus portés par les petits-fils d'immigrés du Maghreb, tout comme « Jérémy » et « Benoît ». Ce sont même plus de 3 % des petits-fils d'immigrés du Maghreb qui seraient prénommés « Nicolas ». Et si l'on reproduit le tableau sur d'autres échantillons, il arrive que « Nicolas » soit le prénom le plus fréquent chez les petits-fils d'immigrés du Maghreb !

Problème : personne ne semble avoir entendu parler d'autant de petits-fils d'immigrés du Maghreb prénommés « Nicolas ». Ni les experts des migrations qui ont bien voulu répondre à mes questions, ni les recherches par entretiens menées auprès des couples « mixtes », ni la presse écrite, ni la radio et la télévision françaises, ni les ouvrages publiés en français depuis 1970, ni les moteurs de recherche Internet ou l'encyclopédie Wikipedia, ni l'annuaire téléphonique et les réseaux sociaux en 2021, ni même les décès en France ne comptent suffisamment de « Nicolas » originaires du Maghreb. Alors, que se passe-t-il ? Un examen de

la méthodologie de l'enquête permet d'envisager que ses bases de données contiennent une erreur. Une partie de ces « Nicolas » ne seraient-ils pas des petits-enfants de *rapatriés* français ou européens de l'ex-empire colonial français en Afrique du Nord (y compris des Français et des Européens d'Algérie), autrement dit des petits-enfants de *pieds-noirs* ? Les *pieds-noirs* ne sont pas considérés comme des immigrés, parce que suite à l'indépendance des pays du Maghreb, ils ont conservé la nationalité française.

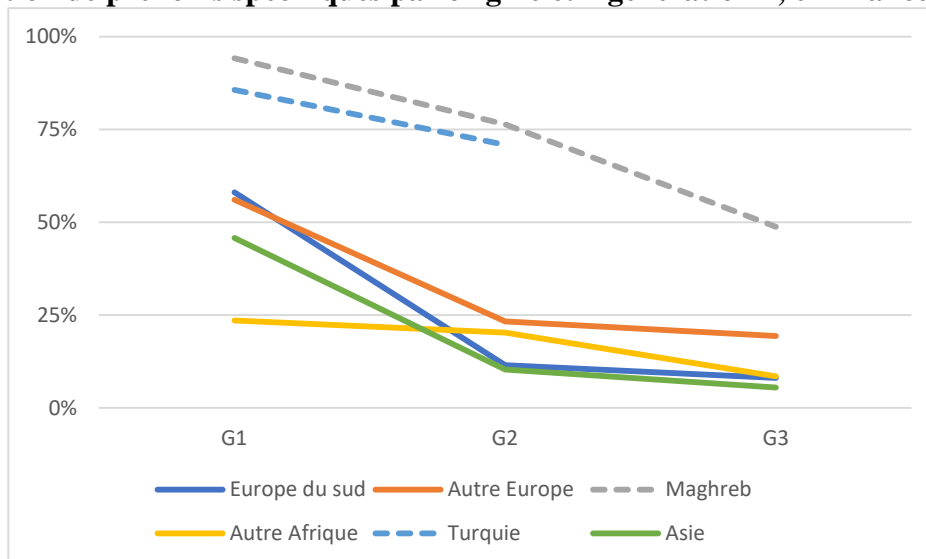
Pour vérifier si ces « Nicolas » ne seraient pas des petits-fils de *pieds-noirs*, j'examine d'abord les prénoms de leur entourage familial. Leur entourage porte des prénoms largement « français » ou européens (87 %) et ils ne sont pas, typiquement, les enfants de « Mohamed » et « Sarah », mais plutôt ceux de « Christine » et « Philippe ». J'estime ensuite combien d'immigrés et de descendants d'immigrés du Maghreb de l'enquête sont vraisemblablement des *pieds-noirs* et descendants de *pieds-noirs*, en combinant 4 critères : (i) l'enquête qui est immigré (G1) ou enfant d'immigrés (G2) du Maghreb se déclare « plutôt d'accord » ou « tout à fait d'accord » avec l'idée que « on le voit comme un Français », (ii) dans son enfance ses parents originaires du Maghreb ne lui parlaient ni arabe ni berbère, et ni (iii) aucun de ses parents originaires du Maghreb ni (iv) lui-même n'est musulman. D'après ces estimations, la plupart (9/11) des « Nicolas » du tableau publié ne seraient pas des petits-fils d'immigrés du Maghreb, mais des petits-fils de *pieds-noirs*. Plus généralement, 1 % des immigrés (G1) et 10 % des enfants d'immigrés (G2) du Maghreb de l'enquête TeO seraient des (descendants de) *pieds-noirs*, mais surtout, 26 % des petits-enfants d'immigrés (G3) du Maghreb seraient des petits-enfants de *pieds-noirs*, ce qui explique pourquoi le tableau publié comprend un prénom comme « Nicolas », mais seulement chez les petits-enfants (G3). Comme les *pieds-noirs* sont par définition nés avant 1962 alors que, dans l'enquête TeO, les immigrés du Maghreb peuvent être nés jusqu'en 1990, les *pieds-noirs* sont en moyenne bien plus âgés que les immigrés du Maghreb, si bien qu'en moyenne ils ont eu plus de temps d'avoir leurs petits-enfants.

Une fois les immigrés du Maghreb et leurs descendants dissociés des *pieds-noirs* et de leurs descendants, les prénoms les plus fréquents chez les petits-fils d'immigrés du Maghreb de l'échantillon retenu sont non pas « Yanis » et « Nicolas », mais « Karim » et « Nassim ». Si les petites-filles d'immigrés du Maghreb portent des prénoms relativement diversifiés, les petits-fils portent pour la plupart des prénoms « arabo-musulmans ». Le Top 20 des prénoms les plus portés par les petits-fils d'immigrés du Maghreb ne comporte *aucun* prénom « français » non ambigu et *aucun* des prénoms du Top 20 des fils d'enquête du groupe « majoritaire » — une situation très différente de celle des descendants d'immigrés d'Europe du Sud (ou d'Asie). Pas étonnant, donc, que le prénom « Nicolas » ne fasse pas partie du Top 3 des prénoms des petits-fils d'immigrés du Maghreb, d'autant qu'il est celui d'un saint catholique qui est associé à la Reconquista. « Nicolas » est en revanche un prénom porté par des descendants de *pieds-noirs*, comme le réalisateur Nicolas Bedos ou le journaliste Nicolas Demorand.

### **Chapitre 8. Essai de reproduction du second résultat spectaculaire : le « 23 % »**

Coulmont et Simon (2019, p. 3) indiquent qu'en France métropolitaine en 2008, seulement 23 % des petits-enfants d'immigrés du Maghreb portent un prénom « arabo-musulman ». Là encore, est-ce vrai ? Même en conservant les petits-enfants de *pieds-noirs* parmi les petits-enfants d'immigrés du Maghreb, ce ne sont pas 23 % mais 37 % des petits-enfants d'immigrés (G3) du Maghreb qui porteraient un prénom « arabo-musulman ». Et une fois les petits-enfants d'immigrés du Maghreb dissociés des petits-enfants de *pieds-noirs*, il apparaît que sur l'échantillon TeO retenu pondéré, 49 % des petits-enfants d'immigrés du Maghreb portent un prénom que la nomenclature de l'INED catégorise comme « arabo-musulman » — loin devant les proportions de prénoms « spécifiques » que l'on peut observer chez les petits-enfants d'immigrés d'Europe du Sud (8 %), d'autre Europe (19 %), d'autre Afrique (9 %) ou d'Asie (5 %) (voir graphique ci-dessous).

### Proportion de prénoms spécifiques par origine et « génération », en France en 2008



Lecture : parmi les personnes originaires du Maghreb, la part de prénoms spécifiques (« arabo-musulmans ») passe de 94 % (G1) à 76 % (G2) puis 49 % (G3), d'après la nomenclature de l'INED.

Champ : personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008, données pondérées, N = 42 208. Les G1, G2 et G3 Maghreb qui en réalité sont probablement (descendants de) rapatriés sont exclus des analyses.

Source : TeO.

Les petits-enfants d'immigrés du Maghreb sont non seulement ceux qui portent le plus souvent un prénom spécifique à leurs origines, mais aussi ceux qui portent le moins souvent un prénom « français » (15 %), loin derrière les petits-enfants d'immigrés d'Europe du Sud (59 %). La sur-rétention des prénoms d'origine s'observe d'ailleurs chez d'autres populations d'origine musulmane en Europe, comme chez les descendants d'immigrés du Pakistan au Royaume-Uni et chez les descendants d'immigrés de Turquie en Allemagne.

### Chapitre 9. Essai de reproduction des autres résultats

La figure 1 de l'article présente des résultats *sélectifs*, ce qui nuit à la connaissance des faits et biaise l'image globale qui s'en dégage. Si, depuis la fin des années 1940, de moins en moins de prénoms attribués en France sont « français », ce n'est pas seulement — comme le montre la figure 1 — parce que le groupe « majoritaire » attribue de moins en moins de prénoms « français » (au profit de prénoms « internationaux ou anglophones »), mais aussi — ce que cette figure 1 ne montre pas — parce que le groupe « minoritaire » (composé des immigrés, des natifs des DOM et de leurs descendants) attribue lui aussi de moins en moins de prénoms français (au profit de prénoms « arabo-musulmans »), et parce que ce groupe « minoritaire » devient démographiquement de plus en plus important au fil du temps. Quant à la figure 3 de l'article, elle sous-estime la part de prénoms « arabo-musulmans » chez les petits-enfants d'immigrés du Maghreb, même s'il est avéré que plus les petits-enfants d'immigrés du Maghreb ont des parents religieux, plus ils se sont vus attribuer des prénoms « arabo-musulmans ».

### Chapitre 10. Une seule option : la rétractation

Les principaux résultats étant faux (ou sélectifs au point que cela biaise l'image globale qui s'en dégage), l'article doit faire l'objet d'une rétractation en bonne et due forme. Contrairement à ce que les auteurs indiquent — un message repris par la plupart des médias en avril 2019 —, les descendants d'immigrés du Maghreb n'abandonnent pas leurs prénoms spécifiques presque autant que les descendants d'immigrés d'Europe du Sud. Au contraire, les



descendants d'immigrés du Maghreb sont, de loin, ceux qui abandonnent *le moins* leurs prénoms spécifiques (et ceux qui se voient le moins souvent attribuer un prénom « français ») : il ne s'agit pas là simplement d'une « idée reçue », mais d'un fait solidement établi.

### **Chapitre 11. Erreurs de bonne foi ou fraude scientifique ?**

Pourquoi des auteurs et une institution de recherche qui considèrent leurs résultats comme fiables, ou méthodologiquement défendables, empêcheraient-ils de reproduire leurs résultats ? Quel besoin auraient-ils de communiquer une nomenclature fallacieuse et un programme informatique incomplet et de violer leurs engagements ainsi que la « [Charte nationale de déontologie des métiers de la recherche](#) » ? Deux chercheurs compétents — l'un spécialiste des prénoms, l'autre spécialiste des immigrés et fin connaisseur des catégorisations statistiques des immigrés du Maghreb et des pieds-noirs — auraient-ils pu croire qu'en métropole en 2008, les prénoms les plus fréquents chez les petits-fils d'immigrés du Maghreb étaient « Yanis » et « Nicolas », sans se poser de questions ni prendre le temps d'examiner les prénoms de l'entourage de la dizaine de « Nicolas » concernés ? Auraient-ils pu ne pas être alertés par le fait que, parmi les petits-enfants d'immigrés du Maghreb de leur figure 3, seulement 50 % sont d'origine musulmane ? Vraisemblablement, les auteurs ne savaient que trop bien que leurs spectaculaires résultats étaient faux ou douteux : c'est pourquoi, dès fin avril, début mai 2019, ils ont empêché leur libre examen. Patrick Simon et l'INED ont d'ailleurs un passif en matière de manipulations statistiques destinées à « démontrer » que, contrairement aux idées reçues, les immigrés du Maghreb et leurs descendants s'intègrent pleinement dans la société française. Tout indique que l'objectif de Patrick Simon — chercheur-militant de la mouvance idéologique « décoloniale », Directeur de recherche à l'INED, directeur de département à l'Institut Convergences Migrations et chercheur associé à Sciences Po Paris — est d'influer favorablement sur les attitudes du grand public à l'égard des immigrés et de l'immigration. Il est regrettable que ces méconduites scientifiques bénéficient de la complaisance du milieu de la recherche, non seulement au sein de l'INED mais plus globalement en sociologie et en démographie de l'immigration.

### **Conclusion**

En publiant ce travail, je vise plusieurs objectifs :

- ✓ faire rétracter l'article de Coulmont et Simon (2019), et rétablir la vérité au sujet des prénoms des descendants d'immigrés en France ;
- ✓ alerter la communauté des chercheurs en sciences sociales sur une imposture scientifique : l'INED, qui est l'institut de recherche censé informer les Français et les pouvoirs publics sur les questions démographiques, non content d'avoir publié — et diffusé aux journalistes — des énormités, a violé les règles de l'intégrité scientifique et m'a empêché de vérifier ses résultats ;
- ✓ alerter les journalistes et le grand public sur une fraude scientifique dont l'objectif est manifestement de biaiser l'image de l'immigration aux yeux des Français ;
- ✓ alerter sur le contexte académique qui a permis tout cela : la complaisance du milieu de la recherche en sociologie et en démographie de l'immigration, et le manque de fiabilité scientifique de ce champ de recherche, qui souffre d'un biais idéologique volontiers « décolonial ».

En biaisant les faits publiés et communiqués aux journalistes, au grand public et aux pouvoirs publics, les méconduites scientifiques sapent la confiance dans la recherche et dégradent le débat démocratique. Si la recherche en sciences sociales doit aider à éclairer le débat public et à lutter contre certaines idées colportées par des militants de tous bords, c'est en apportant aux citoyens des arguments *factuels*, fondés sur une méthode scientifique dénuée d'idéologie.

# Prénoms des descendants d'immigrés en France : Essai de reproduction d'un article scientifique <sup>1</sup>

*Allons bon, on reparle des prénoms en France. Mais cette fois-ci, c'est du lourd, c'est l'Institut national d'études démographiques (Ined).*  
Fil d'actualité du journal [Libération](#), 10 avril 2019

## Introduction

Ce travail présente un essai de reproduction d'un article scientifique portant sur les prénoms des descendants d'immigrés en France. Démographe au Centre national de la recherche scientifique (CNRS), j'ai pris seul l'initiative de conduire cette reproduction et de la mener à terme, si bien qu'elle n'engage que moi et que je suis seul responsable de toute erreur.<sup>2</sup>

Outre la présente introduction (p. 8) et la conclusion (p. 73), le document comprend onze chapitres, les références bibliographiques (p. 75) et onze annexes. Après avoir présenté l'article (1) et sa médiatisation (2), j'expose les raisons qui m'ont conduit à douter de ses résultats (3). Étant donné qu'en violation des règles de l'intégrité scientifique, les auteurs et l'Institut national d'études démographiques (INED) ont refusé de communiquer la méthodologie qui leur a permis de produire leurs résultats (4), j'ai entrepris de reproduire leurs analyses « au plus près » (5 et 6). Aucun des deux résultats les plus spectaculaires de l'article n'est exact (7 et 8), pas plus d'ailleurs que les principaux autres résultats (9). Dans ce contexte, l'article doit être rétracté (10). Il apparaît en outre que l'article est vraisemblablement le fruit d'une fraude scientifique à but idéologique et politique (11).

Ce travail peut se lire de deux façons :

- dans une version synthétique : de la présente introduction à la conclusion, sans les annexes (p. 8-74) ;
- dans sa version détaillée : le chapitre 1 et son Annexe 1 (p. 86), puis le chapitre 2 et son Annexe 2 (p. 92), etc. ; les numéros de page sont cliquables.

Le lecteur qui s'intéresserait aux prénoms en France, mais pas aux méconduites scientifiques de l'INED, peut se contenter de consulter les passages suivants : Annexes 5A (p. 165-182), 6A (p. 267-284) et 6B (p. 296-326), puis chapitres 7 (p. 39-45) et 8 (p. 46-51).

---

<sup>1</sup> Ce travail a bénéficié d'une aide de l'État gérée par l'Agence nationale de la Recherche au titre du programme Investissements d'Avenir portant la référence ANR-10-EQPX-17 (Centre d'accès sécurisé aux données — CASD). Il a aussi bénéficié d'un financement du Groupe d'étude des méthodes de l'analyse sociologique de la Sorbonne (GEMASS, UMR8598). Il repose notamment sur l'exploitation de la source statistique suivante : Trajectoires et origines (TeO) — version réduite — 2008, INSEE, INED [producteurs], ADISP [diffuseur].

<sup>2</sup> Je suis démographe, chargé de recherche au [Centre national de la recherche scientifique \(CNRS\)](#) et membre du [Groupe d'étude des méthodes de l'analyse sociologique de la Sorbonne \(GEMASS\)](#). Je remercie le CNRS et mon laboratoire de m'avoir permis de travailler en toute liberté.



## 1. Présentation de l'article reproduit

En avril 2019 a paru [en français](#) et [en anglais](#) un article scientifique de 4 pages intitulé : « [Quels prénoms les immigrés donnent-ils à leurs enfants en France ?](#) », dont une copie se trouve en Annexe 1A, p. 86.

Cet article est paru dans *Population et sociétés*, un bulletin mensuel d'information scientifique (avec comité de rédaction) qui est destiné au grand public, lu par de nombreux enseignants du secondaire et fréquemment repris par les médias. Précisément, *Population et sociétés* est un bulletin scientifique de l'Institut National d'Études Démographiques (INED), un « établissement public à caractère scientifique et technologique » comme le CNRS ou l'INSERM, chargé d'informer le public et les pouvoirs publics sur les questions démographiques.<sup>3</sup> L'article a été rédigé par deux chercheurs : Baptiste Coulmont, sociologue à l'Université Paris 8 au moment de la publication de l'article et spécialiste des prénoms,<sup>4</sup> et Patrick Simon, démographe à l'INED et spécialiste de l'immigration.<sup>5</sup>

L'article part de la question suivante : « Les prénoms donnés par les immigrés à leurs enfants, et ceux que reçoivent à la génération suivante leurs petits-enfants, sont-ils puisés dans le registre des prénoms des pays d'origine, ou sont-ils au contraire similaires à ceux donnés par la population majoritaire ? » (Coulmont et Simon 2019, p. 1). Le choix d'appeler son fils « Jean-Édouard », plutôt que « Carlos » ou « Mohamed », est gratuit, si bien qu'il peut être considéré comme libre : il permet « d'atteindre le goût débarrassé du coût » (Besnard 1991, p. 56). Depuis la fin des années 1970, le choix du prénom est ainsi devenu, pour nombre de chercheurs en sciences sociales (historiens, anthropologues, sociologues, démographes, économistes, linguistes, etc.), un « outil d'évaluation des processus d'acculturation ou d'assimilation de minorités ethniques ou religieuses » (Grange 2016a, p. 7) et un « révélateur de choix identitaires » (Grange 2016b, p. 65), bref un « indicateur d'assimilation » (Gerhards et Hans 2009, p. 1102).<sup>6</sup> Comme le rappellent les auteurs de l'article, qui s'inscrivent dans ce courant de recherche et étudient le choix du prénom comme « un marqueur culturel » (Coulmont et Simon 2019, p. 1), « Beaucoup d'études interprètent comme une mesure de l'assimilation le degré de proximité [des prénoms de descendants d'immigrés] avec ceux de la population majoritaire (les personnes qui n'ont pas d'ascendance immigrée sur deux générations) » (Coulmont et Simon 2019, p. 1). Dans cette perspective, « l'assimilation intergénérationnelle [...] passe par l'abandon des prénoms du pays d'origine et l'adoption, pour ses enfants, des prénoms du pays d'accueil » (Coulmont 2017a, p. 1).

---

<sup>3</sup> Le Décret n°86-382 du 12 mars 1986 portant organisation et fonctionnement de l'Institut national d'études démographiques dispose : « L'Institut national d'études démographiques a les missions suivantes : [...] Il tient notamment le Gouvernement et les pouvoirs publics informés des connaissances acquises et développe une capacité d'expertise et d'appui aux politiques publiques menées pour répondre en particulier aux défis sociétaux et éducatifs dans son champ d'activité ; [...] Il assure l'information du public sur les questions démographiques » (Titre 1<sup>er</sup>, Article 3).

<sup>4</sup> Baptiste Coulmont était, lors de la publication de l'article, Professeur des universités à l'Université Paris 8 Vincennes-Saint-Denis, affilié au Centre de recherches sociologiques et politiques de Paris (CRESPPA). Depuis septembre 2020, il est Professeur des universités à l'École Normale Supérieure Paris-Saclay. Il est aussi [chroniqueur au journal Le Monde](#).

<sup>5</sup> Patrick Simon est Directeur de recherche à l'INED, directeur de département à l'Institut Convergences Migrations et chercheur associé à Sciences Po Paris (au Centres d'études européennes puis à l'Observatoire sociologique du changement).

<sup>6</sup> J'utilise indifféremment les termes « assimilation », « intégration » et « acculturation » pour désigner la réduction intergénérationnelle des différences entre individus d'origines diverses, tout en considérant que le choix du prénom n'est qu'un révélateur parmi d'autres de ces processus, de pair avec la réduction intergénérationnelle de la ségrégation spatiale, des disparités économiques (métiers, revenus et patrimoine), des relations sociales (amitiés et mariages) et des pratiques culturelles (langue, religion, vêtement, nourriture, musique, etc.). Sur ces notions, voir Héran 2020.

Reposant sur une exploitation des données statistiques de l'enquête *Trajectoires et origines* (TeO) menée par l'INED et l'INSEE en 2008 (Annexe 1B, p. 91), l'article est ainsi résumé :

« La convergence entre population majoritaire et descendants d'immigrés ne se fait pas autour de prénoms typiquement « français », mais de prénoms internationaux auxquels tous et toutes peuvent s'identifier. Pour les enfants ou petits-enfants d'immigrés d'origine maghrébine, des prénoms arabo-musulmans continuent en partie à être donnés ainsi que de nouveaux prénoms (Yanis, Rayane ou Lina) perçus comme d'origine maghrébine par la population majoritaire » (Coulmont et Simon 2019, p. 4).

À la lecture de l'article, au moins deux de ses principaux résultats peuvent apparaître comme particulièrement spectaculaires :

- d'une part, en France métropolitaine en 2008, parmi les petits-fils d'immigrés du Maghreb, les deux prénoms les plus fréquemment portés sont « Yanis » et « Nicolas » (Coulmont et Simon 2019, tableau p. 3 ; c'est moi qui souligne « Nicolas ») ; ainsi, « Les prénoms que reçoivent les petits-enfants [d'immigrés du Maghreb] sont, en 2008, proches de ceux que la population majoritaire donne à ses enfants » (Coulmont et Simon 2019, p. 3) ;
- d'autre part, en France métropolitaine en 2008, alors que 94 % des *immigrés* du Maghreb (« 1<sup>e</sup> génération ») et 84 % des *enfants* d'immigrés du Maghreb (« 2<sup>e</sup> génération ») portent un prénom « arabo-musulman », les *petits-enfants* d'immigrés du Maghreb (« 3<sup>e</sup> génération ») ne sont que 23 % à porter un prénom « arabo-musulman », un ordre de grandeur assez proche des 16 % de petits-enfants d'immigrés d'Europe du Sud qui portent un prénom « latin » (Coulmont et Simon 2019, figure 2, p. 3) ; en conséquence, sur trois « générations » (des immigrés à leurs petits-enfants), les descendants d'immigrés du Maghreb abandonnent les prénoms de leurs pays d'origine (Algérie, Maroc, Tunisie) presque autant que les descendants d'immigrés d'Europe du Sud (Italie, Espagne, Portugal) (Coulmont et Simon 2019, figure 2, p. 3) ; ainsi, « La trajectoire suivie par les originaires du Maghreb mène au même point d'arrivée que celle suivie par les Européens du Sud, mais de manière différée » (Coulmont et Simon 2019, p. 3).

## 2. Des résultats largement médiatisés

Les journalistes ne s'y sont pas trompés, qui ont largement relayé ces spectaculaires résultats (Annexe 2, p. 92) :

- dans des journaux nationaux comme [Le Monde](#), [Le Figaro](#), [Libération](#), [La Croix](#), [20 Minutes](#), [L'Obs](#), [L'Express](#) ou [Valeurs actuelles](#) ;
- dans la presse régionale : [Le Parisien](#), [Ouest France](#), [L'Est républicain](#), [Le Républicain lorrain](#), [Les dernières nouvelles d'Alsace](#), [Le Progrès](#), [Le Dauphiné](#), [La Provence](#), [Corse-Matin](#), etc. ;
- dans la presse étrangère : [El Watan](#), [The Economist](#), [Inside Over](#) ;
- sur des sites Internet d'information : [Huffington Post](#), [France Soir](#), [Le petit journal](#), [Le journal des femmes](#), [vie-publique](#), [Le café pédagogique](#), [l'ENS Lyon](#), [les étudiants du CFJ](#), [Yabiladi](#) et [Oumma.com](#) ;
- sur des radios : [France Inter](#), [France Info](#), [RTL](#), [RMC](#), [Europe 1](#) et [RFI](#) ; et enfin
- sur des chaînes télévisées : [France 2](#), [M6](#), [LCI](#) et [Africa news](#).

Naturellement, les auteurs de l'étude n'ont pas été les derniers à diffuser leurs résultats les plus forts, que ce soit sur [CNEWS](#), [RMC](#) ou [BFM TV](#).

La presse a globalement repris l'article de façon fidèle à son contenu, comme en témoignent les titres d'articles et d'émissions, qui ont pu porter sur :

- la question de recherche : « Quels prénoms les immigrés donnent-ils à leurs descendants ? » ([20 Minutes](#)) ; « Enzo, Nicolas, Inès... Quels prénoms portent les enfants d'immigrés ? » ([Ouest France](#)) ;
- l'utilisation du choix du prénom comme indicateur de degré de désir d'assimilation : « Que révèlent les prénoms des enfants d'immigrés en France ? » ([Le Parisien](#)) ; « **Ce que les prénoms révèlent de l'intégration en France** » ([Le Figaro](#)) ; « La GG du jour : Les prénoms, marqueurs d'intégration ? » ([RMC](#)) ;
- les résultats de l'article : « Yanis, Sarah ou Inès : quand l'intégration se lit dans les prénoms des petits-enfants d'immigrés » ([Le Monde](#)) ; « Après deux générations, les enfants d'immigrés abandonnent les prénoms du pays de leurs parents » ([L'Obs](#)) ; « Choix des prénoms : "Les petits-enfants d'immigrés ont peu souvent des prénoms en référence au pays d'origine de leurs grands-parents" » ([France Info](#)) ; « **ces prénoms qui racontent l'intégration** », une enquête « **qui démontre l'intégration des populations d'origine immigrée par le prénom** » ([Europe 1](#)) ; « Les immigrés privilégient les prénoms "internationaux" pour leurs enfants » ([L'Express](#)) ; « France : De plus en plus de Maghrébins prénomment leurs enfants Yanis ou Nicolas » ([Yabiladi](#)) ;
- l'apport de ces résultats au débat public : [LCI](#) met en avant « **Des chiffres qui tordent le cou à plusieurs idées reçues** » ; Pascal Praud, sur [CNEWS](#), observe que l'étude « va a contrario des idées reçues » ; et [La Croix](#) observe que « l'étude que publie l'institut national d'études démographiques (Ined) mercredi 10 avril va mettre à mal certains clichés ».

En résumé, la présentation dominante de l'article est assez bien illustrée par celle qu'en a donné le journal [Le Monde](#), dans son article intitulé « Yanis, Sarah ou Inès : quand l'intégration se lit dans les prénoms des petits-enfants d'immigrés » :

« Leurs grands-parents s'appelaient Farid, Ahmed et Rachid, Fatima, Fatiha et Khadija. Les petits-enfants s'appellent désormais Yanis, Nicolas et Mehdi, Sarah, Inès et Lina. Ce sont les prénoms les plus portés par les Français descendants d'immigrés originaires du Maghreb. Contrairement à certaines idées reçues, « *les petits-enfants de la première génération s'appellent Abdelkader ou Karim sont très largement minoritaires* », explique Baptiste Coulmont. »

### 3. Comme un doute...

Quelle qu'ait été l'ampleur de leur exposition médiatique, les résultats présentés dans cet article sont surprenants. Si en France en 2008, entre la 2<sup>e</sup> et la 3<sup>e</sup> « générations » de descendants d'immigrés<sup>7</sup> du Maghreb, la part de prénoms « arabo-musulmans » passait de 84 %<sup>8</sup> à 23 %, cela constituerait indéniablement un phénomène social majeur. Les spécialistes, les observateurs et les Français en général, y compris l'auteur de ces lignes, seraient-ils passés à côté d'un phénomène aussi massif que révélateur ? Est-il envisageable qu'en 2008, seulement 23 % des petits-enfants d'immigrés du Maghreb portent un prénom « arabo-musulman », alors qu'à la même date plus de 10 % de la *totalité* des naissances annuelles en France se voyaient attribuer un prénom « arabo-musulman » (Fourquet et Manternach 2019a, p. 136-140) ?<sup>9</sup>

Par ailleurs, serait-il possible qu'en France en 2008 le prénom le plus fréquent chez les petits-fils d'immigrés du Maghreb soit, derrière « Yanis », « Nicolas » ? De fait,

« dans un département comme la Seine-Saint-Denis, où cette troisième génération [issue de l'immigration maghrébine] est numériquement très représentée, le [fichier de l'Insee](#) [qui contient les données exhaustives d'état civil en France sur les prénoms attribués chaque année aux nouveau-nés] ne répertorie que très peu de naissances de Nicolas depuis vingt ans » (Fourquet et Manternach 2019b, p. 124).

Précisément, si on examine les prénoms attribués aux naissances annuelles en Seine-Saint-Denis de 1980 à 2014, le nombre de « Nicolas » n'a cessé de baisser (de 450 en 1980 à moins de 50 dans les années 2010), si bien que depuis les alentours de l'an 2000 on compte chaque année en Seine-Saint-Denis moins de naissances de « Nicolas » que de « Mohamed », « Yan(n)is », « Rayan(e) » ou encore « Adam » (de 100 à 300 pour chacun de ces prénoms des alentours de l'an 2000 aux années 2010) (Fourquet et Manternach 2019b, p. 124). Si les nouveau-nés comptent relativement peu de « Nicolas » en Seine-Saint-Denis, où est-ce que les petits-fils d'immigrés du Maghreb sont souvent prénommés « Nicolas » ?

Pour pouvoir admettre l'exactitude de résultats aussi contre-intuitifs, il est indispensable de connaître et de comprendre la méthodologie qui a permis de les produire. Or, les auteurs ont choisi de publier ces résultats dans une publication de seulement 4 pages, à l'encadré méthodologique nécessairement minimaliste (Coulmont et Simon 2019, p. 2). Il m'est donc apparu comme un *décalage* entre d'une part l'importance et le caractère contre-intuitif des résultats, et d'autre part — on va le voir — le flou de la méthodologie qui conditionne leur fiabilité.

---

<sup>7</sup> Un descendant d'immigrés est un individu qui a au moins un parent ou grand-parent immigré, c'est-à-dire né étranger à l'étranger.

<sup>8</sup> La proportion d'enfants d'immigrés du Maghreb qui portent un prénom spécifiquement « arabo-musulman » est-elle d'environ 84 % comme l'indique la Figure 2 de l'article, ou de « près des deux tiers » comme semble l'indiquer le texte (Coulmont et Simon 2019, p. 3) ? À défaut de pouvoir le savoir, je reprends les informations de la Figure 2.

<sup>9</sup> La part des naissances annuelles masculines en France auxquelles est attribué un prénom « arabo-musulman » (l'un des 4 550 prénoms masculins « arabo-musulmans » recensés par Fourquet et Manternach et indicateurs d'origine culturelle arabo-musulmane) est proche de 0 % jusqu'en 1950 puis dépasse 2 % dans les années 1960, 5 % dans les années 1970, 10 % dans les années 2000 et atteint 19 % en 2016 (Fourquet et Manternach 2019, p. 136-140). D'après d'autres sources, « Between 1970 and 2000, the number of Muslims and other immigrants skyrocketed in France. In 2010, Jean-Baptiste came second as *prénom composé* to Mohamed-Amine (Fatima-Zahra was fourth for girls). In 2011, Mohamed was the most common name given to newborns in Marseilles » (Lawson 2016, p. 175).

### ❖ Une accumulation d'inexactitudes et d'omissions

Comme l'article n'est accompagné d'aucune documentation annexe,<sup>10</sup> les seules informations méthodologiques disponibles sont présentées dans un encadré (Coulmont et Simon 2019, p. 2). Le premier paragraphe de l'encadré porte sur l'enquête TeO, une enquête statistique qui renseigne le prénom et de nombreuses autres caractéristiques de près de 22 000 adultes de 18-60 ans résidant en ménage ordinaire en France métropolitaine en 2008, ainsi que le prénom des enfants de l'enquêté (qu'ils résident ou non avec l'enquêté) et le prénom des autres habitants partageant son logement. Ce paragraphe de l'encadré méthodologique permet de comprendre que, dans leur article, les auteurs comparent les prénoms portés par trois « générations » d'immigrés ou de descendants d'immigrés : tout d'abord des immigrés, qui par définition sont nés étrangers à l'étranger (1<sup>e</sup> génération ou G1) ; puis des enfants d'au moins un immigré (2<sup>e</sup> génération ou G2) ; et enfin des petits-enfants d'au moins un immigré (3<sup>e</sup> génération ou G3). Toutefois, ce paragraphe ne renseigne pas précisément le champ des analyses, ni l'effectif d'individus et de prénoms analysés, ni les variables utilisées — et recodées — pour réaliser les analyses, ni l'utilisation des pondérations, si bien que concrètement un chercheur qui souhaiterait reproduire ces résultats ne le pourrait pas sans devoir demander aux auteurs divers compléments d'information.

Le second paragraphe de l'encadré méthodologique porte sur le classement des prénoms par origine et permet de comprendre que, pour associer chaque prénom à son origine linguistique ou culturelle, les auteurs recourent à « une classification linguistique probabiliste qui catégorise les prénoms en "groupes culturels" », et qui consiste à

« relier chaque prénom à une catégorie linguistique et culturelle à partir d'une classification probabiliste des référents linguistiques des syllabes. Cette méthode automatisée peut également être appliquée aux noms de famille, qui sont plus stables que les prénoms. Le fait que des schèmes linguistiques soient spécifiques à des aires culturelles se retrouve dans les prénoms et autorise cette attribution » (Coulmont et Simon 2019, p. 2).

Sont ainsi catégorisés comme « français » les « prénoms dont les consonances sont caractéristiques des schèmes linguistiques français » (Coulmont et Simon 2019, p. 2), et comme « inclassables » les prénoms dont « l'analyse des syllabes n'a pas permis d'associer le prénom à une catégorie connue » (Coulmont et Simon 2019, p. 2). Cette méthodologie novatrice<sup>11</sup> est intuitivement prometteuse, mais comment fonctionne-t-elle ? Quelles sont précisément les propriétés linguistiques des syllabes qui permettent de distinguer des prénoms « français », « latins », « arabo-musulmans », « africains », « asiatiques », « internationaux ou anglophones », « autre groupe » et « inclassables » ?<sup>12</sup> Concrètement, comment ont été classifiés, par exemple, chacun des prénoms du tableau présenté dans l'article (Coulmont et Simon 2019, p. 3) ? Et comment se fait-il que cette méthodologie « syllabique », indéniablement utile, n'ait jamais été utilisée avant ni depuis l'article de Coulmont et Simon (2019), ni par eux ni par quiconque d'autre ?

<sup>10</sup> Tel n'est pas le cas d'autres publications de l'INED issues de l'enquête TeO, qui ont une documentation annexe comme les [annexes de Beauchemin et al. 2016](#).

<sup>11</sup> La méthode « syllabique » d'attribution d'une origine culturelle aux prénoms peut être considérée comme novatrice parce que Baptiste Coulmont n'y fait pas référence dans son ouvrage de 2014 sur la sociologie des prénoms (Coulmont 2014, p. 72-85), alors qu'il mentionne bien la méthode par « listes » (sous le terme « familles de prénoms ») ainsi que la méthode du « score de distance » (sous le terme « indice de francité ») (Coulmont 2014, p. 74). Même dans une présentation datant de 2017, Baptiste Coulmont ne mentionne pas cette méthode « syllabique » : au contraire, il indique qu'« Il existe deux manières principales de traiter les prénoms » (Coulmont 2017a, p. 1), à savoir la méthode par listes et l'indice susmentionné. On peut donc envisager que Coulmont et Simon auraient conçu ou découvert la méthode « syllabique » au cours de l'année 2018.

<sup>12</sup> Le libellé et le contenu de ces catégories d'origine des prénoms sont examinés en détail dans l'Annexe 6A, p. 267 et dans l'Annexe 6B, p. 285.



C'est pour connaître et comprendre précisément la méthodologie de l'article que, quatre jours après sa publication, j'ai contacté l'un de ses auteurs (message électronique du 14 avril 2019, Annexe 4A, p. 94). La discussion orale que j'ai eue avec Baptiste Coulmont (17 avril 2019) m'a appris certaines informations utiles : la base de données compte environ 105 000 individus dont l'enquête TeO renseigne le prénom, et ces individus portent 18 163 prénoms différents. Mais curieusement, pour attribuer une origine linguistique aux prénoms, Coulmont et Simon n'utilisent pas l'une des méthodes qu'ils citent dans leur article (Decharne et Liedts 2007 ou encore Mazières et Roth 2018<sup>13</sup>) : ils utilisent une méthode achetée par l'INED à [l'entreprise allemande Humpert et Schneiderheinze](#) (Humpert et Schneiderheinze 2000), que leur article omet de mentionner. En outre, contrairement à ce que laisse entendre l'encadré méthodologique, cette méthode n'est pas de nature linguistique : elle ne se fonde pas sur l'analyse des morphèmes ou phonèmes ou n-grams ou syllabes ou « structures syllabiques » des prénoms, mais bien sur des listes et répertoires de prénoms par origine géographique, linguistique ou culturelle (Humpert et Schneiderheinze 2000).<sup>14</sup> Enfin, contrairement à ce que laisse encore entendre l'encadré méthodologique, cette méthode est loin d'être entièrement « automatique » : elle produit un classement des prénoms en plusieurs dizaines de modalités, que Coulmont et Simon ont dû eux-mêmes regrouper en sept ou huit catégories, selon une procédure dont l'encadré ne précise rien.

### ❖ Résumé

L'accumulation d'inexactitudes et d'omissions dans l'encadré méthodologique de l'article de Coulmont et Simon (2019, p. 2) peut paraître troublante, car la méthode d'attribution d'une origine linguistique aux prénoms conditionne directement le résultat selon lequel, de la 2<sup>e</sup> à la 3<sup>e</sup> génération de descendants d'immigrés du Maghreb, la proportion de prénoms dits « arabo-musulmans » passerait de 84 % à seulement 23 %. Cela a confirmé mon sentiment qu'un réexamen de cette étude pourrait être nécessaire.

---

<sup>13</sup> Mazières et Roth (2018) élaborent un modèle statistique (dont le code est disponible [ici](#)) permettant de catégoriser n'importe quel nom de famille selon son origine géolinguistique probable. Pour cela, ils procèdent de la façon suivante : (i) ils associent plusieurs centaines de milliers de noms de famille au pays du monde où chacun de ces noms est relativement le plus fréquent, grâce à la base de données comprenant le nom et le pays d'affiliation de chacun des auteurs des 25 millions de publications de PubMed ; (ii) ils regroupent les pays d'origine de ces noms en sept grandes catégories géolinguistiques (étiquetées respectivement « européenne méridionale et centrale », « européenne septentrionale », « slave », « arabe », « indienne », « asiatique » et « africaine »), distinctes les unes des autres selon la fréquence relative des suites de lettres (n-grams) qui composent le plus couramment les noms ; (iii) leur modèle statistique permet alors de catégoriser n'importe quel nom de famille (y compris un nom de famille qui n'est porté par aucun auteur de PubMed) selon l'aire géolinguistique dont il est le plus probablement issu, étant donné la suite de lettres qui le compose.

<sup>14</sup> J'ai envoyé des messages électroniques à Humpert et Schneiderheinze (le 18 avril et le 14 mai 2019, puis le 15 février 2020) pour connaître précisément leur méthode d'attribution d'une origine linguistique aux prénoms. Ils m'ont communiqué des articles décrivant leur méthode (Humpert et Schneiderheinze 2000 ; Humpert et Schneiderheinze 2013 ; Liebau et al. 2018) mais ils ont refusé de me communiquer leur table de passage des 18 163 prénoms présents dans l'enquête TeO à leur origine linguistique détaillée, m'invitant plutôt à contacter Patrick Simon et l'INED.



#### 4. Une méthodologie inaccessible

La base de données individuelles qu'exploitent Coulmont et Simon (2019) est accessible aux chercheurs qui en font la demande.<sup>15</sup> Dans ce contexte, pour examiner la solidité de leurs résultats, seuls deux éléments méthodologiques sont nécessaires : le programme informatique utilisé par les auteurs, et leur nomenclature des prénoms par origine.

##### ❖ Ce que je demande aux auteurs et à l'INED

Le **programme informatique** des auteurs est le document qui contient les commandes de traitement statistique qui, une fois appliquées aux bases de données brutes accessibles aux chercheurs, ont permis aux auteurs de produire les résultats publiés. Ce programme informatique est indispensable — et suffisant — pour pouvoir reproduire à l'identique le résultat selon lequel les deux prénoms les plus fréquents chez les petits-fils d'immigrés du Maghreb sont « Yanis » et « Nicolas ». Ce programme précise notamment quels individus de quelles tables les auteurs étudient, s'ils les ont pondérés et comment, et de quelle manière ils ont recodé diverses variables pour ensuite réaliser leurs analyses et obtenir leurs résultats.

La table de **nomenclature des prénoms par origine** devrait se présenter comme l'indique le tableau 4.1, avec en lignes 18 163 prénoms ; dans une première colonne, la catégorisation de chaque prénom selon la méthode de Humpert et Schneiderheinze (2000) (plusieurs dizaines de catégories) ; et dans une seconde colonne, la catégorisation du même prénom selon la méthode de Coulmont et Simon (2019), en 8 catégories (« français », « arabo-musulman », « latin », « africain », « asiatique », « internationaux ou anglophones », « autre groupe » et « inclassable »). Cette double table de passage est nécessaire, en plus du programme informatique, pour reproduire à l'identique le résultat selon lequel, de la 2<sup>e</sup> à la 3<sup>e</sup> « génération » de descendants d'immigrés du Maghreb, la proportion de prénoms dits « arabo-musulmans » passerait de 84 % à seulement 23 %.

**Tableau 4.1. Structure de la double table de passage de chaque prénom à son origine**

	Catégorie linguistique <b>détaillée</b> (en plusieurs dizaines de modalités) selon la méthodologie de Humpert et Schneiderheinze (2000)	Catégorie linguistique <b>regroupée</b> (en 8 modalités) selon la méthodologie de Coulmont et Simon (2019)
Prénom 1		
Prénom 2		
...		
Prénom 18 163		

<sup>15</sup> Les tables de l'enquête TeO sont accessibles grâce aux [Archives de données issues de la statistique publique](#) (ADISP), et les tables TeO *nominatives*, c'est-à-dire enrichies de la variable « prénom », sont accessibles par le biais du [Centre d'accès sécurisé aux données](#) (CASD) après accord du [Comité du secret statistique](#). Cet accès est rendu possible par la mise à disposition des données d'enquête quantitatives, à laquelle contribuent l'INSEE et d'autres institutions de la statistique publique, y compris [l'INED](#). D'après mon expérience, la procédure d'obtention de données statistiques individuelles *nominatives* (comme l'enquête TeO enrichie de la variable « prénom ») est, par rapport à l'obtention de données individuelles *non nominatives*, très longue et relativement coûteuse (car elle est ultra-sécurisée), mais elle ne pose pas d'autres difficultés majeures. En demandant la communication d'une base de données, le chercheur s'engage à l'utiliser exclusivement dans une finalité de recherche ; à ne pas céder ces données, sous quelque forme que ce soit, à une tierce personne ; à traiter ces données conformément aux règles de l'art ; à mentionner la source des données dans ses publications ; à détruire les fichiers à l'issue du travail de recherche ; etc.

Cette double table est aussi nécessaire pour évaluer si les classifications de Humpert et Schneiderheinze et les regroupements de Coulmont et Simon sont raisonnables. Si des chercheurs refusaient de communiquer leur nomenclature des prénoms par origine, on ne pourrait totalement exclure que le prénom « Mohamed » y soit classifié comme « asiatique » (et non pas « arabo-musulman ») au motif que ce prénom serait originaire de l'Arabie, qui se trouve en Asie, ni que le prénom « Aminata » soit lui aussi classifié comme « asiatique » au motif que sa terminaison sonnerait japonais. On ne saurait non plus exclure que les auteurs aient fait le choix de classifier les prénoms donnés aux petits-enfants d'immigrés du Maghreb en prénoms « français », « internationaux », « autres » ou « inclassables », dans l'objectif de gonfler le degré auquel les descendants d'immigrés du Maghreb abandonnent les prénoms dits « arabo-musulmans ».

Comme chacun peut le constater en examinant la structure de la double table de passage ci-dessus (tableau 4.1), celle-ci ne porte pas sur des individus (ni a fortiori sur des individus enquêtés par la statistique publique et susceptibles d'être ré-identifiables), mais sur des prénoms. Par conséquent, cette double table de passage n'est pas soumise au secret statistique<sup>16</sup> et rien ne s'oppose à ce qu'elle soit communiquée à un collègue qui en ferait la demande. Bien au contraire, il est largement considéré comme de bonne pratique scientifique ([Charte nationale de déontologie des métiers de la recherche](#) ; [Office français de l'intégrité scientifique](#) ; Corvol 2019 ; Pouzat et al. 2020), y compris en sciences sociales et au moins dans les disciplines quantitatives comme la [démographie](#), de communiquer en toute transparence, aux collègues qui en font la demande, les programmes informatiques et autres éléments méthodologiques qui ont permis de produire — et permettent de reproduire — des résultats publiés. C'est tout particulièrement le cas lorsque ces programmes peuvent être appliqués à des bases de données qui sont communes aux chercheurs (recensement, état civil, enquête nationale comme TeO, enquête internationale). Cette pratique est nécessaire pour assurer la traçabilité et la reproductibilité et donc la *fiabilité* des publications scientifiques. Ainsi le 18 mai 2017, l'unique fois où un collègue — d'ailleurs associé à l'INED — m'a demandé de lui communiquer le programme informatique nécessaire à la reproduction d'une de mes publications (Mignot 2010), je lui ai communiqué ledit programme de façon inconditionnelle et *dans la journée même* : rien que de très normal dans le cadre de bonnes pratiques scientifiques.

#### ❖ Le refus des auteurs de communiquer leur méthodologie

L'écart entre les inexactitudes et omissions méthodologiques entourant l'article et le caractère spectaculaire de ses résultats m'a amené, pour la première fois de ma carrière, à demander à des collègues les éléments nécessaires et suffisants pour reproduire à l'identique leur publication. Précisément, je demande explicitement la double table de passage des prénoms à leurs origines depuis un message électronique du 18 avril 2019 (soit huit jours après la publication de l'article), et le programme informatique depuis un message du 30 septembre 2019. Les échanges de messages électroniques que j'ai eus avec les auteurs et avec l'INED, du 14 avril 2019 au 18 décembre 2020, figurent en Annexe 4A, p. 94.

Suite à mes questions méthodologiques, les auteurs ont commencé par apporter certaines réponses (discussion orale du 17 avril 2019 avec Baptiste Coulmont, résumée dans son message du 14 mai 2019). Mais ils ont rapidement éludé ma demande de communication de leur nomenclature, d'abord en ne répondant pas à mes messages dans le délai auquel ils s'étaient

---

<sup>16</sup> En France, le secret statistique permet d'apporter aux personnes qui fournissent des informations utilisées pour l'établissement de statistiques l'assurance que ces informations ne seront pas utilisées d'une façon susceptible de leur porter tort : il garantit la confidentialité de la vie privée, personnelle et familiale des personnes enquêtées, et le respect du secret commercial et des affaires des entreprises.

engagés (d'où mon message de relance du 6 mai 2019), puis en n'apportant de réponses qu'à des questions que je n'avais pas posées (message de Baptiste Coulmont du 13 mai 2019).

Les auteurs ont ensuite motivé leur refus de communiquer leur nomenclature par le fait qu'ils n'auraient pas *l'autorisation* de la communiquer (messages de Baptiste Coulmont et de Patrick Simon du 14 mai 2019). Or comme on l'a vu, une table de passage de prénoms à leur origine ne portant pas sur des individus mais sur des prénoms, elle n'est pas susceptible d'être utilisée de façon qui puisse porter tort à des individus, si bien qu'elle n'est pas soumise au secret statistique et que rien n'empêche de la communiquer immédiatement à des collègues. Quand j'ai demandé à Baptiste Coulmont sa double table de passage il m'a répondu, en substance, que pour reproduire les analyses de son article, il faudrait passer par une demande d'accès aux données nominatives par le biais du Comité du secret statistique (message du 13 mai 2019), avant que Patrick Simon ne confirme cette assertion (message du 14 mai 2019). Pour accéder aux données de l'enquête TeO enrichies de la variable « prénom », il est effectivement nécessaire d'en faire la demande au [Comité du secret statistique](#) puis au [Centre d'accès sécurisé aux données](#) car l'accès à des données statistiques individuelles directement *nominatives* pourrait permettre à un chercheur indélicat de réidentifier un enquêté et de chercher à lui porter un tort. Mais en quoi ce fait relatif aux conditions d'accès à la *base de données* empêche-t-il un chercheur de communiquer la *nomenclature des prénoms par origine* par laquelle il a exploité ladite base de données ?

Lorsque j'ai fait observer à Baptiste Coulmont et Patrick Simon que, contrairement à ce qu'ils avançaient dans leurs messages des 13 et 14 mai 2019, l'accès à leur double table de passage n'était pas soumis au secret statistique et leur table était donc légalement communicable (message du 16 mai 2019), ils n'ont d'abord pas répondu (d'où mes messages de relance du 23 mai et du 3 juin 2019) puis, lorsque Patrick Simon a finalement répondu, il s'est engagé à me communiquer sa double table plus de trois mois plus tard, en septembre 2019 : « je vous propose d'attendre que vous ayez obtenu l'accord du comité du secret et que nous ayons soumis nos publications (à horizon de la rentrée en septembre) pour que je vous transmette une table avec un identifiant des enquêtés et la position des prénoms dans les classifications que nous avons opérées. Vous citerez bien entendu la source de ces classifications dans les publications qui seront issues de vos travaux » (message du 5 juin 2019). Je me suis donc armé de patience et j'ai attendu pendant trois mois. Mais trois mois plus tard, Patrick Simon n'a pas honoré son engagement : « Nous avons pris du retard dans la rédaction de l'article. Je ne manquerai de vous tenir informé de la mise à disposition des variables quand nous a[u]rons terminé » (message du 25 septembre 2019).<sup>17</sup>

Les auteurs ont alors motivé leur refus de communiquer leur nomenclature par le fait qu'ils n'auraient pas *l'obligation* de communiquer une méthode dont l'INED est propriétaire et par le fait que, par ailleurs, la table serait mise à disposition par le biais du [réseau Quételet](#) en avril 2020 (messages de Patrick Simon du 27 septembre 2019 et du 7 mars 2020). D'après Patrick Simon, le droit de propriété le dispense de communiquer sa nomenclature (message du 7 mars 2020).

Baptiste Coulmont et Patrick Simon refusent donc — pour des motifs variables, mais avec constance — de communiquer leur méthode d'attribution d'une origine aux prénoms, alors que les trois autres groupes de chercheurs auxquels j'ai adressé la même demande l'ont tous acceptée : il s'agit de Head et Mayer 2008 et Algan et al. 2013 pour le premier groupe ; de Behaghel et al. 2015 pour le deuxième ; et de Fourquet et Manternach 2019a pour le troisième.<sup>18</sup>

---

<sup>17</sup> À ce jour (août 2021), cet article n'a toujours pas été publié.

<sup>18</sup> Je tiens à remercier respectivement Clément Malgouyres, Thomas Le Barbanchon et Sylvain Manternach — ainsi que tous leurs coauteurs — d'avoir accepté de me communiquer leurs nomenclatures de prénoms par origine. Les diverses thèses soutenues dans le présent document n'engagent que moi, et je suis seul responsable de toute erreur.

Notons au passage que le refus de communiquer cette nomenclature empêche non seulement d'évaluer la fiabilité de leurs résultats : cela empêche aussi d'appliquer leur nomenclature à d'autres bases de données que l'enquête TeO (comme le [fichier INSEE des prénoms](#)), ce qui ralentit le progrès des connaissances.

Concernant leur programme informatique, les auteurs n'ont jamais jugé utile de justifier leur refus de le communiquer. Répétée de façon parfaitement explicite à de nombreuses reprises depuis le 30 septembre 2019, cette demande n'a jamais reçu aucune réponse de leur part. Pourquoi Baptiste Coulmont refuse-t-il de communiquer cette méthodologie, lui qui publie sur [son compte GitHub](#) des méthodologies qui ne lui ont pas été demandées ?

En 2016, quand une enquête de l'Institut Montaigne montrait qu'en France une proportion non négligeable des adultes musulmans exprimaient des valeurs à l'opposé de celles de la plupart du reste de la société française (El Karoui et Jardin 2016), Patrick Simon a répondu dans le journal *Le Monde* par une tribune intitulée « [Arrêtons de faire dire aux musulmans ce qu'ils ne pensent pas](#) » (2016). Reprochant à cette enquête d'être « entachée d'approximations et d'opacité méthodologique », il affirmait :

« Le questionnaire n'est pas fourni avec le rapport, et les annexes méthodologiques sont pauvres. Contrairement à ce qui est annoncé, il n'est pas possible de retrouver comment les typologies ont été construites, ni de reproduire les résultats pour en vérifier la solidité. C'est embêtant, car il est répété à plusieurs reprises que c'est un travail scientifique à la méthodologie irréprochable et rien ne le prouve. »

On ne saurait mieux dire.

#### **Encadré 4.2. Le point de vue de Patrick Simon**

Dans un entretien accordé à la revue *Tracés* publié en juillet 2020 et qui porte sur le service des enquêtes de l'INED (Beauchemin et al. 2019), Patrick Simon aborde incidemment la demande que je lui ai adressée de me communiquer la méthodologie de son article. Je reproduis ici ses déclarations, avant d'y apporter une réponse.

*TRACES : Ce qui ressort ici, c'est donc un risque de mésusages des données. Mais est-ce que cette question ne se pose pas aussi avec des chercheurs ?*

P. SIMON : Bien sûr que si ! Je vais donner un exemple, qui renvoie à cette question du niveau de données à rendre accessible. À la suite du papier sur les transmissions des prénoms que nous avons publié avec Baptiste Coulmont [Coulmont et Simon 2019], un chercheur a eu envie de reproduire ce qu'on a fait parce qu'il n'était pas convaincu par les résultats. Il a voulu faire une contre-analyse. Assez hostile. Les prénoms eux-mêmes étaient protégés par le Comité du secret statistique, mais ce chercheur a fait une demande au Comité du secret pour les obtenir. Puis il m'a contacté pour me demander notre système de classification des prénoms. Il m'a donc demandé un outil de recherche et a justifié sa demande par un objectif de répliquabilité de l'étude, comme s'il s'agissait d'une obligation de permettre cette réplification. C'est une question déontologique et scientifique passionnante à laquelle je n'ai pas de réponse, mais l'INED l'a fait étudier sur le plan juridique et déontologique. La classification que nous avons fait faire est considérée comme un brevet et doit faire l'objet d'une licence d'exploitation. On voit qu'on sort progressivement de l'ère du bricolage pour entrer dans des systèmes juridiques et déontologiques de production, d'exploitation et de mise à disposition des données plus normés. C'est encore une autre étape de réflexion : dans les strates successives de contribution, il y a ce qui doit être diffusé et ce qui relève de ta création scientifique, qui t'est personnelle.

Cette façon de présenter nos échanges appelle plusieurs observations. D'une part, j'ai demandé à Patrick Simon non seulement sa nomenclature des prénoms par origine, mais aussi son *programme informatique*, ce qu'il omet de mentionner. Or ni lui ni Baptiste Coulmont ni l'INED n'ont jamais justifié leur refus de communiquer ce programme informatique (Annexe 4A, p. 94). La question est donc toujours ouverte : pourquoi refuser de communiquer le programme informatique qui leur a permis de produire les résultats qu'ils ont publiés, et que l'intégrité scientifique exigerait qu'ils communiquent à tout collègue qui en fait la demande ?

D'autre part, concernant ma demande de communication de sa nomenclature des prénoms par origine, Patrick Simon indique que s'il ne me l'a pas communiquée, c'est parce que l'analyse juridique et déontologique de l'INED lui a interdit de la communiquer. Or là encore, il omet de mentionner plusieurs faits. Il s'est initialement engagé à me communiquer sa nomenclature (message du 5 juin 2019), et lorsque trois mois plus tard il a manqué d'honorer son engagement, il a avancé la raison suivante : « Nous avons pris du retard dans la rédaction de l'article. Je ne manquerai de vous tenir informé de la mise à disposition des variables quand nous a[u]rons terminé » (message du 25 septembre 2019). Ce n'est que deux jours plus tard, lorsque j'ai averti le rédacteur en chef de *Population et sociétés* que les auteurs d'un des articles de cette revue refusaient de communiquer leur méthodologie, que Patrick Simon a pour la première fois mis en avant des raisons juridico-déontologiques pour motiver son refus de communiquer sa méthodologie : « Il apparaît que les classifications que nous avons réalisées relèvent de la propriété scientifique (un peu du même ordre qu'un brevet) et que nous devons établir une licence d'exploitation pour vous autoriser à les utiliser. Aux fins d'établir cette licence qui définira les conditions d'exploitation, vous devez faire votre demande à la directrice, Magda Tomasini [...] qui fera suivre au service juridique » (message du 27 septembre 2019). Autrement dit, alors qu'après trois mois d'attente (du 5 juin au 25 septembre) Patrick Simon proposait de repousser encore la date à laquelle il honorerait son engagement, il semblerait qu'il lui a suffi de deux jours (du 25 au 27 septembre) pour réaliser l'analyse juridique selon laquelle il lui était impossible de me communiquer sa nomenclature. Ensuite lorsque, suite à ce message de Patrick Simon du 27 septembre 2019, j'ai contacté la directrice de l'INED, puis le référent à l'intégrité scientifique de l'INED, puis de nouveau les auteurs de l'article et échangé avec eux une vingtaine de messages supplémentaires étalés sur 15 mois (Annexe 4A, p. 94), aucun n'a plus jamais mentionné ladite licence d'exploitation : elle a comme disparu.

Dans l'entretien susmentionné à la revue *Tracés*, un collègue de Patrick Simon rebondit sur ses déclarations, en rappelant au passage les principes qui justifient que l'intégrité scientifique exige de communiquer la méthodologie des publications :

C. BEAUCHEMIN : Par ailleurs, il y a un certain nombre de revues dans notre domaine qui exigent maintenant que tu mettes à disposition, quand ils te publient, tes données et tes programmes, à des fins de répliquabilité. C'est le cas à *Demographic Research* par exemple. Et ça demande un travail supplémentaire, parce qu'effectivement quand tu travailles tu peux bien faire ton travail mais de façon néanmoins un peu brouillonne. Dès lors que tu dois rendre ton fichier de données et ton programme disponibles, il faut faire le ménage ! C'est comme si tu devais ranger ton bureau à chaque fois. Mais sur le fond je trouve ça bien de pouvoir garantir la répliquabilité, de même que, sur le principe, je trouve ça bien qu'on mette nos données à disposition parce qu'elles sont produites avec de l'argent public. Cependant l'application de ces principes-là est plus difficile. En fait, on est d'accord sur le principe mais ça a un coût énorme qui est souvent non estimé ou sous-évalué et dont la récompense peut être parfois douteuse : tu n'es pas forcément cité, tu peux être démolé, etc. Il y a une mise en danger.



Si une méthodologie est scientifiquement défendable, comment la communiquer à des collègues pourrait-il mettre un chercheur en danger ? Et les chercheurs et leurs institutions ne sont-ils pas, simplement, *tenus* de respecter les règles de l'intégrité scientifique ?

### ❖ Le refus de l'INED de communiquer sa méthodologie

Face au refus des auteurs de communiquer leur méthodologie, j'ai dû m'adresser au rédacteur en chef du bulletin d'information scientifique qui a publié l'article, Gilles Pison (26 septembre 2019, Annexe 4A, p. 94). Informé du refus des auteurs de me communiquer leur méthodologie et du fait que Patrick Simon avait violé son engagement à me communiquer sa nomenclature (message du 26 septembre 2019), le rédacteur en chef de la revue s'est contenté de me *souhaiter* de trouver avec les auteurs une formule qui permette de prolonger leur étude. Pourtant, de l'avis du référent à l'intégrité scientifique de l'INED (message du 14 novembre 2019), ce sont principalement les revues qui doivent se porter garantes des bonnes pratiques des auteurs de publications. De fait, un rédacteur en chef d'une revue scientifique de bonne tenue peut aller jusqu'à rétracter un article, s'il apparaît que ses auteurs refusent de communiquer une méthodologie qui est communicable.

En l'absence de réaction j'ai donc dû écrire à la directrice de la publication et directrice de l'INED, Magda Tomasini (message du 30 septembre 2019), qui m'a indiqué que la table de passage serait disponible par le biais du réseau Quételet un an après la publication, en avril 2020. Qu'est-ce qui justifie un délai d'une année pour accepter de communiquer à un collègue une simple nomenclature des prénoms, nécessaire à la reproduction d'une publication et que les revues scientifiques de bonne tenue communiquent en ligne, en annexe de la publication ? Et pourquoi les auteurs et l'INED refusent-ils, sans justification aucune, de communiquer leur programme informatique ? Pour le savoir j'ai répondu au message de la directrice de l'INED (7 octobre 2019), qui n'y a apporté aucune réponse.

J'ai donc dû envoyer un message au référent à l'intégrité scientifique de l'INED, Henri Leridon (message du 22 octobre 2019), qui près d'un mois plus tard m'a répété le même message : concernant la nomenclature, elle sera disponible aux chercheurs en avril 2020 (ce qui n'explique pas le refus de me la communiquer plus tôt) ; et concernant le programme informatique des auteurs, toujours aucune réponse ni aucune justification de cette absence de réponse. Concrètement, l'INED n'exige pas des auteurs de ses publications scientifiques qu'ils communiquent la méthodologie qui leur a permis de produire leurs résultats (message de Magda Tomasini et Henri Leridon du 14 novembre 2019, répété dans le message de Patrick Simon du 7 mars 2020).

Il semblerait pourtant que l'INED, [comme les autres institutions de la recherche publique française](#), est bel et bien signataire de la « [Charte nationale de déontologie des métiers de la recherche](#) » du 26 janvier 2015, dont la partie relative à la « Fiabilité du travail de recherche » indique :

« Les méthodes mises en œuvre pour la réalisation du projet de recherche doivent être les plus appropriées. La description détaillée du protocole de recherche, dans le cadre de cahiers de laboratoire ou de tout autre support, doit permettre la reproductibilité des travaux expérimentaux. Tous les résultats bruts (qui appartiennent à l'institution) ainsi que l'analyse des résultats doivent être conservés de façon à permettre leur vérification. »

Si la méthodologie des articles de recherche scientifique doit être reproductible, n'est-ce pas pour qu'elle puisse être communiquée à des collègues lorsque ces derniers le demandent, afin qu'ils puissent — précisément — reproduire lesdits résultats ?

Interloqué par la situation, j'ai contacté la Mission à l'intégrité scientifique (MIS) de mon employeur, le [Centre national de la recherche scientifique \(CNRS\)](#) (le 20 novembre 2019), pour lui signaler que depuis plus de sept mois des collègues et l'INED refusaient de communiquer



la méthodologie nécessaire à la reproduction d'une publication scientifique. La MIS m'a mis en contact avec des chargés de mission (le 3 décembre 2019) qui ont contacté le référent à l'intégrité scientifique de l'INED pour faire aboutir mes demandes. Ils m'ont indiqué que suite aux conseils du référent à l'intégrité scientifique de l'INED je pouvais recontacter les auteurs (message du 17 février 2020). Quand j'ai recontacté les auteurs ainsi que la directrice de l'INED et le référent à l'intégrité scientifique de l'INED pour leur redemander de bien vouloir me communiquer leur méthodologie (message du 17 février 2020), ils ont de nouveau commencé par ne pas répondre (d'où mon message de relance du 3 mars 2020). Puis Baptiste Coulmont a répondu qu'il suivrait la décision de l'INED, comme s'il revenait à l'INED de choisir s'il souhaite respecter les règles de l'intégrité scientifique (message du 4 mars 2020). Et enfin Patrick Simon a répété le même message : la nomenclature sera disponible en avril 2020, et toujours aucune réponse concernant le programme informatique (message du 7 mars 2020).

Dans un message récapitulatif envoyé plus d'une année après la parution de l'article et après mes premiers doutes, je demandais aux auteurs, à la directrice de l'INED et au référent à l'intégrité scientifique de l'INED pourquoi « depuis plus d'une année maintenant, vous refusez de communiquer les éléments méthodologiques qui permettraient de reproduire les résultats que vous avez publiés dans votre article de 2019, alors même qu'**aucun de ces éléments méthodologiques ne comprend des données individuelles et n'est donc soumis au secret statistique** » (message du 20 avril 2020, en gras et souligné dans le message lui-même). Je terminais le même message ainsi : « Qu'est-ce qui vous empêche de me communiquer cette double table de passage alors qu'elle est légalement communicable [...] ? D'autre part, pourquoi refusez-vous de me communiquer le programme informatique qui, une fois appliqué aux bases TeO (auxquelles j'ai accès par le biais du CASD), permettrait de reproduire les résultats de votre publication ? Depuis un an que vous refusez de me communiquer ces deux éléments méthodologiques que la plupart des chercheurs auraient communiqué en moins d'une semaine, vous n'avez jamais cru bon de justifier votre refus » (message du 20 avril 2020). Je n'ai jamais reçu de réponse de la part des auteurs, ni du référent à l'intégrité scientifique de l'INED, ni de la directrice de l'INED. Ils n'ont jamais exprimé d'intérêt pour la discussion de fond.

Le 18 décembre 2019, soit dix semaines après que j'aie écrit à la directrice de l'INED pour lui demander la méthodologie de sa publication — et après dix semaines sans réponse positive de sa part sur la double table de passage et sans réponse *tout court* sur le programme informatique —, la directrice de l'INED déclarait, dans un [cycle de conférences organisé par la section du rapport et des études du Conseil d'État](#) :

« La recherche c'est la transparence des méthodes, le respect des questions d'éthique et déontologiques, et une qualité scientifique par une évaluation des pairs des travaux conduits » ([vidéo](#) à 19:25).

C'est cela.

### ❖ Dernières combines

L'INED, qui s'était finalement engagé à mettre à disposition sa nomenclature des prénoms en avril 2020, n'a en réalité communiqué une nomenclature à l'INSEE qu'en mai 2020, ce qui compte tenu des délais de traitement de ces informations (par l'INSEE ainsi que par le réseau Quételet et le CASD), ne m'a permis d'y accéder que le 30 novembre 2020. J'ai alors pu découvrir la nomenclature des prénoms (appelée « CAT0 ») que l'INED a communiquée à la communauté scientifique, et que je présente ci-dessous en regard de la nomenclature utilisée dans l'article de Coulmont et Simon (2019) :

**Tableau 4.3. Modalités de la nomenclature utilisée par Coulmont et Simon (2019), et de la nomenclature qu'ils ont communiquée au réseau Quételet et au CASD (en 2020)**

Nomenclature des prénoms utilisée par Coulmont et Simon (2019, p. 2)	Nomenclature des prénoms communiquée au réseau Quételet et au CASD (en 2020)
« Français »	« France »
« Latin »	
« Internationaux ou anglophones »	« Prénoms internationaux (principalement anglo-américains) »
« Arabo-musulman »	« États de langue arabe (hors Maghreb) »
« Africain »	
« Asiatique »	
	« Langue associée à un pays »
	« Prénoms largement diffusés »
« Inclassable »	« Impossible à classer »
« Autre groupe »	« Erreur »

Dans la nomenclature des prénoms communiquée au réseau Quételet et au CASD, quelles peuvent bien être les origines des « langues associées à un pays » et des « prénoms largement diffusés » ? Où peuvent bien être catégorisés les prénoms originaires du Maghreb ? Comment peut-il se faire que cette nomenclature regroupe, parmi les prénoms originaires des « États de langue arabe (hors Maghreb) » (CAT0=641), des prénoms grecs (CAT1=134), turcs (CAT1=163) et espagnols et portugais (CAT1=661) ? L'INED a non seulement communiqué à la communauté scientifique une nomenclature dont certaines modalités n'ont pas de sens, dont les modalités ne couvrent pas la totalité du champ des possibles, et dont les regroupements sont farfelus : comme chacun peut le constater, l'INED a surtout communiqué une nomenclature différente de celle utilisée par Coulmont et Simon dans leur article, puisque ces deux nomenclatures ne comportent même pas les mêmes modalités (tableau 4.3). Où sont passés les prénoms que leur article catégorise comme « latins », « africains » ou encore « asiatiques » ?

Coulmont, Simon et l'INED m'ont donc fait patienter pendant plus de 18 mois (Annexe 4A, p. 94) pour finalement communiquer à la communauté scientifique une nomenclature fallacieuse, qui ne permet pas de reproduire leurs analyses ni même de mener une quelconque analyse. Il va sans dire que l'INED devra (i) retirer la variable CAT0 des bases TeO mises à disposition des chercheurs, puisque cette variable ne peut avoir aucune utilité pour quiconque ; (ii) remplacer cette variable CAT0 par la variable CAT1 regroupée en 8 modalités selon les règles exposées dans le message de Patrick Simon du 4 décembre 2020 ; (iii) supprimer de la documentation méthodologique mise à disposition des chercheurs les références à une « classification probabiliste des référents linguistiques des syllabes » et à un « algorithme » ; et (iv) remplacer ces inexactitudes par une citation de la véritable source méthodologique de la nomenclature des prénoms CAT1, à savoir la méthodologie de Humpert et Schneiderheinze (2000 et 2013 ; Liebau et al. 2018).

Dans ce contexte, j'ai réécrit à la directrice de l'INED et aux auteurs, tout en prenant à témoin plusieurs responsables de l'intégrité scientifique (à l'INED, au CNRS mais aussi à l'ENS Paris-Saclay dont dépend désormais Baptiste Coulmont), pour leur signaler que leur manipulation ne m'avait pas échappé (messages du 1<sup>er</sup> et du 2 décembre 2020). J'ai ainsi été amené à demander à la directrice de l'INED si « l'INED ne compte finalement pas respecter les engagements pris » (message du 1<sup>er</sup> décembre 2020), suite à quoi Coulmont et Simon ont consenti à me communiquer une nomenclature qui pourrait s'approcher de celle qu'ils ont utilisée dans leur article (message du 4 décembre 2020). Comme la nomenclature communiquée comprend toutefois une modalité « divers » qui semble absente de la nomenclature de l'article,

j'ai dû de nouveau leur écrire pour leur demander comment ils étaient passés de la nomenclature qu'ils m'ont communiquée à celle de leur article, et j'ai profité de l'occasion pour leur demander, aussi, ce qu'il en était de leur programme informatique : « Que vais-je devoir faire, cette fois-ci, pour que votre magnanimité vous pousse jusqu'à respecter la déontologie de notre métier ? » (message du 5 décembre 2020). Comme ils n'ont pas répondu, j'ai écrit à la directrice de l'INED pour lui demander « Madame la directrice, est-ce que l'INED assume de refuser de communiquer la méthodologie de cette publication ? Si non, quand comptez-vous mettre un terme à cette mascarade ? » (message du 18 décembre 2020). Ni les auteurs ni la direction de l'INED n'ont répondu — personne n'a contesté la qualification de la situation comme une « mascarade » —, mais ils ont finalement communiqué au CASD trois programmes informatiques R (6 janvier 2021), présentés en Annexe 4B, p. 124.<sup>19</sup>

Alors, ces programmes permettent-ils de reproduire les résultats publiés par Coulmont, Simon et l'INED ? Non, parce qu'ils sont à la fois inutilisables et incomplets. D'une part, je ne peux pas exécuter la totalité de ces programmes car ils exploitent des bases de données qui ne sont pas mises à disposition des chercheurs (par le réseau Quételet ou le CASD).<sup>20</sup> D'autre part, même si j'avais pu exécuter la totalité de ces programmes, ils ne permettraient pas de produire les tableaux et graphiques publiés ou leurs résultats sous-jacents, puisqu'ils ne contiennent pas toutes les variables nécessaires à la production des résultats publiés. Par exemple, aucun de ces programmes ne contient la variable de religiosité des enquêtés (R\_IMPVIIE), pourtant nécessaire à la construction de la figure 3 de l'article de Coulmont et Simon (2019).<sup>21</sup> Ces programmes ne permettent pas, non plus, de connaître les champs sur lesquels les tableaux et graphiques ont pu être calculés. Autrement dit, les programmes que m'ont communiqué les auteurs et l'INED sont incomplets, si bien qu'ils ne permettent pas de reproduire leurs analyses ni de vérifier si elles produisent effectivement les résultats qu'ils ont publiés. Préférant ne pas perdre plus de temps et d'énergie à démonter cette manipulation, je reprends toute l'analyse des prénoms en France et j'entreprends de reproduire leur article, sinon à l'identique, du moins « au plus près ».

### ❖ Résumé

Au bout de 50 messages électroniques échangés sur une période de 20 mois (Annexe 4A, p. 94), les auteurs de l'article et l'INED ont empêché la reproduction à l'identique de leurs analyses. S'ils m'ont finalement communiqué leur nomenclature des prénoms — non sans lui avoir préalablement substitué une nomenclature fallacieuse —, ils ne m'ont jamais communiqué leur programme informatique complet, ni jamais justifié ce refus. Entre tentatives d'esquive, manœuvres dilatoires, multiplication de fausses excuses, faux-semblants, mensonges purs et simples, non-respect des engagements pris et autres combines, les auteurs et l'INED n'ont pas cessé de violer les règles de l'intégrité scientifique et de la « [Charte nationale de déontologie des métiers de la recherche](#) ». Hormis lors de ma première conversation avec Baptiste Coulmont (17 avril 2019), ni les auteurs ni l'INED n'ont jamais accepté de voir leur travail examiné scientifiquement. Peut-être n'étaient-ils pas totalement confiants dans la fiabilité de leurs résultats ?

---

<sup>19</sup> Il s'agit d'un premier programme appelé « 01-importation.R » (de 837 lignes), d'un deuxième appelé « 02-recodage.R » (de 545 lignes), et d'un troisième appelé « 03-tables.R » (de 295 lignes).

<sup>20</sup> Les programmes de Coulmont, Simon et l'INED (Annexe 4B, p. 124) utilisent deux tables « individus » (« indiv\_moa.sas7bdat » et « base\_indiv\_28052012.sas7bdat »), une table « THL » (« thl\_moa.sas7bdat ») et une table « enfants » (« enfants\_moa.sas7bdat »), dont aucune n'est mise à disposition. J'ai bien essayé de faire tourner ces programmes avec les tables SAS dont je dispose renommées comme les leurs, mais leur 1<sup>er</sup> programme (sur 3) affiche une erreur incorrigible dès la ligne 444.

<sup>21</sup> C'est ce que chacun peut vérifier en recherchant « R\_IMPVIIE » dans l'Annexe 4B, p. 124. D'autres variables nécessairement utilisées par les auteurs, comme les variables d'affiliation religieuse (RELEGO1, RELCA1 et RELPC1), sont, elles aussi, omises des programmes informatiques qui m'ont été communiqués.

## 5. Méthodologie de la reproduction « au plus près » : la base de données

Les prénoms les plus portés par les petits-fils d'immigrés du Maghreb sont-ils vraiment « Yanis » et « Nicolas » ? Et plus généralement est-il vrai que seulement 23 % des petits-enfants d'immigrés du Maghreb portent des prénoms « arabo-musulmans » ? J'ai entrepris de reproduire les analyses de Coulmont et Simon (2019), c'est-à-dire le tableau et les trois figures de l'article, sur la même base de données qu'eux : l'enquête TeO enrichie de la variable « prénom » (sans nom de famille).

### ❖ Synthèse des analyses méthodologiques présentées dans l'Annexe 5A, p. 147

La reproduction « au plus près » utilise 24 variables issues de 4 des 22 tables de la version nominative de l'enquête TeO fournies par le CASD. Je rassemble ces informations concernant les enquêtés, les autres habitants de leur ménage et leurs enfants non cohabitants (y compris décédés) en une seule base de données, qui contient un échantillon maximal de 83 045 individus. Cette table permet de reproduire à l'identique une trentaine de tris à plat du dictionnaire des codes de l'enquête (Insee 2011), ce qui indique que sa construction ne semble pas y avoir introduit d'erreur.

Dans ce contexte, le présent travail se fonde sur la méthodologie suivante :

- **l'échantillon** retenu pour la plupart des analyses est représentatif des personnes de 18-50 / 18-60 ans en ménage ordinaire et de leurs enfants de 0-17 ans, habitant en France métropolitaine en 2008 (N = 42 208) ; il est composé de tous les enquêtés et de certains enfants de l'enquêté vivant en métropole, mais d'aucun autre membre du ménage de l'enquêté, afin d'éviter les doubles comptes et pour s'approcher d'un échantillon représentatif des habitants de 0-50 ans / 0-60 ans en France métropolitaine en 2008 ;
- **la pondération** utilisée consiste à pondérer les enquêtés avec la variable POIDSI, et leurs enfants avec la variable POIDSE ; POIDSE vaut la moitié de POIDSI, car un enfant d'enquêté a deux fois plus de chances qu'un enquêté d'être échantillonné : par son père, et par sa mère.

Cet échantillon pondéré permet de reproduire de nombreux mouvements de la mode des prénoms en France depuis les années 1950, ce qui indique que l'échantillon retenu et sa pondération, mais aussi la variable « prénom », sont de bonne qualité. Confiant dans la fiabilité des données TeO et de cette méthodologie, je les utilise dans les chapitres suivants.

Le programme informatique de la reproduction « au plus près » figure dans l'Annexe 5B, p. 183.

## 6. Méthodologie de la reproduction « au plus près » : la nomenclature des prénoms par origine

### ❖ Synthèse de la revue de littérature présentée dans l'Annexe 6A, p. 267

Dans leur article, Coulmont et Simon (2019, encadré p. 2) catégorisent des prénoms comme « français », « latins » ou « arabo-musulmans », y compris certains de ceux portés en France par les personnes originaires d'Europe du Sud (Italie, Espagne, Portugal) et du Maghreb (Maroc, Algérie, Tunisie). Mais qu'est-ce qu'un prénom « français », « latin » ou « arabo-musulman » ? L'histoire comparée des prénoms en Europe et au Maghreb indique qu'on peut les définir de la façon suivante — le caractère relativement large de ces catégorisations n'empêchant pas d'affiner l'analyse au cas par cas :

- un prénom « **français** » est un prénom usuellement attribué par les locuteurs de langue française, à une période donnée<sup>22</sup> ; depuis les environs du 12<sup>e</sup> siècle, ces prénoms sont le plus souvent issus du prénom d'un personnage du Nouveau Testament ou du prénom d'un saint ou martyr chrétien, leur origine étymologique pouvant elle-même être très variée, notamment hébraïque, grecque, latine, germanique ou celtique ;
- un prénom « **latin** » (ou « **anglophone** ») est un prénom usuellement attribué par les locuteurs des langues italienne, espagnole et portugaise (ou de la langue anglaise) ; ces prénoms sont eux aussi le plus souvent issus du Nouveau Testament ou d'un saint chrétien, et d'origines étymologiques variées ;
- un prénom « **arabo-musulman** » est un prénom usuellement attribué par les locuteurs de langue arabe ou par des populations non arabophones mais de religion musulmane (berbérophones, turcophones, kurdophones, persanophones, de langues austronésiennes, etc.) ; certains de ces prénoms sont usuellement attribués non seulement par des Arabes musulmans (comme les prénoms de personnages du Coran, y compris les prénoms musulmans théophores ou les prénoms des proches et successeurs du prophète Mohamed), mais aussi par des Arabes non musulmans (comme les prénoms arabes profanes ou chrétiens) et par des musulmans non arabes (prénoms berbères, turcs, persans, etc., dont certains sont des adaptations de prénoms arabes mais d'autres sont sans rapport avec l'arabe voire préislamiques) ; l'origine étymologique dominante des prénoms « arabo-musulmans » est arabe, même si elle peut aussi être hébraïque, berbère, turque, persane, etc.

### ❖ Synthèse des analyses méthodologiques présentées dans l'Annexe 6B, p. 285

La reproduction « au plus près » utilise la même nomenclature des prénoms par origine que Coulmont et Simon (2019). Une fois expurgée de leurs erreurs de recodage, cette nomenclature de l'INED concorde globalement bien avec une autre nomenclature conçue indépendamment, et ces nomenclatures sont globalement conformes aux connaissances établies en onomastique. Confiant dans la fiabilité de cette nomenclature de l'INED, je l'utilise dans les chapitres suivants.

---

<sup>22</sup> Une définition rigoureuse exigerait de préciser à partir de quelle fréquence et de quelle durée d'attribution — ou à partir de quel degré d'exclusivité — un prénom devient « français », « latin », « anglophone », etc. mais une telle définition étant impraticable, j'admets les catégorisations réalisées par la littérature onomastique ainsi que par divers auteurs qui ont constitué et utilisé des nomenclatures de prénoms selon la population qui, de nos jours, les attribue usuellement.

## 7. Essai de reproduction du premier résultat spectaculaire : « Nicolas »

On a vu que l'échantillon TeO retenu pondéré et sa variable « prénom » sont fiables (Annexe 5A, p. 147), tout comme la nomenclature des prénoms de l'INED (Annexe 6B, p. 285). Je peux donc les utiliser pour la reproduction « au plus près ». Dans leur article, Coulmont et Simon (2019) indiquent qu'en France métropolitaine en 2008, le prénom le plus fréquent chez les petits-fils d'immigrés du Maghreb serait, derrière « Yanis », « Nicolas » (Coulmont et Simon 2019, tableau p. 3). Est-ce vrai ?

### ❖ Synthèse des analyses présentées dans l'Annexe 7, p. 327

L'échantillon maximal TeO des petits-fils d'immigrés du Maghreb ne contient que 23 « Yanis » et 11 « Nicolas » — des effectifs trop faibles pour permettre d'établir un palmarès des prénoms comme celui qui a été publié. Mais sur l'échantillon retenu pondéré (représentatif des personnes de 18-50 / 18-60 ans en ménage ordinaire et de leurs enfants de 0-17 ans habitant en France métropolitaine en 2008), « Nicolas » est effectivement l'un des prénoms les plus portés par les petits-fils d'immigrés du Maghreb, tout comme « Jérémy » et « Benoît ». Ce sont même plus de 3 % des petits-fils d'immigrés du Maghreb qui seraient prénommés « Nicolas ». Et si l'on reproduit le tableau de Coulmont et Simon (2019, p. 3) sur d'autres échantillons — peut-être proches de l'échantillon utilisé par les auteurs, mais qui contiennent des doubles comptes —, il arrive que « Nicolas » soit le prénom *le plus fréquent* chez les petits-fils d'immigrés du Maghreb !

Problème : personne ne semble avoir entendu parler d'autant de petits-fils d'immigrés du Maghreb prénommés « Nicolas ». Ni les experts des migrations qui ont bien voulu répondre à mes questions, ni les recherches par entretiens menées depuis les années 1980 auprès des couples « mixtes », ni la presse écrite, ni la radio et la télévision françaises, ni les ouvrages publiés en français depuis 1970, ni les moteurs de recherche Internet ou l'encyclopédie Wikipedia, ni l'annuaire téléphonique et les réseaux sociaux en 2021, ni même les décès en France ne comptent suffisamment de « Nicolas » originaires du Maghreb. Alors, que se passe-t-il ? Un examen de la méthodologie de l'enquête permet d'envisager que les bases de données de l'enquête TeO contiennent une erreur. Une partie de ces « Nicolas » ne seraient-ils pas des petits-enfants de *rapatriés* français ou européens de l'ex-empire colonial français en Afrique du Nord (y compris des Français et des Européens d'Algérie), autrement dit des petits-enfants de *pieds-noirs* ? Les pieds-noirs ne sont pas considérés comme des immigrés, parce que suite à l'indépendance des pays du Maghreb, ils ont conservé la nationalité française.

### ❖ Les petits-fils d'immigrés du Maghreb prénommés « Nicolas » seraient-ils, en réalité, des petits-fils de pieds-noirs ?

Alors, les petits-fils d'immigrés du Maghreb prénommés « Nicolas » du tableau de Coulmont et Simon (2019, p. 3) sont-ils bien des descendants d'immigrés du Maghreb, ou sont-ils plutôt des descendants de rapatriés d'Afrique du Nord ? Pour le savoir, je commence par examiner les prénoms de leur entourage,<sup>23</sup> à partir de l'idée suivante : si l'entourage familial des petits-fils d'immigrés du Maghreb prénommés « Nicolas » porte plutôt des prénoms « arabo-musulmans », c'est que ces « Nicolas » sont effectivement des petits-fils d'immigrés du Maghreb, tandis que si leur entourage porte plutôt des prénoms « français », « latins » ou « internationaux ou anglophones » c'est que ces « Nicolas » sont plus probablement des petits-fils de rapatriés d'Afrique du Nord (parmi lesquels on comptait beaucoup de Français d'origine espagnole, italienne, allemande, maltaise, etc.).

---

<sup>23</sup> Sur l'échantillon TeO maximal, les prénoms de l'entourage d'un petit-fils d'immigrés du Maghreb prénommé « Nicolas » sont renseignés pour (i) tous les habitants du logement de son parent enquêté et (ii) tous les autres enfants de son parent enquêté, vivants ou non et qu'ils habitent dans le logement de leur parent enquêté ou non.



L'échantillon maximal non pondéré de l'enquête TeO compte 11 petits-fils d'immigrés du Maghreb prénommés « Nicolas ». On peut s'étonner que Coulmont, Simon et l'INED aient publié un tableau de résultats dans lequel les prénoms des petits-enfants d'immigrés du Maghreb présentés comme les plus fréquents en France métropolitaine en 2008 reposent sur des effectifs aussi réduits, mais ces résultats ayant été publiés, il convient de les examiner. Sur l'échantillon maximal non pondéré de l'enquête TeO, les 11 petits-fils d'immigrés du Maghreb prénommés « Nicolas » comptent dans leur entourage 30 personnes originaires du Maghreb, parmi lesquelles 26 portent un prénom que la nomenclature de l'INED catégorise comme « français », « latin » ou « international ou anglophone », et seulement 3 portent un prénom « arabo-musulman » (tableau 7.10). Par contraste, les 23 petits-fils d'immigrés du Maghreb prénommés « Yanis » — prénom masculin le plus fréquent chez les petits-fils d'immigrés du Maghreb, juste devant « Nicolas », d'après le tableau de Coulmont et Simon (2019, p. 3) — comptent dans leur entourage 67 personnes originaires du Maghreb, parmi lesquelles 45 portent un prénom « arabo-musulman » et 19 portent un prénom « français », « latin » ou « international ou anglophone » (tableau 7.10).

Autrement dit, sur l'échantillon maximal non pondéré, les petits-fils d'immigrés du Maghreb qui sont prénommés « Nicolas » ont un entourage qui porte des prénoms massivement français ou européens (87 %), comme s'ils étaient petits-fils de pieds-noirs, ce qui les distingue des petits-fils d'immigrés du Maghreb prénommés « Yanis », dont l'entourage porte des prénoms majoritairement « arabo-musulmans » (67 %), comme s'ils étaient effectivement petits-fils d'immigrés du Maghreb. Il s'agit là d'un indice flagrant que **dans le tableau publié par l'INED (Coulmont et Simon 2019, p. 3), les petits-fils d'immigrés du Maghreb prénommés « Nicolas » ne sont probablement pas des petits-fils d'immigrés du Maghreb, mais sans doute plutôt des petits-fils de rapatriés d'Afrique du Nord.** Donc contrairement à ce que laisse entendre ce tableau, les petits-fils d'immigrés du Maghreb prénommés « Nicolas » ne sont pas typiquement les enfants de « Mohamed » et « Sarah » : ils sont plutôt les enfants de « Christine » et « Philippe ». Difficile, dès lors, de soutenir la thèse selon laquelle cette forte prévalence du prénom « Nicolas » chez les petits-fils d'immigrés du Maghreb serait un signe de désir d'assimilation : les parents de ces « Nicolas » n'ont sans doute jamais envisagé de choisir, pour leur enfant, un prénom « arabo-musulman ».

**Tableau 7.10. Prénoms de l'entourage maghrébin des petits-fils d'immigrés du Maghreb prénommés « Nicolas » et « Yanis », en France en 2008 (enquête TeO)**

		N	ORIGINE DU PRÉNOM DES INDIVIDUS (NOMENCLATURE DE L'INED)							
			Français	Latin	International ou anglophone	Autre Europe	Arabo-musulman	Africain	Asiatique	Inclassable
« NICOLAS » (N=11)	Grands-parents (G1)	0	0	0	0	0	0	0	0	0
	Parents (G2)	11	5	0	2	0	3	0	0	1
	Frères et sœurs (G3)	19	10	1	8	0	0	0	0	0
	Total	30	15	1	10	0	3	0	0	1
« YANIS » (N=23)	Grands-parents (G1)	3	0	0	0	0	2	0	0	1
	Parents (G2)	35	2	1	2	0	29	1	0	0
	Frères et sœurs (G3)	29	4	1	9	0	14	0	0	1
	Total	67	6	2	11	0	45	1	0	2

Lecture : les 11 petits-fils d'immigrés du Maghreb prénommés « Nicolas » comptent dans leur entourage 30 personnes originaires du Maghreb, parmi lesquelles 15 portent un prénom « français » et seulement 3 portent un prénom « arabo-musulman » (d'après la nomenclature de l'INED).

Champ : individus du champ maximal qui ont dans leur entourage un petit-fils d'immigrés du Maghreb prénommé « Nicolas » ou « Yanis », données non pondérées, N = 97.

Source : TeO (voir Annexe 5B, p. 183).

Le fait que les petits-fils d'immigrés du Maghreb prénommés « Nicolas » ont un entourage qui porte largement des prénoms « français » et « européens », tandis que l'entourage des « Yanis » porte largement des prénoms « arabo-musulmans », est suggestif, mais peut-être pas décisif. Après tout, si un descendant d'immigrés du Maghreb prénommait effectivement l'un de ses fils « Nicolas », serait-il étonnant qu'il attribue à ses autres enfants des prénoms « français » ou « européens » ? Certes, les prénoms des *parents* des « Nicolas » sont, pour la plupart, eux aussi « français » ou « européens », alors que les prénoms des *parents* des « Yanis » sont massivement « arabo-musulmans » (tableau 7.10). Mais un examen des *seuls prénoms* de l'entourage de ces petits-fils d'immigrés du Maghreb prénommés « Nicolas » permet-il d'être confiant que la plupart d'entre eux sont en réalité des petits-fils de pieds-noirs ?

Pour évaluer quels petits-fils d'immigrés du Maghreb sont bien, en probabilité, des descendants d'immigrés du Maghreb plutôt que des descendants de rapatriés, mais *sans utiliser pour cela leurs prénoms*, je propose de combiner 4 critères :

- i. premièrement, comme la plupart des immigrés du Maghreb sont arabophones ou berbérophones (en plus d'être, pour certains, francophones), mais que la plupart des rapatriés ne le sont pas,<sup>24</sup> j'examine les langues que l'immigré (G1) du Maghreb parlait à ses enfants (G2) dans leur enfance (variables 17 à 20 du tableau 5.1) : s'il leur parlait arabe ou berbère, même en seconde langue, alors il n'est probablement pas un rapatrié ;
- ii. deuxièmement, comme les immigrés (G1) du Maghreb — appelés pendant la période coloniale « indigènes musulmans », « Français musulmans » d'Algérie ou « musulmans originaires d'Algérie » — sont majoritairement musulmans (quand ils ont une religion), mais que les rapatriés ne le sont pas,<sup>25</sup> j'examine leur affiliation religieuse (variable 21 du tableau 5.1) : s'ils sont musulmans, alors ils ne sont probablement pas des rapatriés ;
- iii. troisièmement, comme la religion du grand-parent de « Nicolas » qui est immigré (G1) n'est pas toujours renseignée,<sup>26</sup> j'examine aussi celle de l'enfant d'immigrés (G2) du Maghreb, le parent de « Nicolas » (variable 21 du tableau 5.1) : s'il est musulman, alors il n'est probablement pas un descendant de rapatriés ;
- iv. quatrièmement j'examine l'opinion du parent de « Nicolas » qui a été enquêté par rapport à l'affirmation « On me voit comme un Français » (variable 24 du tableau 5.1) : s'il n'est « plutôt pas d'accord » ou « pas du tout d'accord » avec l'idée qu'on le voit comme un Français, alors il n'est probablement pas un descendant de rapatriés.<sup>27</sup>

<sup>24</sup> En effet, sur l'échantillon pondéré des enquêtés, seulement 5 % des enquêtés enfants de rapatriés d'Afrique du Nord ont au moins un parent qui leur parlait dans leur enfance en arabe ou en berbère (comme 1<sup>e</sup> ou 2<sup>e</sup> langue), contre 65 % des enquêtés enfants d'immigrés du Maghreb.

<sup>25</sup> En effet, sur l'échantillon pondéré des enquêtés, seulement 6 % des enquêtés enfants de rapatriés d'Afrique du Nord sont d'origine musulmane, contre 77 % des enquêtés enfants d'immigrés du Maghreb. Cette condition a pour conséquence que, parmi les rapatriés d'Algérie, les harkis (rapatriés musulmans) resteraient comptabilisés parmi les immigrés du Maghreb. Comme on sait par ailleurs que certains d'entre eux, porteurs de prénoms arabes, ont pu attribuer à certains de leurs enfants des prénoms français (Kydjian 2016, p. 116-118), cela contribuerait à surestimer la part de prénoms français parmi les descendants d'immigrés du Maghreb, relativement à la part de prénoms français parmi les descendants de rapatriés d'Afrique du Nord.

<sup>26</sup> L'enquête TeO renseigne l'affiliation religieuse de l'enquêté et de ses parents, ainsi que celle de son conjoint actuel (s'il en a un) et celle de son premier conjoint (s'il est différent du conjoint actuel), mais pas celle des parents de son conjoint actuel ni de son premier conjoint. En conséquence, l'affiliation religieuse du grand-parent de « Nicolas » qui est immigré (G1) du Maghreb n'est pas renseignée lorsque le(s) seul(s) grand(s)-parent(s) immigré(s) (G1) se trouve(nt) du côté du parent (G2) de « Nicolas » qui n'a *pas* été enquêté.

<sup>27</sup> En effet, sur l'échantillon pondéré des enquêtés, seulement 5 % des enquêtés enfants de rapatriés d'Afrique du Nord se déclarent « plutôt pas d'accord » ou « pas d'accord du tout » avec le fait qu'on les voit comme des

Pris ensemble, ces 4 critères peuvent être considérés comme stricts. Quelle serait la probabilité qu'un habitant prénommé « Nicolas » en métropole en 2008 soit petit-fils *d'immigrés* du Maghreb, plutôt que petit-fils *de rapatriés*, si son grand-parent né en Afrique du Nord ne parlait ni arabe ni berbère à ses enfants, si ni son grand-parent (G1) ni son parent (G2) n'est musulman, et si son parent enquêté considère qu'« on le voit comme un Français » ? Le porteur de toutes ces caractéristiques à la fois serait plus probablement un descendant de rapatriés.

En examinant les profils des 11 petits-fils d'immigrés du Maghreb prénommés « Nicolas », il apparaît que pour la plupart d'entre eux : leur grand-parent immigré (G1) du Maghreb n'a parlé à son enfant (G2) qu'en français ; ni leur grand-parent immigré (G1) ni leur parent enquêté (G2) n'est musulman ; et leur parent enquêté (G2) se dit « tout à fait d'accord » avec l'idée qu'on le voit comme un « Français » (tableau 7.11). J'en conclus que, sur l'échantillon maximal non pondéré de l'enquête TeO, 9 des 11 « Nicolas » ne présentent aucun des 4 indicateurs de descendant d'immigrés : ils sont plus probablement des descendants de rapatriés. Seuls 2 des 11 « Nicolas » pourraient être correctement catégorisés comme descendants d'immigrés (tableau 7.11).

Mais peut-être cette méthode, aussi stricte puisse-t-elle paraître, surestime-t-elle le nombre de descendants de rapatriés que TeO a catégorisés à tort comme descendants d'immigrés ? Pour le savoir, j'applique cette méthode de détection des descendants de rapatriés d'Afrique du Nord aux 23 petits-fils d'immigrés du Maghreb prénommés « Yanis ». Il apparaît alors que, sur l'échantillon maximal non pondéré de l'enquête TeO, seulement 2 des 23 « Yanis » sont probablement descendants de rapatriés, c'est-à-dire que 21 des 23 « Yanis » seraient correctement catégorisés comme descendants d'immigrés (tableau 7.12). Concrètement, pour la plupart des « Yanis », leur grand-parent immigré (G1) du Maghreb a parlé à son enfant en arabe ainsi qu'en français ; leur grand-parent immigré (G1) est musulman, tout comme leur parent enquêté (G2) ; et leur parent enquêté (G2) se dit « pas du tout d'accord » ou « plutôt pas d'accord » avec l'idée qu'on le voit comme un « Français ».

De ces analyses (tableaux 7.10, 7.11 et 7.12), je conclus que **9 des 11 « Nicolas » que l'enquête TeO catégorise comme petits-fils d'immigrés du Maghreb sont vraisemblablement des descendants de rapatriés d'Afrique du Nord**. Cela ne paraît pas particulièrement surprenant : si le lecteur de ces lignes ne connaît probablement pas de descendant d'immigrés du Maghreb prénommé « Nicolas », il connaît possiblement au moins deux descendants de rapatriés d'Algérie prénommés « Nicolas » : le journaliste Nicolas Demorand, né en 1971 d'une mère pied-noir, et le réalisateur Nicolas Bedos, né en 1979 d'un père pied-noir.

---

Français, contre 39 % des enquêtés enfants d'immigrés du Maghreb.

**Tableau 7.11. Profils des 11 petits-fils d'immigrés du Maghreb prénommés « Nicolas », d'après l'enquête TeO**

	Langue(s) que le G1 Maghreb parlait au G2 Maghreb dans son enfance	Religion du G1 Maghreb	Religion du G2 Maghreb	Opinion de l'enquête parent de « Nicolas » sur : « On me voit comme un Français »	Conclusion : ce « Nicolas » est-il descendant de rapatriés ?
<b>1</b>	s	s	s	s	non
<b>2</b>	s	s	s	s	non
<b>3</b>	français seulement	sans religion	sans religion	tout à fait d'accord	oui
<b>4</b>	français seulement		chrétien	tout à fait d'accord	oui
<b>5</b>			chrétien	tout à fait d'accord	oui
<b>6</b>	français seulement	sans religion	chrétien	tout à fait d'accord	oui
<b>7</b>	français seulement	sans religion	chrétien	tout à fait d'accord	oui
<b>8</b>	français seulement	chrétien	chrétien	tout à fait d'accord	oui
<b>9</b>	français seulement	chrétien	chrétien	tout à fait d'accord	oui
<b>10</b>	français seulement	chrétien	chrétien	tout à fait d'accord	oui
<b>11</b>	s	s	s	s	oui

Lecture : Pour aboutir à la conclusion qu'un petit-fils d'immigrés (G3) du Maghreb prénommé « Nicolas » est probablement un descendant de rapatriés plutôt qu'un descendant d'immigrés, il doit remplir 4 conditions : (i) l'immigré (G1) du Maghreb ne parlait ni arabe ni berbère à ses enfants, ni en 1<sup>e</sup> ni en 2<sup>e</sup> langue ; (ii) l'immigré (G1) du Maghreb n'est pas musulman ; (iii) l'enfant d'immigrés (G2) du Maghreb n'est pas musulman non plus ; et (iv) l'enquête parent de « Nicolas » est « plutôt d'accord » ou « tout à fait d'accord » avec l'idée que « on le voit comme un Français ».

Champ : individus du champ maximal qui sont petits-fils d'immigrés (G3) du Maghreb et qui se prénommant « Nicolas », le 2<sup>e</sup> prénom le plus souvent porté par les petits-fils d'immigrés du Maghreb d'après Coulmont et Simon 2019 (tableau p. 3), données non pondérées, N = 11.

Source : TeO (voir Annexe 5B, p. 183).

NB : Conformément aux règles du secret statistique du [CASD](#) et de l'[INSEE](#), et pour qu'aucune information ne permette d'identifier directement ou indirectement une personne physique (ni d'inférer de nouvelles informations sur elle), les résultats obtenus sur les deux premiers et le dernier « Nicolas » sont notés « s » (comme « secret »).

**Tableau 7.12. Profils des 23 petits-fils d'immigrés du Maghreb prénommés « Yanis », d'après l'enquête TeO**

	Langue(s) que le G1 Maghreb parlait au G2 Maghreb dans son enfance	Religion du G1 Maghreb	Religion du G2 Maghreb	Opinion de l'enquête parent de « Yanis » sur : « On me voit comme un Français »	Conclusion : ce « Yanis » est-il descendant de rapatriés ?
1	arabe seulement	musulman	musulman	pas du tout d'accord	non
2	arabe seulement	musulman	musulman	plutôt pas d'accord	non
3	arabe et français	musulman	musulman	pas du tout d'accord	non
4	arabe et français	musulman	musulman	pas du tout d'accord	non
5	arabe et français	musulman	musulman	pas du tout d'accord	non
6	arabe et français	musulman	musulman	pas du tout d'accord	non
7	arabe et français	musulman	musulman	plutôt pas d'accord	non
8	arabe et français	musulman	musulman	tout à fait d'accord	non
9	arabe et français	musulman	sans religion	plutôt pas d'accord	non
10	arabe et français	musulman	sans religion	tout à fait d'accord	non
11	français et arabe	musulman	musulman	plutôt d'accord	non
12	français et arabe	musulman	musulman	tout à fait d'accord	non
13	français et arabe	musulman		pas du tout d'accord	non
14	français et arabe	musulman		plutôt pas d'accord	non
15	français et arabe	musulman	sans religion	pas du tout d'accord	non
16			musulman	tout à fait d'accord	non
17			musulman	pas du tout d'accord	non
18	français seulement	musulman	sans religion	pas du tout d'accord	non
19	français seulement	musulman	sans religion	tout à fait d'accord	non
20	français seulement	sans religion	musulman	pas du tout d'accord	non
21	français seulement	sans religion	sans religion	pas du tout d'accord	non
22	s	s	s	s	oui
23	s	s	s	s	oui

Lecture : Pour aboutir à la conclusion qu'un petit-fils d'immigrés (G3) du Maghreb prénommé « Yanis » est probablement un descendant de rapatriés plutôt qu'un descendant d'immigrés, il doit remplir 4 conditions : (i) l'immigré (G1) du Maghreb ne parlait ni arabe ni berbère à ses enfants, ni en 1<sup>e</sup> ni en 2<sup>e</sup> langue ; (ii) l'immigré (G1) du Maghreb n'est pas musulman ; (iii) l'enfant d'immigrés (G2) du Maghreb n'est pas musulman non plus ; et (iv) l'enquête parent de « Yanis » est « plutôt d'accord » ou « tout à fait d'accord » avec l'idée que « on le voit comme un Français ».

Champ : individus du champ maximal qui sont petits-fils d'immigrés (G3) du Maghreb et qui se prénommant « Yanis », le prénom le plus souvent porté par les petits-fils d'immigrés du Maghreb d'après Coulmont et Simon 2019 (tableau p. 3), données non pondérées, N = 23.

Source : TeO (voir Annexe 5B, p. 183).

NB : Conformément aux règles du secret statistique du [CASD](#) et de l'[INSEE](#), et pour qu'aucune information ne permette d'identifier directement ou indirectement une personne physique (ni d'inférer de nouvelles informations sur elle), les résultats obtenus sur les deux derniers « Yanis » sont notés « s » (comme « secret »).



❖ **Combien de descendants d’immigrés du Maghreb de l’enquête TeO sont, en réalité, des descendants de pieds-noirs ?**

Pour savoir quels immigrés (G1) ou descendants d’immigrés (G2 ou G3) du Maghreb sont plus probablement des rapatriés ou des descendants de rapatriés d’Afrique du Nord — et pour pouvoir ensuite reproduire le tableau de Coulmont et Simon (2019, p. 3) corrigé de certaines de ses erreurs —, j’applique les 4 critères de détection des rapatriés et descendants de rapatriés d’Afrique du Nord non plus seulement aux petits-enfants d’immigrés (G3) du Maghreb prénommés « Yanis » ou « Nicolas », mais à tous les individus de l’échantillon retenu que TeO catégorise comme originaires du Maghreb (G1, G2 ou G3). Je catégorise un individu originaire du Maghreb (G1, G2 ou G3) comme probablement *rapatrié ou descendant de rapatriés d’Afrique du Nord* si son entourage familial<sup>28</sup> remplit les 4 conditions suivantes :

- i. les parents originaires du Maghreb de l’enquête qui est immigré (G1) ou enfant d’immigrés (G2) du Maghreb ne lui parlaient, dans son enfance, ni arabe ni berbère ;
- ii. les parents originaires du Maghreb de l’enquête qui est immigré (G1) ou enfant d’immigrés (G2) du Maghreb ne sont pas musulmans ;
- iii. l’enquête qui est immigré (G1) ou enfant d’immigrés (G2) du Maghreb n’est pas musulman non plus ; et
- iv. l’enquête qui est immigré (G1) ou enfant d’immigrés (G2) du Maghreb se déclare « plutôt d’accord » ou « tout à fait d’accord » avec l’idée que « on le voit comme un Français ».

Si l’entourage familial d’un individu originaire du Maghreb ne remplit pas chacune de ces 4 conditions, je conserve la catégorisation de l’individu comme *immigré ou descendant d’immigrés du Maghreb*.

Sur l’échantillon TeO retenu pondéré (représentatif des personnes de 18-50 / 18-60 ans en ménage ordinaire et de leurs enfants de 0-17 ans habitant en France métropolitaine en 2008), il apparaît que 9 % des individus catégorisés comme immigrés ou descendants d’immigrés du Maghreb seraient probablement des descendants de rapatriés d’Afrique du Nord (273 / (2 881 + 273)) (tableau 7.13). Précisément, 1 % des immigrés (G1) du Maghreb seraient des pieds-noirs et 10 % des enfants d’immigrés (G2) du Maghreb seraient des enfants de pieds-noirs, mais surtout **26 % des petits-enfants d’immigrés (G3) du Maghreb seraient des petits-enfants de pieds-noirs** (tableau 7.13). En effet, comme les rapatriés d’Afrique du Nord sont par définition nés avant 1962, alors que dans l’enquête TeO les immigrés du Maghreb peuvent être nés jusqu’en 1990,<sup>29</sup> les rapatriés d’Afrique du Nord sont en moyenne bien plus âgés que les immigrés du Maghreb, si bien qu’en moyenne ils ont eu plus de temps d’avoir leurs petits-enfants, ce qui explique la sur-présence de leurs petits-enfants parmi les petits-enfants du groupe réunissant rapatriés d’Afrique du Nord et immigrés du Maghreb.<sup>30</sup> On comprend mieux pourquoi le Top 3 de Coulmont et Simon (2019, p. 3) comprend le prénom « Nicolas » chez les

---

<sup>28</sup> Sur l’échantillon TeO retenu (comme sur l’échantillon maximal), les informations sur l’entourage d’un individu originaire du Maghreb sont renseignées pour (i) tous les habitants du logement de l’enquête, (ii) tous les autres enfants de l’enquête, vivants ou non et qu’ils habitent dans le logement de leur parent enquêté ou non, et (iii) le conjoint actuel de l’enquête (s’il en a un) et son premier conjoint (s’il est différent du conjoint actuel) ainsi que les parents de l’enquête et de son conjoint actuel et de son premier conjoint, vivants ou non et qu’ils habitent en France métropolitaine ou non.

<sup>29</sup> En outre, les rapatriés pouvaient, lors de leur arrivée en France, être de tous âges alors que les immigrés sont le plus souvent de jeunes adultes.

<sup>30</sup> Sur l’échantillon maximal pondéré, les petits-fils d’immigrés (G3) du Maghreb qui sont prénommés « Nicolas » et « Yanis » (tous nés en métropole de parents eux-mêmes presque tous nés en métropole) ont ainsi des âges très différents : lors de l’enquête (2008), les « Nicolas » ont en moyenne 21 ans, tandis que les « Yanis » ont en moyenne 4 ans.

petits-enfants d’immigrés du Maghreb *mais seulement chez les petits-enfants* : c’est que, par rapport aux membres des autres « générations » originaires du Maghreb, les petits-enfants incluent relativement beaucoup plus de descendants de rapatriés mal catégorisés.

D’après ces mêmes estimations, l’enquête TeO sous-estimerait le nombre de rapatriés et de leurs descendants de 17 % ( $273 / (273 + 1\,332)$ ) (tableau 7.13). Les sources indépendantes de données sur l’effectif de pieds-noirs ne permettent malheureusement pas de vérifier si l’enquête TeO sous-estime effectivement d’environ 17 % le nombre de rapatriés et descendants de rapatriés en métropole en 2008.<sup>31</sup> Il semble en revanche établi que 26 % des petits-enfants d’immigrés du Maghreb de l’enquête TeO sont en réalité des petits-enfants de pieds-noirs. On notera qu’une erreur d’ampleur globalement modeste (qui porte sur 9 % des immigrés et descendants d’immigrés du Maghreb et 17 % des rapatriés et descendants de rapatriés d’Afrique du Nord de l’enquête TeO) a pu suffire à mal catégoriser 26 % des *petits-enfants* d’immigrés du Maghreb et donc à affecter le Top 3 et le Top 20 de leurs prénoms.

Au total, plusieurs analyses (tableaux 7.10 à 7.13) semblent converger vers la conclusion suivante : manifestement, non seulement **la plupart (9/11) des « Nicolas » du tableau de Coulmont et Simon (2019, p. 3) sont des petits-fils de rapatriés d’Afrique du Nord** que l’enquête TeO a catégorisés à tort comme petits-fils d’immigrés du Maghreb, mais **plus généralement, 26 % des petits-enfants d’immigrés du Maghreb de l’enquête TeO sont vraisemblablement des petits-enfants de pieds-noirs.**

---

<sup>31</sup> D’après mes recherches, les deux seules sources qui estiment l’effectif de pieds-noirs en France autour de 2008 sont Couto 2013 et Fourquet 2014. D’une part, Couto (2013) exploite l’Échantillon démographique permanent pour suivre, de 1968 à 1999, les effectifs de « pieds-noirs », d’« immigrés intercensitaires » et de « métropolitains », tous ces individus étant nés avant 1953 et âgés de moins de 60 ans. Elle définit les « pieds-noirs » comme les individus nés français (avant 1953) en Algérie, y habitant au 1<sup>er</sup> janvier 1962 et recensés en France métropolitaine en 1968 (mais pas désignés comme « musulmans » ou nationaux algériens), et âgés de moins de 60 ans (Couto 2013, p. 103). D’autre part, Fourquet (2014) exploite un cumul d’enquêtes de l’IFOP qui demandaient aux enquêtés : « Êtes-vous pied-noir ? » et « Un ou plusieurs de vos parents / grands-parents sont-ils / étaient-ils pieds-noirs ? ». Cette question ne précise pas si, par « pied-noir », l’enquêté était censé n’inclure que les pieds-noirs d’Algérie ou aussi ceux du Maroc et de Tunisie. Quoi qu’il en soit, Fourquet (2014) observe que parmi les personnes inscrites sur liste électorale en France en 2012, 1,8 % seraient pied-noir et 5,2 % auraient un parent ou grand-parent pied-noir (Fourquet 2014, p. 2).

**Tableau 7.13. 26 % des petits-enfants d’immigrés du Maghreb de l’enquête TeO sont vraisemblablement des petits-enfants de pieds-noirs**

			G1	G2	G3	Total
a	<b>Individu que TeO catégorise comme immigré ou descendant d’immigrés du Maghreb</b>	Individu <b>bien catégorisé</b> comme immigré ou descendant d’immigrés du Maghreb	1 031	1 602	248	2 881
b		Individu <b>mal catégorisé</b> comme immigré ou descendant d’immigrés du Maghreb, et qui est probablement rapatrié ou descendant de rapatriés d’Afrique du Nord	11	175	87	273
c	<b>Individu que TeO catégorise comme rapatrié ou descendant de rapatriés d’Afrique du Nord</b>		181	739	412	1 332
d	<b>Individu que TeO catégorise autrement</b>		2 441	27 050	8 231	37 721
e	<b>Total</b>		3 665	29 566	8 977	42 208
b / (a+b)	<b>Part d’individus que l’enquête TeO catégorise comme immigré ou descendant d’immigrés du Maghreb mais qui sont probablement rapatrié ou descendant de rapatriés d’Afrique du Nord</b>		1 %	10 %	26 %	9 %
b / (b+c)	<b>Part de rapatriés et descendants de rapatriés d’Afrique du Nord qui, d’après mes analyses, ne sont pas catégorisés comme tels par l’enquête TeO</b>		6 %	19 %	17 %	17 %

Lecture : 26 % des petits-enfants d’immigrés du Maghreb de l’enquête TeO sont vraisemblablement des petits-enfants de rapatriés d’Afrique du Nord, et 17 % des rapatriés et descendants de rapatriés d’Afrique du Nord sont vraisemblablement catégorisés comme immigrés et descendants d’immigrés du Maghreb.

Champ : personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008, données pondérées, N = 42 208.

Source : TeO (voir Annexe 5B, p. 183).

#### ❖ La reproduction « au plus près » du tableau p. 3, sans les erreurs

Me voilà désormais en mesure de distinguer le Top 20 des prénoms des petits-enfants d’immigrés du Maghreb (tableau 7.4), selon qu’ils sont effectivement petits-enfants d’immigrés du Maghreb ou plus probablement petits-enfants de rapatriés d’Afrique du Nord. Une telle opération produit des résultats relativement impressionnants (tableau 7.14).

En effet, sur l’échantillon TeO retenu pondéré (représentatif des personnes de 18-50 / 18-60 ans en ménage ordinaire et de leurs enfants de 0-17 ans habitant en France métropolitaine en 2008), les prénoms les plus portés par les individus que notre méthode identifie comme petits-enfants de pieds-noirs sont, d’après la nomenclature de l’INED, presque tous (38/40) « français », « latins » ou « internationaux ou anglophones » (en grisé sur le tableau 7.14), tandis que les prénoms les plus portés par les petits-enfants d’immigrés du Maghreb sont pour la majorité (24/40) « arabo-musulmans » (en blanc). Notre méthode de détection des descendants de rapatriés d’Afrique du Nord ne tient pas compte de leur prénom pour déterminer s’ils sont probablement descendants d’immigrés ou de rapatriés, mais elle a bel et bien pour effet de séparer deux stocks de prénoms différents : **seuls 1 prénom masculin et 2 prénoms féminins se retrouvent à la fois dans le Top 20 des petits-enfants d’immigrés du Maghreb et dans le Top 20 des petits-enfants de pieds-noirs**. Ces stocks de prénoms largement disjoints vont de pair avec des origines religieuses très différentes : les petits-enfants d’immigrés du Maghreb sont en majorité d’origine musulmane (64 %), tandis que ceux qui sont plus probablement petits-enfants de pieds-noirs sont, pour la grande majorité, soit d’origine

chrétienne (51 %) soit sans origine religieuse (34 %), c'est-à-dire qu'aucun de leurs parents n'est affilié à une religion.

Les petits-fils de pieds-noirs portent presque exclusivement des prénoms « français », « latins » ou « internationaux ou anglophones », tandis que les petits-fils d'immigrés du Maghreb portent presque exclusivement des prénoms « arabo-musulmans ». Chez les petites-filles, toutefois, la situation est sensiblement différente. Si les prénoms les plus portés par les petites-filles de pieds-noirs sont presque tous (19/20) « français », « latins » ou « internationaux ou anglophones » (en grisé sur le tableau 7.14), les prénoms les plus portés par les petites-filles d'immigrés du Maghreb sont moins souvent arabo-musulmans (7/20) que ceux de leurs frères (17/20) : elles portent relativement souvent des prénoms d'origine multiple ou ambiguë comme « Sarah » et « Inès », mentionnés dans le tableau de Coulmont et Simon (2019, p. 3). Ces résultats sont commentés plus en détail ci-après.

Si « Nicolas » est le prénom masculin le plus porté dans ce groupe de petits-fils de pieds-noirs, en revanche il n'apparaît pas dans le Top 20 des prénoms les plus portés par les petits-fils d'immigrés du Maghreb (tableau 7.14).<sup>32</sup> Plus généralement, « Nicolas » et les autres prénoms « français » non ambigus sont précisément le type de prénom qui, dans le Top 20 des prénoms des petits-fils d'immigrés du Maghreb, brille par son absence, alors que tel n'est pas le cas chez les descendants d'immigrés d'Europe du Sud (tableau 7.3). **Chez les petits-fils d'immigrés (G3) du Maghreb en métropole en 2008, les prénoms les plus fréquents seraient non pas « Yanis » et « Nicolas » mais « Karim » et « Nassim ».**

Que l'on admette *ou non* que bon nombre des petits-enfants d'immigrés du Maghreb de l'enquête TeO sont en réalité des petits-enfants de pieds-noirs, on notera que les petits-enfants d'immigrés du Maghreb qui portent des prénoms « français » non ambigus ont une ascendance assez particulière : ni arabophone, ni berbérophone, ni musulmane, l'enquêté de la famille étant en outre « plutôt d'accord » ou « tout à fait d'accord » avec l'idée qu'« on le voit comme un Français ». Voilà des informations qu'il n'aurait pas été inutile de mentionner dans l'article de Coulmont et Simon (2019) et de l'INED.

---

<sup>32</sup> Précisément, sur les 11 petits-fils d'immigrés (G3) du Maghreb prénommés « Nicolas » que compte l'échantillon maximal non pondéré (tableau 7.11), « seuls » 10 sont inclus dans l'échantillon retenu (tableaux 7.14 et 7.15), car le 11<sup>e</sup> est le fils d'un enquêté du groupe « majoritaire » et d'un conjoint d'enquêté qui est enfants d'immigrés (G2) du Maghreb, si bien qu'il est exclu de l'échantillon retenu pour éviter les doubles comptes (tableau 5.2). Sur les 10 petits-fils d'immigrés (G3) du Maghreb prénommés « Nicolas » que compte l'échantillon retenu non pondéré, 8/10 sont probablement descendants de rapatriés (comme indiqué dans le tableau 7.14), et 2/10 sont probablement descendants d'immigrés (ce qui n'est pas indiqué dans le tableau 7.14, car ces 2 « Nicolas » ne suffisent pas pour que le prénom « Nicolas » fasse partie du Top 20 des prénoms des petits-enfants d'immigrés du Maghreb).

**Tableau 7.14. Top 20 des prénoms des petits-enfants d’immigrés du Maghreb, selon qu’ils sont effectivement petits-enfants d’immigrés du Maghreb ou plus probablement petits-enfants de rapatriés d’Afrique du Nord, en France en 2008 (enquête TeO)**

		G3 Maghreb - total (N = 335)			G3 Maghreb probablement petits-enfants de rapatriés (N = 87)			G3 Maghreb probablement petits-enfants d’immigrés (N = 248)		
		Prénom	Effectif non pondéré	% pondéré	Prénom	Effectif non pondéré	% pondéré	Prénom	Effectif non pondéré	% pondéré
TOP 20 DES PRÉNOMS MASCULINS	1	Karim	s	4,5 %	Nicolas	8	11,2 %	Karim	s	6,3 %
	2	Nassim	9	3,6 %	Jérémy	s	9,3 %	Nassim	9	5,0 %
	3	Farès	8	3,5 %	Antony	s	2,2 %	Farès	8	5,0 %
	4	Nicolas	10	3,5 %	Romain	5	2,1 %	Mourad	s	3,5 %
	5	Jérémy	s	2,7 %	Thomas	5	2,1 %	Yanis	18	2,5 %
	6	Mourad	s	2,5 %	Ethan	s	1,9 %	Mehdi	14	2,3 %
	7	Yanis	19	1,9 %	Benoît	s	1,9 %	Kevin	8	1,5 %
	8	Mehdi	16	1,8 %	Théo	5	1,6 %	Yacine	8	1,4 %
	9	Kevin	9	1,1 %	Martin	s	1,4 %	Ilyès	6	1,3 %
	10	Yacine	8	1,0 %	Yannick	s	1,3 %	Rayan	7	1,3 %
	11	Ilyès	6	1,0 %	Guillaume	s	1,3 %	Mohamed	9	1,1 %
	12	Rayan	7	0,9 %	Oscar	s	1,1 %	Rayane	6	1,1 %
	13	Benoît	s	0,9 %	Aurélien	s	1,0 %	Adam	7	1,0 %
	14	Driss	6	0,8 %	Jonathan	s	1,0 %	s	s	1,0 %
	15	Mohamed	9	0,8 %	s	s	1,0 %	s	s	1,0 %
	16	Rayane	6	0,8 %	s	s	1,0 %	Marwan	s	0,9 %
	17	Sofian	6	0,7 %	s	s	1,0 %	Ryad	s	0,9 %
	18	Adam	7	0,7 %	Gaëtan	s	1,0 %	Sofian	5	0,8 %
	19	s	s	0,7 %	Tristan	s	1,0 %	Driss	5	0,8 %
	20	s	s	0,7 %	Eliott	s	1,0 %	Younès	s	0,8 %
TOP 20 DES PRÉNOMS FÉMININS	1	Sarah	25	5,0 %	s	s	3,5 %	Sarah	22	6,1 %
	2	Sophia	s	3,0 %	Virginie	s	2,9 %	Sophia	s	3,8 %
	3	Sabrina	s	2,6 %	Jessica	s	2,9 %	Lina	17	3,1 %
	4	Lina	17	2,4 %	Marlène	s	2,0 %	Sabrina	s	2,9 %
	5	Inès	17	2,3 %	Clara	s	1,9 %	Inès	16	2,8 %
	6	Sofia	9	1,7 %	Carla	s	1,8 %	Sofia	8	1,9 %
	7	Celia	11	1,5 %	Ludivine	s	1,7 %	Célia	10	1,7 %
	8	Jessica	7	1,1 %	Louna	s	1,7 %	Leila	6	1,4 %
	9	Leila	6	1,1 %	Camille	s	1,6 %	Myriam	10	1,3 %
	10	Myriam	10	1,0 %	s	s	1,5 %	Anaïs	8	1,1 %
	11	Anaïs	9	1,0 %	Rachelle	s	1,5 %	Lilia	5	1,1 %
	12	Virginie	s	0,9 %	Maud	s	1,5 %	Aïcha	s	1,1 %
	13	Melissa	8	0,9 %	Sabrina	s	1,4 %	s	s	1,1 %
	14	Lilia	5	0,9 %	Kathleen	s	1,4 %	Yasmine	5	1,0 %
	15	Aïcha	s	0,9 %	s	s	1,4 %	s	s	1,0 %
	16	Lisa	7	0,8 %	s	s	1,4 %	Melissa	6	0,9 %
	17	Emma	8	0,8 %	Alizée	s	1,3 %	Kaina	s	0,9 %
	18	s	s	0,8 %	s	s	1,3 %	Farah	s	0,9 %
	19	Lola	5	0,8 %	Lisa	s	1,3 %	Jasmine	6	0,8 %
	20	s	s	0,8 %	Lola	s	1,3 %	Maeva	5	0,8 %

Lecture : le prénom masculin le plus porté par les petits-fils d’immigrés (G3) du Maghreb qui sont probablement petits-fils de rapatriés est « Nicolas » (porté par 11,2 % de ces probables descendants de rapatriés), tandis que le prénom masculin le plus porté par les petits-fils d’immigrés (G3) du Maghreb qui sont effectivement petits-fils d’immigrés est « Karim » (porté par 6,3 % de ces probables descendants d’immigrés). Sont grisés les prénoms qui, d’après la nomenclature de l’INED, ne sont pas « arabo-musulmans ».

Champ : individus du champ retenu (personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008) qui sont petits-enfants d’immigrés du Maghreb, données pondérées, N = 335.

Source : TeO (voir Annexe 5B, p. 183).

NB : Conformément aux règles du secret statistique du [CASD](#) et de l’[INSEE](#), et pour qu’aucune information ne permette d’identifier directement ou indirectement une personne physique (ni d’inférer de nouvelles informations sur elle), les résultats obtenus sur des effectifs faibles (<5) sont notés « s » (comme « secret »).

**Tableau 7.15. Top 20 des prénoms des habitants originaires du Maghreb selon leur « génération », en France en 2008 (enquête TeO)**

		G1 Maghreb (N = 1 031)			G2 Maghreb (N = 1 602)			G3 Maghreb (N = 248)		
		Prénom	Effectif non pondéré	% pondéré	Prénom	Effectif non pondéré	% pondéré	Prénom	Effectif non pondéré	% pondéré
TOP 20 DES PRÉNOMS MASCULINS	1	Mohamed	103	9,5 %	Mohamed	105	4,0 %	Karim	s	6,3 %
	2	Ahmed	30	3,2 %	Karim	52	2,3 %	Nassim	9	5,0 %
	3	Rachid	20	2,1 %	Farid	37	2,2 %	Farès	8	5,0 %
	4	Mohammed	16	1,8 %	Rachid	27	1,7 %	Mourad	s	3,5 %
	5	Abdelkader	18	1,5 %	Mehdi	46	1,6 %	Yanis	18	2,5 %
	6	Abdellah	12	1,4 %	Kamel	21	1,3 %	Mehdi	14	2,3 %
	7	Khalid	10	1,3 %	Samir	31	1,3 %	Kevin	8	1,5 %
	8	Mourad	13	1,3 %	Adam	35	1,2 %	Yacine	8	1,4 %
	9	Mustapha	16	1,3 %	Sofiane	36	1,2 %	Ilyès	6	1,3 %
	10	Ali	16	1,3 %	Yanis	37	1,1 %	Rayan	7	1,3 %
	11	Brahim	9	1,2 %	Yassine	31	1,0 %	Mohamed	9	1,1 %
	12	Hamid	9	1,2 %	Saïd	20	1,0 %	Rayane	6	1,1 %
	13	Hicham	10	1,1 %	Amine	36	1,0 %	Adam	7	1,0 %
	14	Saïd	13	1,1 %	Youssef	25	1,0 %	s	s	1,0 %
	15	Omar	10	1,1 %	Ahmed	20	1,0 %	s	s	1,0 %
	16	Walid	s	1,0 %	Yacine	20	0,8 %	Marwan	s	0,9 %
	17	Nabil	7	1,0 %	Nordine	16	0,7 %	Ryad	s	0,9 %
	18	Driss	12	1,0 %	Samy	21	0,7 %	Sofian	5	0,8 %
	19	Hassan	8	0,9 %	Hakim	14	0,6 %	Driss	5	0,8 %
	20	Karim	6	0,9 %	Abdelkader	13	0,6 %	Younès	s	0,8 %
TOP 20 DES PRÉNOMS FÉMININS	1	Fatima	70	6,1 %	Nadia	70	3,1 %	Sarah	22	6,1 %
	2	Khadija	27	3,1 %	Sarah	91	2,9 %	Sophia	s	3,8 %
	3	Samira	16	2,5 %	Myriam	72	2,3 %	Lina	17	3,1 %
	4	Fatiha	28	2,5 %	Fatima	60	2,1 %	Sabrina	s	2,9 %
	5	Aïcha	23	1,9 %	Yamina	14	1,7 %	Inès	16	2,8 %
	6	Malika	17	1,6 %	Inès	61	1,5 %	Sofia	8	1,9 %
	7	Nadia	14	1,5 %	Sonia	35	1,5 %	Célia	10	1,7 %
	8	Zohra	17	1,5 %	Malika	28	1,4 %	Leila	6	1,4 %
	9	Naima	17	1,5 %	Sabrina	33	1,4 %	Myriam	10	1,3 %
	10	Karima	17	1,4 %	Nora	28	1,3 %	Anaïs	8	1,1 %
	11	Samia	18	1,4 %	Samira	25	1,3 %	Lilia	5	1,1 %
	12	Houria	13	1,2 %	Anissa	33	1,1 %	Aïcha	s	1,1 %
	13	Rachida	14	1,2 %	Leila	30	1,1 %	s	s	1,1 %
	14	Meriem	7	1,1 %	Houria	s	1,0 %	Yasmine	5	1,0 %
	15	Saïda	13	1,1 %	Samia	28	1,0 %	s	s	1,0 %
	16	Fatma	13	1,0 %	Yasmina	25	0,9 %	Melissa	6	0,9 %
	17	Bouchra	8	1,0 %	Lina	29	0,8 %	Kaina	s	0,9 %
	18	Yamina	12	1,0 %	Yasmine	28	0,8 %	Farah	s	0,9 %
	19	Jamila	8	1,0 %	Aïcha	21	0,8 %	Jasmine	6	0,8 %
	20	Najat	10	1,0 %	Karima	20	0,7 %	Maeva	5	0,8 %

Lecture : le prénom masculin le plus porté par les immigrés (G1) du Maghreb est « Mohamed », porté par 9,5 % d'entre eux. Sont grisés les prénoms qui, d'après la nomenclature de l'INED, ne sont pas « arabo-musulmans ».

Champ : individus du champ retenu (personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008) qui sont immigrés (G1), enfants d'immigrés (G2) ou petits-enfants d'immigrés (G3) du Maghreb, données pondérées, N = 2 881.

Source : TeO (voir Annexe 5B, p. 183).

NB : Conformément aux règles du secret statistique du [CASD](#) et de l'[INSEE](#), et pour qu'aucune information ne permette d'identifier directement ou indirectement une personne physique (ni d'inférer de nouvelles informations sur elle), les résultats obtenus sur des effectifs faibles (<5) sont notés « s » (comme « secret »).



On peut maintenant reproduire le Top 3 (Coulmont et Simon 2019, p. 3) et le Top 20 (tableau 7.4) des prénoms des individus originaires du Maghreb, expurgés de leurs erreurs (tableau 7.15). Je reproduis ces analyses sur les individus de l'échantillon retenu pondéré (personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008) qui sont immigrés (G1), enfants d'immigrés (G2) ou petits-enfants d'immigrés (G3) du Maghreb. Dans les commentaires des résultats de ce tableau 7.15, j'appelle « arabe » un prénom usuellement attribué par des arabophones ; j'appelle « d'origine (ou d'étymologie) arabe » un prénom sémantiquement transparent en arabe (Annexe 6A, p. 267) ; et j'appelle « arabo-musulman » un prénom que la nomenclature des prénoms de l'INED catégorise comme tel (Annexe 6B, p. 285).<sup>33</sup>

### Les prénoms des immigrés (G1) du Maghreb

Parmi les immigrés (G1) du Maghreb, les 20 prénoms masculins et les 20 prénoms féminins les plus portés sont presque tous catégorisés comme « arabo-musulmans » (en blanc), comme « Mohamed », « Ahmed » et « Rachid » ou encore « Fatima », « Khadija » et « Samira » (tableau 7.15). Le seul de ces 40 prénoms qui ne soit pas catégorisé par la nomenclature de l'INED comme « arabo-musulman » est « Nadia », un prénom européen d'origine slave qui signifie « espoir », mais qui est aussi l'orthographe française courante du prénom arabe « Nādiya » (qui appelle) (Annexe 6B, p. 285). La très grande majorité des prénoms les plus portés par les immigrés du Maghreb sont des prénoms arabes d'étymologie arabe. Conformément aux connaissances acquises sur les prénoms portés au Maghreb (Annexe 6A, p. 267), les prénoms les plus portés ne sont que rarement berbères : seul le prénom « Fatma » peut être une forme berbérisée du prénom arabe « Fatima » (comme « Fadna »), mais c'est aussi la prononciation courante en arabe maghrébin du prénom arabe classique « Fatima ». Ces résultats confirment aussi que peu d'immigrés du Maghreb ont francisé leur prénom, par contraste avec les immigrés d'Europe du Sud (tableau 7.3) ou d'Asie (Annexe 6B, p. 285).

Parmi les prénoms les plus portés par les immigrés du Maghreb (tableau 7.15), se trouvent ainsi :

- ✓ des prénoms coraniques d'origine arabe, surtout masculins, comme « Mohamed » et « Mohammed » (le très loué, du nom du prophète de l'islam), et « Ahmed » (le plus loué, autre nom du prophète), mais aussi « Houria » (femmes au fond de l'œil très blanc, faisant apparaître la prunelle très noire<sup>34</sup>) ; des prénoms coraniques d'origine hébraïque, comme « Brahim » ou « Meriem » ; des prénoms théophores d'origine arabe, comme « Abdelkader » (serviteur du puissant, du nom du résistant du 19<sup>e</sup> siècle à la colonisation française de l'Algérie) et « Abdellah » (serviteur de Dieu) ; et d'autres prénoms religieux comme « Mustapha » (l' élu), « Driss » (du nom d'un prophète traditionnellement associé à Énoch) et « Fatiha » (du nom de la sourate débutant le Coran) ;
- ✓ des prénoms arabes de proches ou successeurs du prophète Mohamed, comme « Rachid » (de bon conseil, le pluriel désignant les 4 premiers califes « bien-guidés »), « Ali » (cousin et gendre du prophète et 4<sup>e</sup> calife), « Saïd » (compagnon du prophète), « Omar » (2<sup>e</sup> calife), « Hassan » (petit-fils du prophète), « Fatima »

<sup>33</sup> Les thèses soutenues dans le présent document n'engagent personne d'autre que moi et je suis seul responsable de toute erreur. Je tiens toutefois à remercier chaleureusement Manel Khelifa, arabophone, Mohamed Lahrouchi, linguiste spécialiste des langues berbères, Pierre Larcher, linguiste spécialiste de l'arabe, et Muriel Roiland, spécialiste d'onomastique arabe médiévale, de m'avoir aidé dans l'analyse des résultats présentés ci-dessous.

<sup>34</sup> L'expression hūr 'în, qu'on trouve dans le Coran en 44, 54 et 52, 20, désigne des femmes au fond de l'œil très blanc (hūr) faisant apparaître la prunelle très noire ('în), à l'origine des « houris » (les vierges du Paradis).

- ou « Fatma » (fille du prophète, épouse du 4<sup>e</sup> calife et mère de Hasan et Husayn), « Khadija » (1<sup>e</sup> épouse du prophète) et « Aïcha » (3<sup>e</sup> épouse du prophète) ;
- ✓ des prénoms profanes d'origine arabe (majoritairement féminins) comme « Khalid » (éternel), « Mourad » (vœu), « Hamid » (loué), « Hicham » (générosité), « Walid » (né), « Nabil » (talentueux), « Karim » (généreux), « Samira » (compagnon de discussion du soir), « Malika » (reine), « Nadia » (qui appelle), « Zohra » (fleur en floraison), « Naima » (tranquille), « Karima » (généreuse), « Samia » (sublime), « Rachida » (bien guidée), « Saïda » (heureuse), « Bouchra » (joie à l'annonce d'une bonne nouvelle), « Yamina » (fortunée), « Jamila » (belle) et « Najat » (salut).

### **Les prénoms des enfants d'immigrés (G2) du Maghreb**

Parmi les enfants d'immigrés (G2) du Maghreb, les 20 prénoms masculins et les 20 prénoms féminins les plus portés sont là aussi, pour la plupart, catégorisés comme « arabo-musulmans », comme « Mohamed », « Karim » et « Farid » ou encore « Fatima », « Yamina » et « Malika » (tableau 7.15). Le seul des 20 prénoms masculins les plus portés par les fils d'immigrés du Maghreb qui ne soit pas catégorisé par la nomenclature de l'INED comme « arabo-musulman » est « Adam », catégorisé comme « français » mais qui pour les musulmans est le prénom arabe (coranique) d'un prophète (Annexe 6A, p. 267). Les prénoms les plus portés par les filles d'immigrés du Maghreb sont en revanche moins massivement « arabo-musulmans » que ceux de leurs frères<sup>35</sup> : certains sont catégorisés comme « internationaux ou anglophones » (« Nadia », « Sarah », « Sonia » et « Lina »), d'autres comme « latins » (« Inès » et « Sabrina ») et d'autres encore comme « français » (« Myriam ») (Annexe 6A, p. 267 et Annexe 6B, p. 285). Au total, le contraste avec les prénoms des enfants d'immigrés (G2) d'Europe du Sud reste frappant : alors que les enfants d'immigrés d'Europe du Sud n'ont que rarement reçu des prénoms spécifiques à leurs origines (« latins ») et ont même massivement reçu des prénoms « français » (tableau 7.3), les enfants d'immigrés du Maghreb — et tout particulièrement les fils — ont largement reçu des prénoms spécifiques « arabo-musulmans » et n'ont que rarement reçu des prénoms « français » (tableau 7.15).

Parmi les prénoms les plus portés par les enfants d'immigrés du Maghreb (tableau 7.15), se trouvent :

- ✓ des prénoms coraniques d'origine arabe, comme « Mohamed », « Ahmed » et « Houria » (voir ci-dessus) ; des prénoms coraniques d'origine hébraïque, comme « Adam » et « Youssef » ; des prénoms théophores d'origine arabe, comme « Abdelkader » ; et d'autres prénoms religieux d'origine arabe comme « Mehdi » (guidé),<sup>36</sup> « Yassine » et « Yacine » (du nom des deux lettres « yâ » et « sîn » qu'on trouve au début de la 36<sup>e</sup> sourate du Coran) ou « Nordine » (lumière de la religion) ;

<sup>35</sup> « La transmission des origines s'effectue donc davantage au travers des garçons que des filles » (Auger et Moïse 2007, p. 6). « La tendance des couples mixtes à "brouiller les cartes" [attribuer des prénoms d'origine ambiguë comme « Sonia », « Nadia » ou « Sabrina »] plus facilement pour les filles que pour les garçons ne ferait donc que redoubler la tendance plus générale des sociétés et particulièrement des sociétés du Maghreb à tolérer plus facilement l'innovation et l'importation pour les prénoms féminins que masculins, ces derniers restant plus traditionnels » (Streiff-Fénart 1990, p. 13-14). De même à Java (Indonésie), les filles reçoivent plus souvent que les garçons des prénoms hybrides arabo-javanais, alors que les garçons reçoivent plus souvent que les filles des prénoms exclusivement arabes (Kuipers et Askuri 2017, p. 45).

<sup>36</sup> « Mehdi » est un prénom arabe qui signifie « guidé ». *Mahdī* (« guidé ») est le participe passif du verbe *hadā* (« guider, diriger »). Même s'il est absent du Coran, il appartient à une famille lexicale, celle de *hudā* (« direction, guidance »), omniprésente dans le Coran, si bien que sous la forme *al-Mahdī* il est devenu le surnom de personnages religieux, ainsi que le personnage eschatologique attendu par les musulmans.

- ✓ des prénoms arabes de proches ou successeurs du prophète Mohamed, comme « Rachid » (4 premiers califes « bien-guidés »), « Saïd » (compagnon du prophète), « Sofiane » (chef de la tribu des Quraysh), « Amine » (calife abbaside), « Fatima » (fille du prophète, épouse du 4<sup>e</sup> calife et mère de Hassan et Husayn) et « Aïcha » (3<sup>e</sup> épouse du prophète) ;
- ✓ des prénoms profanes d'origine arabe comme « Karim » (généreux), « Farid » (unique), « Kamel » (parfait), « Samir » (compagnon de discussion du soir), « Samy » (sublime), « Hakim » (sage), « Yamina » (fortunée), « Malika » (reine), « Nora » (lumière), « Samira » (compagnon de discussion du soir), « Anissa » (amicale), « Leila » (nuit), « Samia » (sublime), « Lina » (douce) et « Karima » (généreuse) ; et d'autres prénoms profanes d'origine persane comme « Yasmina » et « Yasmine » (jasmin) ;
- ✓ et des prénoms, surtout féminins, mixtes ou ambigus (arabes mais pas seulement), comme « Nadia » (« qui appelle » en arabe, et « espoir » dans des langues slaves), « Sarah » et « Myriam » (prénoms d'origine hébraïque qui peuvent être considérés comme arabes mais aussi comme français), « Inès » (prononciation dialectale du prénom arabe « Īnās » qui signifie « amabilité », mais aussi forme française d'un prénom espagnol d'origine grecque<sup>37</sup>) ; des prénoms d'origine non arabe mais qui pourraient passer pour coraniques, comme « Yanis » (prénom grec d'origine hébraïque, qui pourrait désigner le personnage coranique de Jean le Baptiste<sup>38</sup>) ; des prénoms qui pourraient éventuellement avoir une racine plus ou moins acceptable en arabe, comme « Sabrina » (forme française d'un prénom latin d'origine galloise, mais dans laquelle une oreille arabe pourrait reconnaître « sabr » (patience), « -ina » pouvant alors, par un bricolage linguistique, passer pour une espèce de suffixe) ; d'autres prénoms sans origine arabe ni racine arabe crédible, mais dont tous les phonèmes sont arabes, et qui se terminent par le suffixe féminin de l'arabe (-a), comme « Sonia » (prénom européen d'origine slave).<sup>39</sup>

<sup>37</sup> « Inès » est la forme française du prénom « Inés », qui est la forme espagnole d'un prénom grec qui signifie « chaste » ; sainte Agnès (« Inés » en espagnol, « Inês » en portugais) était une martyre des 3<sup>e</sup> et 4<sup>e</sup> siècles. Mais au Maghreb (où l'on prononce souvent les â longs é), « Ines » est aussi la prononciation dialectale du prénom arabe « Īnās », qui signifie « amabilité, affabilité, courtoisie, urbanité ». Le succès de ce prénom chez les personnes originaires du Maghreb semble donc relever du jeu de mots. C'est ainsi que le prénom « Inès » peut être catégorisé comme « latin » par la nomenclature de l'INED mais comme « arabo-musulman » par la nomenclature synthétique. Comme l'écrit Coulmont (2017b, p. 3) à propos de la France en 2008, « Some first names are given by *Christians and by Muslims*: they are relatively rare Christian names: *Ines* is one of the most frequent names given to daughters of Muslims in France, it is below the 120<sup>th</sup> name given by Christian or nonreligious parents. Same numbers for *Sarah* (one of the top 5 names for daughters of muslims, below the 90<sup>th</sup> name for Christian parents). »

<sup>38</sup> « Yanis » est le cognat grec de « Jean », prénom d'origine hébraïque qui signifie « Dieu est bon ». Jean le Baptiste étant un personnage coranique, la forme « Yanis » a pu paraître à la fois religieusement acceptable tout en ayant l'avantage, comme « Nadia » ou « Inès », de n'être pas exclusivement musulmane. Quoi qu'il en soit, « Yanis » est catégorisé comme « arabo-musulman » par la nomenclature de l'INED aussi bien que par la nomenclature synthétique.

<sup>39</sup> Dans l'Algérie des années 1970 et 1980, des prénoms d'origine slave comme « Sonia », mais aussi « Ludmilla » (Varro et Lesbet 1986, p. 144) ou « Katia », ont connu un certain succès. Les relations de l'Algérie avec l'URSS et l'Europe communiste ayant entraîné des mariages mixtes entre des hommes algériens et des femmes slaves, cela a diffusé en Algérie des prénoms slaves, attribués d'abord à des enfants de couples mixtes, puis plus largement. Même si ces prénoms n'ont pas de racine arabe crédible, leur phonétique leur permet d'être transcrits en arabe avec une prononciation identique à la prononciation occidentale, si bien qu'ils pourraient passer pour arabes.

### Les prénoms des petits-enfants d'immigrés (G3) du Maghreb

Parmi les petits-enfants d'immigrés (G3) du Maghreb, les 40 prénoms les plus portés sont là encore pour la plupart catégorisés comme « arabo-musulmans », comme « Karim », « Nassim » et « Farès » ou encore « Leila », « Lilia » et « Aïcha » (tableau 7.15). Parmi les 40 prénoms les plus fréquents chez les petits-enfants du groupe « majoritaire » (natifs sans parent ni grand-parent immigré), 15 se retrouvent dans le Top 40 des petits-enfants d'immigrés d'Europe du Sud (tableau 7.3), mais 1 seul (« Anaïs ») se retrouve dans le Top 40 des petits-enfants d'immigrés du Maghreb (tableau 7.15).

Parmi les 20 prénoms les plus portés par les petits-fils d'immigrés du Maghreb, seuls 3 ne sont pas catégorisés par la nomenclature de l'INED comme « arabo-musulmans » : « Kevin », « Adam » et un prénom noté « s » comme « secret » car trop rare pour pouvoir être mentionné sans risque de réidentification. En revanche, parmi les 20 prénoms les plus portés par les petites-filles d'immigrés du Maghreb, c'est une majorité de prénoms qui ne sont pas catégorisés comme « arabo-musulmans » : « Sarah », <sup>40</sup> « Sophia », « Lina », « Sabrina », « Inès », « Sofia », « Célia », « Myriam », « Anaïs », deux prénoms trop rares pour pouvoir être mentionnés sans risque de réidentification, « Melissa » et « Maeva ». Certes, ces prénoms féminins — dont plusieurs ont été repérés par la littérature qualitative comme attribués par des couples « mixtes » — peuvent, pour certains, être considérés comme arabes ou à tout le moins avoir une racine acceptable en arabe, ou être couramment attribués au Maghreb. Toutefois, l'impression générale qui se dégage des prénoms les plus portés par les petites-filles d'immigrés du Maghreb est assez largement différente de celle qui se dégage des prénoms des petits-fils. Chez les descendants d'immigrés du Maghreb en France, comme plus généralement en Occident et dans le monde entier (Annexe 5A, p. 147), les filles tendent manifestement à se voir attribuer des prénoms moins traditionnels (dans leur groupe d'origine) que les garçons.

Parmi les prénoms les plus portés par les petits-enfants d'immigrés du Maghreb (tableau 7.15), se trouvent :

- ✓ des prénoms coraniques d'origine arabe, comme « Mohamed » ; des prénoms coraniques d'origine hébraïque, comme « Ilyès », « Adam » et « Younès » ; et d'autres prénoms religieux d'origine arabe, comme « Mehdi » (guidé), « Yacine » (du nom de la 36<sup>e</sup> sourate du Coran) et « Driss » (du nom d'un prophète traditionnellement associé à Énoch) ;
- ✓ des prénoms arabes de proches ou successeurs du prophète Mohamed, comme « Marwan » (califes omeyyades), « Sofian » (chef de la tribu des Quraysh) et « Aïcha » (3<sup>e</sup> épouse du prophète) ;
- ✓ des prénoms profanes d'origine arabe comme « Karim » (généreux), « Nassim » (brise, prénom plus attribué au Proche-Orient qu'au Maghreb), « Farès » (chevalier), « Mourad » (vœu), « Ryad » (jardins), « Lina » (douce), « Leila » (nuit) et « Farah » (joie) ; d'autres prénoms profanes, d'origine persane comme « Yasmine » et « Jasmine » (jasmin), ou qui font référence à un personnage berbère comme « Kaina » (du nom d'une reine berbère du 7<sup>e</sup> siècle qui résista contre l'expansion arabe) ;
- ✓ des prénoms « passe-partout », <sup>41</sup> c'est-à-dire des prénoms d'origines multiples ou ambiguës (arabe mais pas seulement), comme « Rayan » et « Rayane »

<sup>40</sup> Dans leur article, Coulmont et Simon (2019) notent que chez les petits-enfants d'immigrés du Maghreb « on remarque, comme pour les petits-enfants des Européens du Sud, le choix croissant de prénoms appartenant au registre des prénoms maintenant considérés comme communs (Inès, Sarah) » (Coulmont et Simon 2019, p. 3). De fait, en métropole en 2008, « Inès » et « Sarah » sont portés à peu près aussi souvent par des membres du groupe « majoritaire » que par des personnes originaires du Maghreb, mais par rapport à la moyenne des prénoms ils sont beaucoup moins portés par le groupe « majoritaire » et beaucoup plus par les personnes originaires du Maghreb et les personnes d'origine musulmane.

<sup>41</sup> Conformément à la littérature qualitative sur les prénoms attribués par les couples « mixtes » (Annexe 7, p. 327),



(abondamment arrosé), « Sarah » et « Myriam » (prénoms d'origine hébraïque qui peuvent être considérés comme arabes mais aussi comme français), « Inès » (prononciation dialectale du prénom arabe « Īnās » qui signifie « amabilité », mais aussi forme française d'un prénom espagnol d'origine grecque), « Celia » (« consolée » en arabe, mais aussi forme féminine anglaise et espagnole d'un nom de famille latin qui signifie « ciel ») et « Lilia » (« nocturne » en arabe, mais aussi forme italienne et espagnole d'un prénom d'origine latine qui signifie « lys ») ; et des prénoms d'origine non arabe mais qui pourraient passer pour coraniques comme « Yanis » (prénom grec d'origine hébraïque, qui pourrait désigner le personnage coranique de Jean le Baptiste), ou qui tout au moins peuvent avoir une racine arabe crédible, comme « Sabrina » (forme française d'un prénom latin d'origine galloise, mais dans laquelle une oreille arabe pourrait reconnaître « sabr » (patience)) ;

- ✓ des prénoms non attribués au Maghreb et d'origine celtique comme « Kevin », grecque comme « Sophia » et « Sofia »<sup>42</sup> et « Melissa », ou tahitienne comme « Maeva » ;
- ✓ peut-être quelques prénoms « français » féminins non ambigus comme « Anaïs », mais *aucun prénom « français » masculin non ambigu* comme « Nicolas » : c'est là précisément le type de prénom qui brille par son absence du Top 20 des prénoms des petits-fils d'immigrés du Maghreb, alors que les descendants d'immigrés d'Europe du Sud (tableau 7.3) et d'Asie (Annexe 6B, p. 285) portent largement des prénoms « français » et des prénoms portés par les descendants de la population majoritaire (tableau 7.2).

Au total, quand on passe des immigrés (G1) du Maghreb à leurs enfants (G2) et à leurs petits-enfants (G3), certains prénoms très répandus au Maghreb figurent continuellement dans le Top 20 des prénoms les plus portés : « Mohamed », « Karim », « Aïcha » et « Meriem » / « Myriam ». « Mohamed », porté par 10 % des immigrés, n'est plus porté que par 4 % des fils d'immigrés et 1 % des petits-fils d'immigrés. Plus généralement, les prénoms coraniques ainsi que les prénoms théophores, qui sonnent typiquement arabes, ainsi que les prénoms de proches ou successeurs du prophète Mohamed, deviennent relativement moins nombreux, pour être en partie remplacés par des prénoms « passe-partout », voire parfois par quelques prénoms féminins « français » non ambigus. Mais **le Top 20 des prénoms les plus portés par les petits-fils d'immigrés du Maghreb ne comporte aucun prénom « français » non ambigu et aucun prénom porté par les fils d'enquête du groupe « majoritaire »**. Dans ce contexte, il n'est peut-être pas particulièrement surprenant que n'y figure pas un prénom masculin « français » non ambigu et typiquement chrétien comme « Nicolas » — prénom qui se trouve de surcroît être celui d'un saint catholique associé à la Reconquista.

---

j'appelle « passe-partout » des prénoms qui « ne sonnent pas trop arabe » tout en pouvant passer pour l'être : il peut s'agir de prénoms arabes homophones de prénoms anglophones (« Rayan » et « Rayane » sont homophones de « Ryan », « Samy » de « Sammy », etc.), ou de prénoms d'origine « mixte » ou ambiguë (arabe mais pas seulement), ou encore de prénoms non arabes (« anglophones » ou « latins » voire « français ») mais qui peuvent avoir une racine arabe crédible. De fait, la situation linguistique du Maghreb permet toutes sortes d'hybridations, puisque coexistent deux variétés d'arabe, classique et dialectale (dont les systèmes phonologiques ne coïncident pas entièrement), ainsi que les langues berbères et le français.

<sup>42</sup> On ne saurait exclure que les prénoms des petits-enfants d'immigrés du Maghreb en France suivent des modes islamiques mondiales, comme pourrait en témoigner le fait que dans un pays aussi éloigné de la France que la Malaisie, les Malais (Chinois et Indiens exclus) attribuent en 2016 à leurs enfants des prénoms qui ressemblent fort à ceux qui sont portés par les petits-enfants d'immigrés du Maghreb en France en 2008, qu'il s'agisse de prénoms arabes (« Mohamed », « Ahmad », « Adam » et « Rayyan » pour les garçons et « Nor », « Aishah », « Maryam » et « Nadia » pour les filles) ou de prénoms non arabes (« Sophia ») (Robustova 2016, p. 207-208).

### Encadré 7.16. Des descendants d’immigrés du Maghreb rarement prénommés « Nicolas »... ou « Pierre »

Le 20 avril 2021, une mère (voilée) du quartier de La Mosson, à Montpellier, interpellait le Président Macron comme suit :

« Monsieur le Président, j’ai mon fils qui a 8 ans, il m’a demandé si le prénom de "Pierre" existait vraiment ou si c’est que dans les livres, tellement qu’il y a un manque de mixité dans le quartier. C’est vraiment grave, quoi. Je lui ai dit non, mon fils, "Pierre" ça existe vraiment, il y a des gens qui s’appellent "Pierre". Moi je fais sortir mes enfants, mais c’est vrai que cette question m’a choquée » ([vidéo](#)).

Le fait qu’un enfant puisse se demander si le prénom « Pierre » est encore porté en France est bien entendu révélateur de niveaux élevés de concentration spatiale d’immigrés et de descendants d’immigrés. En l’occurrence, à Montpellier, la part des 0-18 ans qui vivent avec au moins un parent immigré extra-européen est passée de 4 % en 1968 à 38 % en 2017, et dans certains quartiers de la ville jusqu’à 71 % des 0-18 ans vivent avec au moins un parent immigré extra-européen, les *petits-enfants* d’immigrés n’étant pas inclus.<sup>43</sup> Mais le fait qu’un enfant puisse se demander si « Pierre » est encore porté dépend aussi du fait que, comme « Nicolas », ce prénom est précisément le type de prénom — masculin, et « français » non ambigu — que ne portent presque jamais les descendants d’immigrés du Maghreb en France.

#### ❖ Résumé

La plupart des « Nicolas » du tableau de Coulmont et Simon (2019, p. 3) ne sont pas des petits-fils d’immigrés du Maghreb, mais des petits-fils de pieds-noirs. Leur entourage porte des prénoms largement « français » ou européens (87 %) et ils ne sont pas, typiquement, les enfants de « Mohamed » et « Sarah », mais plutôt ceux de « Christine » et « Philippe ». Plus généralement, 1 % des immigrés (G1) du Maghreb de l’enquête TeO seraient des pieds-noirs et 10 % des enfants d’immigrés (G2) du Maghreb seraient des enfants de pieds-noirs, mais surtout, 26 % des petits-enfants d’immigrés (G3) du Maghreb seraient des petits-enfants de pieds-noirs, ce qui explique pourquoi le tableau publié comprend un prénom comme « Nicolas », mais seulement chez les petits-enfants (G3). Comme les rapatriés d’Afrique du Nord sont par définition nés avant 1962 alors que, dans l’enquête TeO, les immigrés du Maghreb peuvent être nés jusqu’en 1990, les rapatriés d’Afrique du Nord sont en moyenne bien plus âgés que les immigrés du Maghreb, si bien qu’en moyenne ils ont eu plus de temps d’avoir leurs petits-enfants.

Une fois les immigrés du Maghreb et leurs descendants dissociés des pieds-noirs et de leurs descendants, il apparaît que sur l’échantillon TeO retenu pondéré (représentatif des personnes de 18-50 / 18-60 ans en ménage ordinaire et de leurs enfants de 0-17 ans, habitant en France métropolitaine en 2008), les prénoms les plus fréquents chez les petits-fils d’immigrés du Maghreb sont non pas « Yanis » et « Nicolas », mais « Karim » et « Nassim ». Si les petites-filles d’immigrés du Maghreb portent des prénoms relativement diversifiés, les petits-fils portent pour la plupart des prénoms « arabo-musulmans ». Le Top 20 des prénoms les plus portés par les petits-fils d’immigrés du Maghreb ne comporte *aucun* prénom « français » non ambigu et *aucun* prénom porté par les fils d’enquête du groupe « majoritaire » — une situation très différente de celle des descendants d’immigrés d’Europe du Sud ou d’Asie. Pas étonnant, donc, que le prénom « Nicolas » ne fasse pas partie du Top 20 des prénoms des petits-fils

<sup>43</sup> Voir [ici](#) l’outil de datavisualisation de France Stratégie.



d'immigrés du Maghreb, ni a fortiori du Top 3, d'autant qu'il se trouve être celui d'un saint catholique qui est associé à la Reconquista. Le prénom « Nicolas » est en revanche porté par des descendants de pieds-noirs, comme le réalisateur Nicolas Bedos ou le journaliste Nicolas Demorand.

Le 10 avril 2019, jour de la parution de l'article de Coulmont et Simon (2019), dans [l'émission « L'heure des pros » sur CNEWS](#), quand la journaliste Élisabeth Lévy fait part de son sentiment selon lequel le rang du prénom « Nicolas » dans les prénoms des petits-fils d'immigrés du Maghreb lui semble « assez contre-intuitif », Patrick Simon, interrogé en direct, lui répond :

« C'est à ça que ça sert ces études, c'est contre-intuitif, c'est vrai, peut-être, mais en l'occurrence on n'a pas décidé de leur donner ce prénom-là [Nicolas], c'est ce qu'on a constaté en prenant les prénoms des personnes dont les grands-parents étaient originaires du Maghreb » (vidéo [ici](#), anciennement [là](#) de 8:36 à 9:30).

Ce ne sont certes pas les auteurs de l'article qui ont « décidé de leur donner ce prénom-là », mais de la part de deux chercheurs — l'un Professeur des universités à l'Université Paris 8 Vincennes-Saint-Denis, aujourd'hui à l'ENS Paris-Saclay et spécialiste des prénoms, l'autre Directeur de recherche à l'INED, directeur de département à l'Institut Convergences Migrations, chercheur associé à Sciences Po Paris et spécialiste des immigrés —, on aurait peut-être pu s'attendre à ce qu'ils vérifient des résultats qu'ils admettent « contre-intuitifs ». Quant à l'INED, organisme public de recherche qui est censé faire référence en France sur les questions de population et d'immigration, n'aurait-il pas pu lui aussi (faire) vérifier de tels résultats avant de les diffuser au grand public ?

## 8. Essai de reproduction du second résultat spectaculaire : le « 23 % »

On sait désormais que le premier résultat spectaculaire de Coulmont et Simon (2019, p. 3) est faux : la plupart des « Nicolas » de leur tableau ne sont pas des petits-fils d'immigrés du Maghreb mais, à l'évidence, des petits-fils de pieds-noirs (Chapitre 7, p. 26). Les auteurs indiquent plus généralement qu'en France métropolitaine en 2008, seulement 23 % des petits-enfants d'immigrés du Maghreb portent un prénom « arabo-musulman » (Coulmont et Simon 2019, figure 2, p. 3).<sup>44</sup> Est-ce vrai ?

### ❖ La reproduction « au plus près » de la figure 2, page 3

Pour y voir plus clair dans les résultats publiés par Coulmont et Simon (2019), je reproduis leur figure 2 sur l'échantillon retenu, selon 2 variantes de pondération (avec ou sans) et en tenant compte *ou non* du fait que certains (descendants d') immigrés du Maghreb sont vraisemblablement des (descendants de) pieds-noirs, soit au total 4 combinaisons d'hypothèses envisageables (tableau 8.1a de l'Annexe 8, p. 355).<sup>45</sup> Par commodité, je reproduis cette figure 2 non pas sous forme de graphique mais — dans un premier temps — sous forme de tableau, et je ne reproduis pas les résultats spécifiques aux « G2 mixtes », qui sur la figure 2 pourraient avoir été comptés deux fois.<sup>46</sup> De ces résultats (tableau 8.1a de l'Annexe 8, p. 355), je tire plusieurs enseignements :

- tout d'abord, cette reproduction « au plus près » ne permet pas de reproduire à l'identique les résultats publiés par Coulmont et Simon (2019, p. 3) : comme leur article n'indique pas le champ sur lequel ils ont calculé leurs résultats, et comme les programmes informatiques qu'ils m'ont finalement communiqués sont incomplets (Annexe 4B, p. 124), je continue d'ignorer le champ sur lequel leur figure 2 est calculée ;
- quoi qu'il en soit, les résultats varient assez peu d'une pondération à l'autre, et aucun d'entre eux ne s'approche des « 23 % » publiés :
  - *même en conservant les petits-enfants de pieds-noirs parmi les petits-enfants d'immigrés du Maghreb*, ce sont 37 % des petits-enfants d'immigrés du Maghreb qui, sur l'échantillon retenu pondéré et d'après la nomenclature de l'INED, portent un prénom spécifique (tableau 8.1a de l'Annexe 8, p. 355)<sup>47</sup> ; par conséquent, si dans le Chapitre 7 (p. 26) la reproduction « au plus près » du tableau page 3 a permis de retrouver le résultat le plus spectaculaire (le « Nicolas » des tableaux 7.4 et 7.5), dans ce Chapitre 8 la reproduction « au plus près » de la figure 2 page 3 ne permet *pas* de retrouver son résultat le plus spectaculaire. Alors, comment

<sup>44</sup> Ce résultat, visible sur la Figure 2 page 3 de l'article, est confirmé dans le texte : « en moyenne, 23 % des petits-enfants d'immigrés maghrébins portent un prénom arabo-musulman » (Coulmont et Simon 2019, p. 4).

<sup>45</sup> J'y ajoute des estimations (avec ou sans pondération) sur 2 autres échantillons (échantillon maximal et échantillon composé des enquêtés et de tous leurs enfants vivants) (tableaux 8.1b à 8.1c de l'Annexe 8, p. 355), que je commente en note de bas de page.

<sup>46</sup> Je considère comme « spécifiques » les prénoms suivants : pour le groupe « majoritaire » et les personnes originaires des DOM, ce sont les prénoms que la nomenclature de l'INED catégorise comme « français » ; pour les personnes originaires d'Europe du Sud, les prénoms « latins » ; pour les personnes originaires d'autre Europe, les prénoms « internationaux ou anglophones » et d'« autre Europe » ; pour les personnes originaires du Maghreb et de Turquie, les prénoms « arabo-musulmans » ; pour les personnes originaires d'autre Afrique, les prénoms « africains » ; et pour les personnes originaires d'Asie, les prénoms « asiatiques ». Je ne peux attribuer de catégorie de prénoms « spécifiques » ni aux personnes originaires d'« ailleurs » ni à celles qui n'ont pas répondu aux questions relatives à leurs origines (ou auxquelles ces questions n'ont pas été posées).

<sup>47</sup> En conservant les petits-enfants de pieds-noirs parmi les petits-enfants d'immigrés du Maghreb sur les autres échantillons (échantillon maximal pondéré ou non, échantillon composé des enquêtés et de tous leurs enfants vivants pondéré ou non, et échantillon retenu non pondéré), ce sont de 34 % à 39 % des petits-enfants d'immigrés du Maghreb qui portent un prénom spécifique (tableaux 8.1a à 8.1c de l'Annexe 8, p. 355).

les auteurs ont-ils fait pour estimer qu'en métropole en 2008, seuls 23 % des petits-enfants d'immigrés du Maghreb portaient un prénom « arabo-musulman » ? Je ne peux pas le savoir — et ce sera à eux de le préciser — mais **pour parvenir à un tel résultat les auteurs ne se sont sans doute pas contentés d'inclure dans leurs calculs de nombreux petits-enfants de pieds-noirs ;**

- et une fois les petits-enfants d'immigrés du Maghreb dissociés des petits-enfants de pieds-noirs, 49 % des petits-enfants d'immigrés du Maghreb portent un prénom que la nomenclature de l'INED catégorise comme « arabo-musulman » ; précisément, sur l'échantillon retenu pondéré (représentatif des personnes de 18-50 / 18-60 ans en ménage ordinaire et de leurs enfants de 0-17 ans habitant en France métropolitaine en 2008), **les petits-enfants d'immigrés du Maghreb sont ceux qui portent le plus souvent un prénom spécifique (49 %), loin devant les petits-enfants d'immigrés d'Europe du Sud (8 %), d'autre Europe (19 %), d'autre Afrique (9 %) ou d'Asie (5 %) (tableau 8.1a de l'Annexe 8, p. 355).**<sup>48</sup>

#### ❖ Analyses complémentaires : proportion de prénoms spécifiques par origine et par génération

Pour pouvoir se faire une image globale de ces résultats, il est utile de les visualiser sous forme de graphiques. Sur l'échantillon retenu pondéré (représentatif des personnes de 18-50 / 18-60 ans en ménage ordinaire et de leurs enfants de 0-17 ans habitant en France métropolitaine en 2008), si l'on compare la proportion de prénoms spécifiques par origine et par « génération » chez les personnes d'origine étrangère<sup>49</sup> (figure 8.2a), apparaissent deux groupes d'origine distincts :

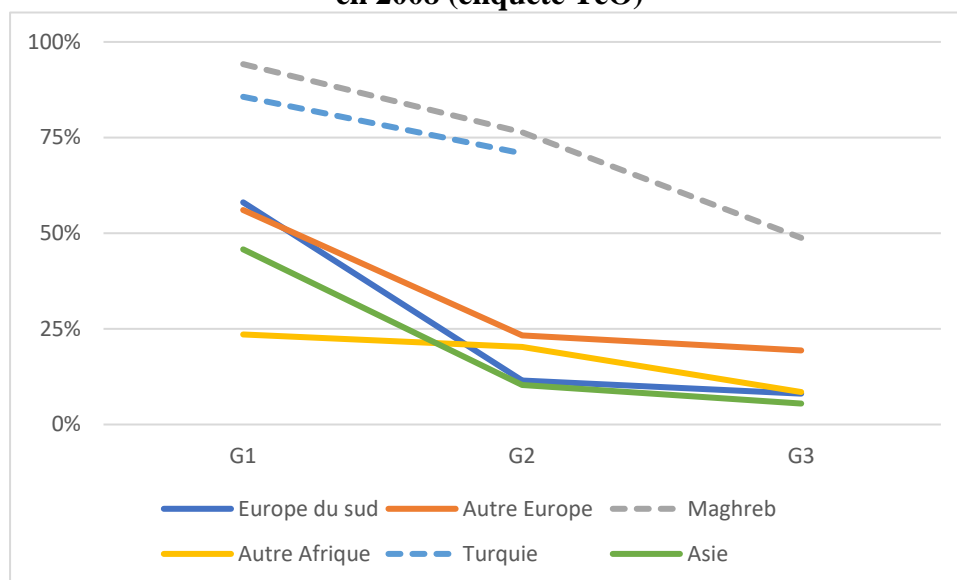
- d'une part, les personnes originaires du Maghreb (et de Turquie), chez qui plus de 75 % des immigrés (G1) et des enfants d'immigrés (G2) portent des prénoms spécifiques, et chez qui près de 50 % des petits-enfants d'immigrés (G3) continuent de porter des prénoms « arabo-musulmans » ;
- d'autre part, les autres personnes d'origine étrangère (originaires d'Europe du Sud, d'autre Europe, d'autre Afrique et d'Asie), chez qui seulement 25 % à 60 % des immigrés (G1) portent des prénoms spécifiques, et chez qui environ 25 % des enfants d'immigrés (G2) et moins de 25 % des petits-enfants d'immigrés (G3) continuent de porter des prénoms spécifiques.

<sup>48</sup> Toujours sur l'échantillon retenu pondéré (tableau 8.1a de l'Annexe 8, p. 355), les petits-enfants d'immigrés de Turquie sont 39 % à porter un prénom spécifique (« arabo-musulman »), mais cette estimation repose sur un effectif trop faible pour pouvoir être considérée comme fiable. Sur l'échantillon pondéré composé des enquêtés et de tous leurs enfants vivants, qui comprend de nombreux doublons mais qui produisait les résultats du tableau page 3 les plus proches de ceux publiés par les auteurs (Chapitre 7, p. 26), les petits-enfants d'immigrés du Maghreb (non dissociés des petits-enfants de pieds-noirs) sont là aussi les petits-enfants d'origine étrangère qui portent le plus souvent un prénom spécifique (38 %), avec les petits-enfants d'immigrés de Turquie (39 %) et loin devant les petits-enfants d'immigrés d'Europe du Sud (7 %), d'autre Europe (21 %), d'autre Afrique (19 %) ou d'Asie (3 %) (tableau 8.1c de l'Annexe 8, p. 355).

<sup>49</sup> Pour information, sur l'échantillon retenu pondéré (tableau 8.1a de l'Annexe 8, p. 355), 90 % des enquêtés du groupe « majoritaire » et 67 % de leurs enfants du groupe « majoritaire » portent un prénom spécifique, c'est-à-dire « français » ; et 79 % des natifs (G1) des DOM, 64 % de leurs enfants (G2) et 46 % de leurs petits-enfants (G3) portent un prénom spécifique, c'est-à-dire là encore « français ». Si le groupe « majoritaire » et les personnes originaires des DOM abandonnent partiellement les prénoms « français », c'est largement au profit de prénoms « internationaux ou anglophones » (mais aussi de certains prénoms « latins », d'« autre Europe » ou « inclassables »).

Les descendants d'immigrés du Maghreb (et de Turquie) se distinguent donc nettement des autres descendants d'immigrés en ce que, au fil des « générations », ils conservent leurs prénoms spécifiques beaucoup plus massivement (figure 8.2a). De même, en France en 1992, les enfants (G2) et les petits-enfants d'immigrés (G3) d'Algérie étaient plus nombreux à porter un prénom « du pays d'origine » (84 % et 35 %) que les enfants (G2) et les petits-enfants d'immigrés (G3) d'Espagne (25 % et 7 %) (Valetas et Bringé 2009).

**Figure 8.2a. Proportion de prénoms spécifiques par origine et « génération » en France en 2008 (enquête TeO)**



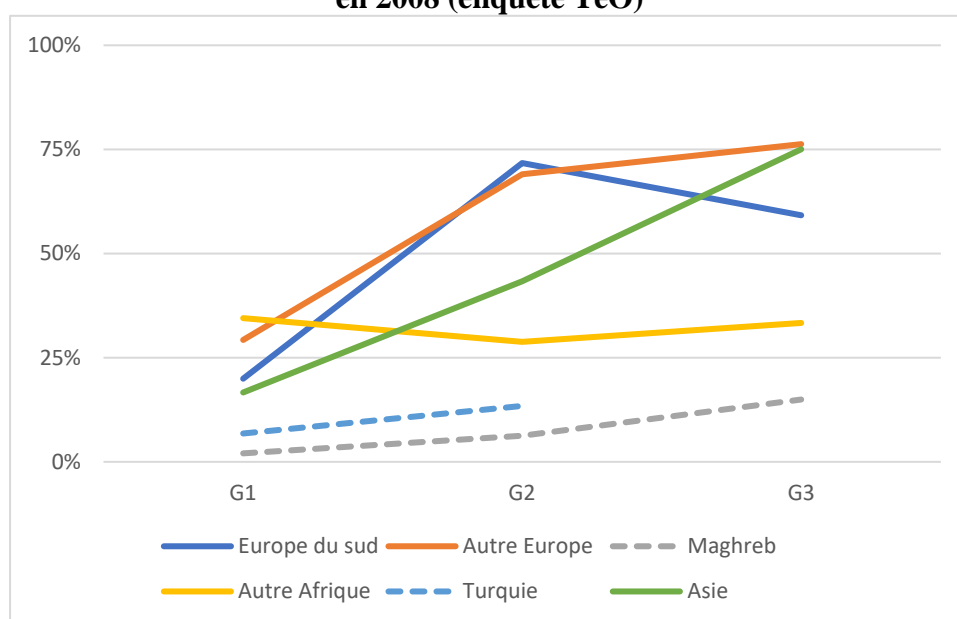
Lecture : parmi les personnes originaires du Maghreb, la part de prénoms spécifiques (« arabo-musulmans ») passe de 94 % (G1) à 76 % (G2) puis 49 % (G3), d'après la nomenclature de l'INED.

Champ : personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008, données pondérées, N = 42 208. Les G1, G2 et G3 Maghreb qui en réalité sont probablement (descendants de) rapatriés sont exclus des analyses.

Source : TeO (voir Annexe 5B, p. 183).

Parallèlement, les descendants d'immigrés du Maghreb (et de Turquie) sont ceux qui adoptent des prénoms « français » le moins souvent (figure 8.2b). De fait, seuls 15 % des petits-enfants d'immigrés du Maghreb portent un prénom que la nomenclature de l'INED catégorise comme « français », loin derrière les petits-enfants d'immigrés d'Europe du Sud (59 %), d'autre Europe (76 %), d'autre Afrique (33 %) et d'Asie (75 %) (figure 8.2b). Les descendants d'immigrés du Maghreb se distinguent donc des autres descendants d'immigrés en ce que, au fil des « générations », ils se voient attribuer des prénoms « français » beaucoup moins massivement (figure 8.2b). De même, en France en 1992, les enfants (G2) et les petits-enfants d'immigrés (G3) d'Algérie étaient moins nombreux à porter un prénom « français » (3 % et 22 %) que les enfants (G2) et les petits-enfants d'immigrés (G3) d'Espagne (61 % et 56 %) (Valetas et Bringé 2009).

**Figure 8.2b. Proportion de prénoms « français » par origine et « génération » en France en 2008 (enquête TeO)**



Lecture : parmi les personnes originaires du Maghreb, la part de prénoms « français » passe de 2 % (G1) à 6 % (G2) puis 15 % (G3), d'après la nomenclature de l'INED.

Champ : personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008, données pondérées, N = 42 208. Les G1, G2 et G3 Maghreb qui en réalité sont probablement (descendants de) rapatriés sont exclus des analyses.

Source : TeO (voir Annexe 5B, p. 183).

Coulmont et Simon (2019) indiquent à plusieurs reprises que les prénoms portés par les descendants d'immigrés du Maghreb convergent avec ceux de la population « majoritaire ». Dans le texte de leur article, on peut lire :

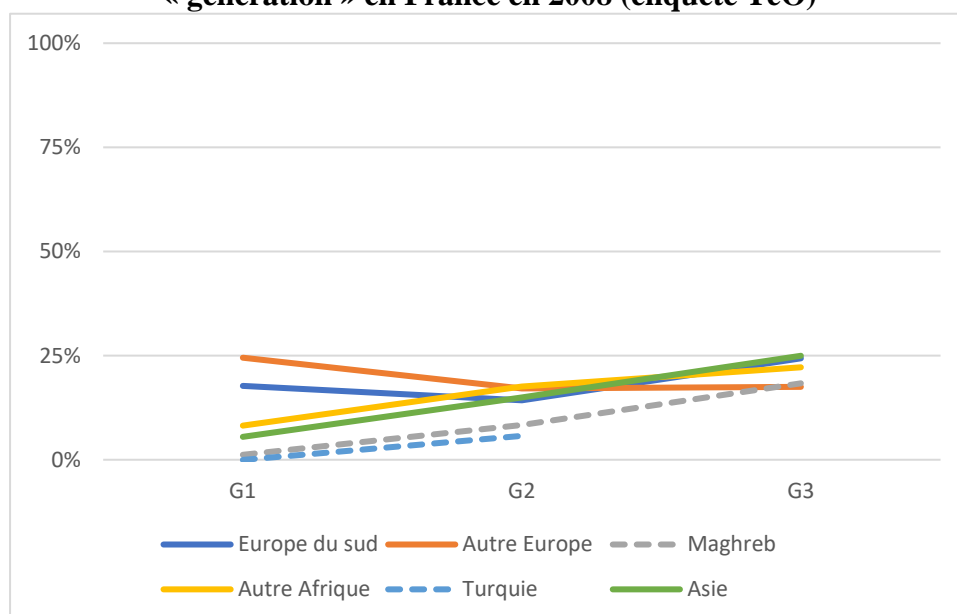
« Les prénoms que reçoivent les petits-enfants [d'immigrés du Maghreb] sont, en 2008, proches de ceux que la population majoritaire donne à ses enfants. [...] La trajectoire suivie par les originaires du Maghreb mène au même point d'arrivée que celle suivie par les Européens du Sud, mais de manière différée » (Coulmont et Simon 2019, p. 3).

Ces affirmations sont clairement démenties par les figures 8.2a et 8.2b. D'une part, il est difficile de soutenir que les prénoms des petits-enfants d'immigrés du Maghreb sont proches de ceux que la population majoritaire donne à ses enfants : les premiers sont plus souvent « arabo-musulmans » (49 %) que « français » (15 %), tandis que les seconds sont beaucoup plus souvent « français » (67 %) que « arabo-musulmans » (2 %). D'autre part, il est difficile de soutenir que la trajectoire suivie par les originaires du Maghreb mène au même point d'arrivée que celle suivie par les Européens du Sud : parmi les petits-enfants d'immigrés, ceux d'Europe du Sud ne sont que 8 % à porter des prénoms spécifiques, alors que ceux du Maghreb sont 49 % ; et parallèlement, les petits-enfants d'immigrés d'Europe du Sud sont 59 % à porter un prénom « français », alors que ceux du Maghreb ne sont que 15 %. Comme on l'a vu, parmi les 40 prénoms les plus fréquents chez les enfants « majoritaires » des enquêtés (natifs sans parent ni grand-parent immigré ou des DOM), 1 seul se retrouve dans le Top 40 des prénoms des petits-enfants d'immigrés du Maghreb (tableau 7.15), contre 15 dans le Top 40 des prénoms des petits-enfants d'immigrés d'Europe du Sud (tableau 7.3).

Puis la conclusion et le résumé de l'article reparlent, chacun, d'une « convergence entre population majoritaire et descendants d'immigrés », qui « ne se fait pas autour de prénoms typiquement "français", mais de prénoms internationaux auxquels tous et toutes peuvent

s'identifier » (Coulmont et Simon 2019, p. 4). En réalité, en métropole en 2008, les petits-enfants d'immigrés du Maghreb sont plus nombreux à porter un prénom spécifique (49 %) (figure 8.2a) qu'un prénom « international ou anglophone » (18 %) (figures 8.2c). Par contraste, les petits-enfants d'immigrés d'Europe du Sud — comme ceux d'autre Europe ou même d'Asie — sont bien plus nombreux à porter un prénom « français » (de 59 % à 76 %) (figure 8.2b) qu'un prénom « international ou anglophone » (18 % à 25 %) (figure 8.2c). Difficile, donc, de parler d'une « convergence entre population majoritaire et descendants d'immigrés » qui se ferait autour de prénoms internationaux. La convergence avec les prénoms de la population « majoritaire » n'est réelle que pour les descendants d'immigrés d'Europe du Sud, d'autre Europe ou d'Asie — mais pas du Maghreb.

**Figure 8.2c. Proportion de prénoms « internationaux ou anglophones » par origine et « génération » en France en 2008 (enquête TeO)**



Lecture : parmi les personnes originaires du Maghreb, la part de prénoms « internationaux ou anglophones » passe de 1 % (G1) à 8 % (G2) puis 18 % (G3), d'après la nomenclature de l'INED.

Champ : personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008, données pondérées, N = 42 208. Les G1, G2 et G3 Maghreb qui en réalité sont probablement (descendants de) rapatriés sont exclus des analyses.

Source : TeO (voir Annexe 5B, p. 183).

Au final, les descendants d'immigrés du Maghreb se distinguent des autres descendants d'immigrés en ce que, au fil des « générations », ils sont ceux qui, de loin, abandonnent le moins souvent leurs prénoms spécifiques (figure 8.2a) et adoptent le moins souvent des prénoms « français » (figure 8.2b). Cela n'empêche pas que les petits-enfants d'immigrés du Maghreb, comme les autres petits-enfants d'immigrés et comme les petits-enfants de la population « majoritaire », portent plus de prénoms « internationaux ou anglophones » que leurs aïeux (figure 8.2c). Mais contrairement à ce qu'indiquent Coulmont et Simon (2019), les descendants d'immigrés du Maghreb sont les descendants d'immigrés dont les prénoms convergent *le moins* avec ceux du groupe « majoritaire ».

#### ❖ Des résultats spécifiques à la France ?

Chez les Pakistanais musulmans immigrés à Glasgow (Royaume-Uni) et leurs descendants dans les années 2000, presque tous les prénoms portés sont spécifiques à cette population et d'origine arabe (Bramwell 2011, p. 45-46 ; voir aussi encadré 6.1). Et parmi les immigrés en Allemagne des années 1980 aux années 2000, les immigrés de Turquie sont ceux



qui attribuent le plus de prénoms spécifiques à leurs enfants : 90 % des immigrés de Turquie attribuent à leurs enfants des prénoms spécifiques, contre 43 % des immigrés d'ex-Yougoslavie et 35 % des immigrés d'Europe du Sud (Italie, Espagne et Portugal) (Gerhards et Hans 2009, p. 1116). Précisément, en Allemagne, des années 1980 aux années 2000, les immigrés donnent plus souvent à leurs enfants des prénoms allemands s'ils sont en mariage mixte avec un Allemand, s'ils ont plus de contacts avec des Allemands,<sup>50</sup> si leur niveau de diplôme est plus élevé, si leur enfant naît allemand et si leur enfant est une fille plutôt qu'un garçon — mais « toutes choses égales par ailleurs », les immigrés de Turquie sont bel et bien ceux qui attribuent le moins souvent des prénoms allemands, derrière les immigrés d'ex-Yougoslavie et a fortiori d'Europe du Sud (Gerhards et Hans 2009, p. 1116-1124 ; voir aussi Tuppatt et Gerhards 2021). Par conséquent, si en France les immigrés du Maghreb (et de Turquie) et leurs descendants continuent — beaucoup plus largement que les autres immigrés — d'attribuer des prénoms spécifiques à leurs origines plutôt que des prénoms « français » (ou « internationaux ou anglophones »), il ne s'agit pas là d'une exception française : le même phénomène s'observe chez les immigrés d'origine musulmane au Royaume-Uni et en Allemagne.

### ❖ Résumé

Même en conservant les petits-enfants de pieds-noirs parmi les petits-enfants d'immigrés du Maghreb, ce ne sont pas 23 % mais 37 % des petits-enfants d'immigrés (G3) du Maghreb qui porteraient un prénom « arabo-musulman ». Et une fois les petits-enfants d'immigrés du Maghreb dissociés des petits-enfants de pieds-noirs, il apparaît que sur l'échantillon TeO retenu pondéré (représentatif des personnes de 18-50 / 18-60 ans en ménage ordinaire et de leurs enfants de 0-17 ans, habitant en France métropolitaine en 2008), 49 % des petits-enfants d'immigrés du Maghreb portent un prénom que la nomenclature de l'INED catégorise comme « arabo-musulman » — loin devant les proportions de prénoms « spécifiques » que l'on peut observer chez les petits-enfants d'immigrés d'Europe du Sud (8 %), d'autre Europe (19 %), d'autre Afrique (9 %) ou d'Asie (5 %). Les petits-enfants d'immigrés du Maghreb sont non seulement ceux qui portent le plus souvent un prénom spécifique à leurs origines, mais aussi ceux qui portent le moins souvent un prénom « français » (15 %), loin derrière les petits-enfants d'immigrés d'Europe du Sud (59 %). La sur-rétention des prénoms d'origine s'observe d'ailleurs chez d'autres populations d'origine musulmane en Europe, comme chez les descendants d'immigrés du Pakistan au Royaume-Uni et chez les descendants d'immigrés de Turquie en Allemagne.

---

<sup>50</sup> Plus une minorité est géographiquement concentrée, plus elle tend à attribuer des prénoms spécifiques à ses origines : c'est ce qu'on observe aussi chez la minorité russophone de Finlande dans les années 2000 et 2010 (Eskola et Hämäläinen 2019, p. 211-213).

## 9. Essai de reproduction des autres résultats

On sait désormais que les deux résultats les plus spectaculaires de Coulmont et Simon (2019) sont faux (Chapitre 7, p. 26 et Chapitre 8, p. 46). Qu'en est-il des autres résultats ?

### ❖ Reproduction « au plus près » de la figure 1 (page 2)

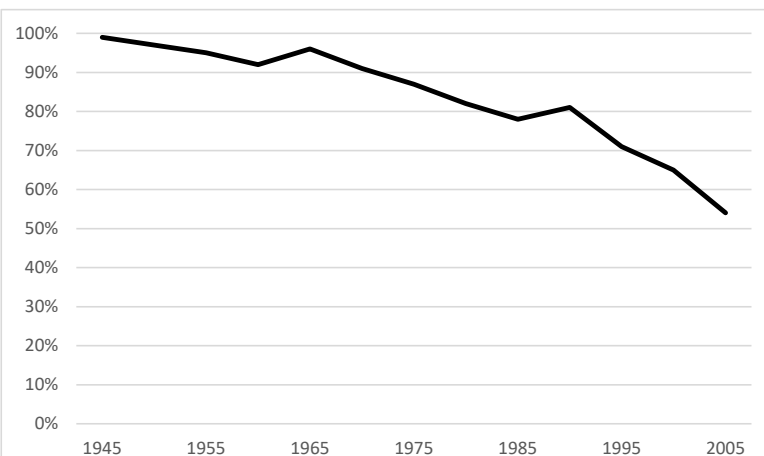
Les auteurs indiquent que de moins en moins de nouveau-nés de la population « majoritaire » recevraient un prénom « français », de près de 100 % à la fin des années 1940 à un peu plus de 50 % à la fin des années 2000 (Coulmont et Simon 2019, figure 1 p. 2). Est-ce vrai ?

Oui. Si l'on compare le résultat publié (figure 9.1a) à la même figure reproduite « au plus près » (figure 9.1b), les résultats obtenus sont similaires. Sur les individus de l'échantillon retenu pondéré (représentatif des personnes de 18-50 / 18-60 ans en ménage ordinaire et de leurs enfants de 0-17 ans habitant en France métropolitaine en 2008) qui sont nés en France de 1945-1949 à 2005-2009 et qui sont du groupe « majoritaire », la proportion de nouveau-nés qui reçoivent un prénom « français » est passée, d'après la nomenclature de l'INED, de 99 % en 1945-1949 à 54 % en 2005-2009. Il s'agit d'une des raisons majeures pour lesquelles, depuis les années 1950, les prénoms attribués en France sont de moins en moins souvent « français » (figures 6.10 et 6.11).

## Comparaison entre la figure 1 publiée page 2 et la même figure reproduite « au plus près »

### PUBLIÉ

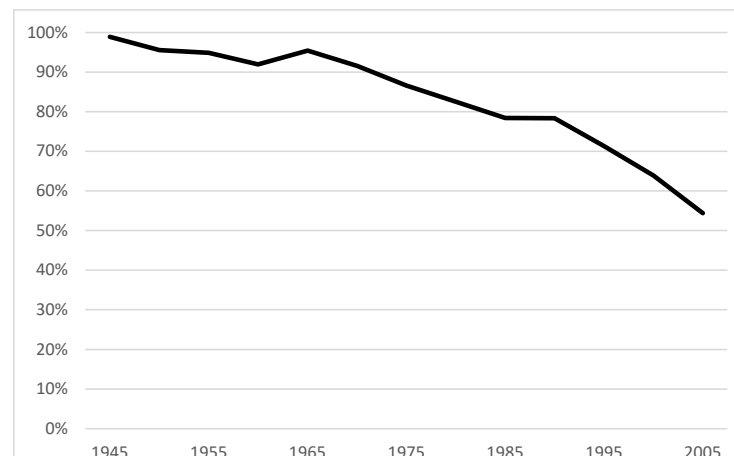
Figure 9.1a. Proportion de nouveau-nés recevant un prénom « français » par année de naissance - France, population majoritaire (%)



Champ enfants de la population « majoritaire », pondération non indiquée, N = non indiqué.

### REPRODUIT « AU PLUS PRÈS »

Figure 9.1b. Proportion de nouveau-nés recevant un prénom « français » par année de naissance - France, population majoritaire (%)



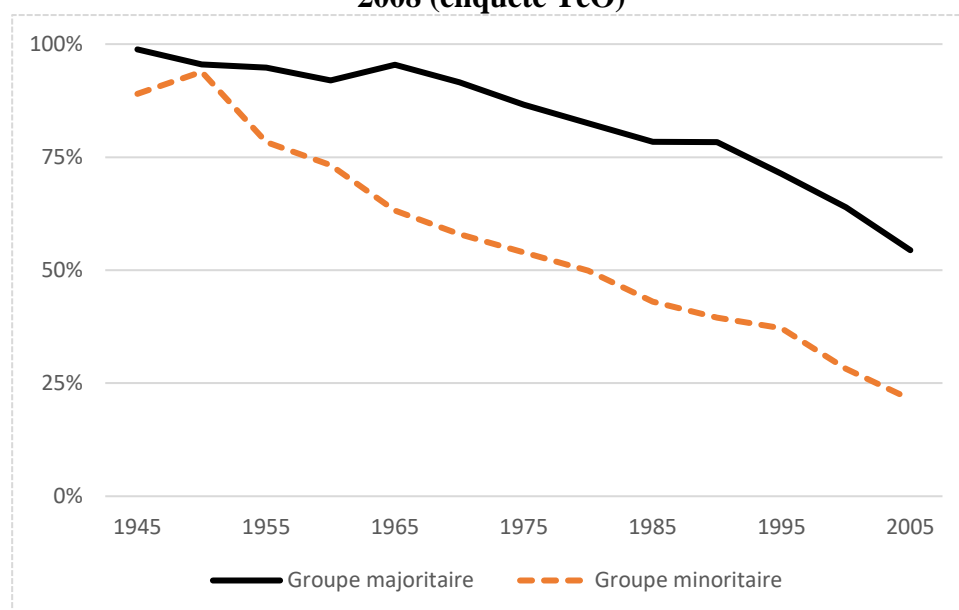
Champ : individus du champ retenu (personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008) qui sont nés en France de 1945-1949 à 2005-2009 et qui sont du groupe « majoritaire », données pondérées, N = 32 185.

### ❖ Analyses complémentaires autour de la figure 1 (page 2)

Pour un œil averti, la figure 1 de l'article (Coulmont et Simon 2019, p. 2) peut paraître légèrement surprenante : pourquoi présenter la proportion de nouveau-nés recevant un prénom « français » exclusivement pour la population « majoritaire » (c'est-à-dire sans parent ni grand-parent immigré ni DOMien), et non pas aussi pour la population « minoritaire » (composée des immigrés et DOMiens et de leurs descendants), de façon à présenter un graphique couvrant la totalité de la population ? Les nouveau-nés de la population « minoritaire » reçoivent-ils eux aussi de moins en moins de prénoms « français », ou de plus en plus ?

Si de la fin des années 1940 à la fin des années 2000 le groupe « majoritaire » se voit attribuer des prénoms de moins en moins souvent « français » (de 99 % à 54 %), c'est encore plus vrai pour le groupe « minoritaire » (de 89 % à 21 %) (figure 9.2). En effet, ce groupe est de moins en moins composé de personnes originaires d'Europe du Sud ou d'autre Europe (ou de personnes originaires des DOM), et de plus en plus composé de personnes originaires du Maghreb et d'autre Afrique (et de Turquie), dont on a vu que ce sont celles qui se voient attribuer le moins souvent des prénoms « français » (figure 8.2b). L'origine de plus en plus fortement maghrébine et africaine (et turque) des immigrés et de leurs descendants a ainsi pour effet de réduire la part du groupe « minoritaire » qui se voit attribuer un prénom « français ».

**Figure 9.2. Proportion de nouveau-nés recevant un prénom « français » par année de naissance, dans le groupe « majoritaire » et dans le groupe « minoritaire », en France en 2008 (enquête TeO)**



Lecture : d'après la nomenclature de l'INED, entre 1945-1949 et 2005-2009, la part de prénoms « français » passe de 99 % à 54 % dans le groupe « majoritaire » (sans parent ni grand-parent immigré ni DOMien), mais — ce que ne montre pas la figure 1 publiée page 2 — elle passe aussi de 89 % à 21 % dans le groupe « minoritaire » (immigrés et DOMiens, et natifs de métropole ayant au moins un parent ou grand-parent immigré ou DOMien).

Champ : individus du champ retenu (personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008) qui sont nés en France de 1945-1949 à 2005-2009, données pondérées, N = 32 185. Les G1, G2 et G3 Maghreb qui en réalité sont probablement (descendants de) rapatriés sont exclus des analyses.

Source : TeO (voir Annexe 5B, p. 183).

Ces observations permettent de clarifier pourquoi de moins en moins de prénoms attribués en France sont « français », de 93 % en 1900 et en 1950 à seulement 29 % en 2019 (figure 6.11a) :

- non seulement le groupe « majoritaire » (sans parent ni grand-parent immigré ni DOMien) attribue de moins en moins de prénoms « français » (et plus de prénoms « internationaux ou anglophones »), comme le montre la figure 1 de Coulmont et Simon (2019, p. 2) ;
- mais en outre, ce que ne montre pas la figure 1 de Coulmont et Simon (2019, p. 2), le groupe « minoritaire » (composé des immigrés et DOMiens et de leurs descendants) est de plus en plus composé de personnes attribuant — et continuant d'attribuer — des prénoms « arabo-musulmans » ;
- et enfin, ce que ne montre pas non plus la figure 1 de Coulmont et Simon (2019, p. 2), le groupe « minoritaire » devient, par rapport au groupe « majoritaire », relativement plus important au fil du temps : alors qu'en 1945-1949, 93 % des naissances en France provenaient du groupe « majoritaire », en 2005-2009 ce n'est plus le cas que de 66 % des naissances.

C'est pour ces deux dernières raisons que, même si le prénom « Mohamed » baisse en fréquence (et en rang d'attribution) lorsqu'on passe des hommes immigrés (G1) du Maghreb à leurs fils (G2) et à leurs petits-fils (G3) (tableau 7.15), il peut croître en fréquence (et en rang d'attribution) parmi les naissances enregistrées à l'état civil en France depuis les années 1950 — tout en ayant régressé parmi les naissances enregistrées du début des années 1980 à la fin des années 1990. Alors que dans les années 1950 moins de 0,1 % des nouveau-nés masculins en France se voyaient attribuer le prénom « Mohamed », depuis la fin des années 2010 c'est plus de 1 % ([Fichier INSEE des prénoms](#)). Le prénom « Mohamed », qui est entré dans le Top 100 des prénoms masculins les plus attribués dans les années 1960, figure dans le Top 20 depuis 2018.<sup>51</sup>

Au total, la figure 1 de Coulmont et Simon (2019, p. 2) présente donc des résultats *sélectifs*, ce qui nuit à la connaissance des faits et biaise l'image globale qui s'en dégage. Si de moins en moins de prénoms attribués en France sont « français », ce n'est pas seulement parce que le groupe « majoritaire » attribue de moins en moins de prénoms « français » (au profit de prénoms « internationaux ou anglophones »), mais aussi parce que le groupe « minoritaire » attribue de plus en plus de prénoms « arabo-musulmans » et parce que ce groupe « minoritaire » devient démographiquement de plus en plus important au fil du temps.

#### ❖ **Reproduction « au plus près » de la figure 3 (page 4)**

Les auteurs indiquent qu'en France métropolitaine en 2008, 63 % des petits-enfants d'immigrés du Maghreb dont le parent enquêté accorde beaucoup d'importance à la religion dans sa vie portent un prénom « arabo-musulman », contre seulement 7 % des petits-enfants d'immigrés du Maghreb dont le parent est sans religion (Coulmont et Simon 2019, figure 3 p. 4). Est-ce vrai ?

Si l'on compare le résultat publié (figure 9.3a) au résultat obtenu *en conservant les petits-enfants de pieds-noirs parmi les petits-enfants d'immigrés du Maghreb*, les résultats sont plus ou moins similaires. En effet, sur l'échantillon retenu pondéré, 57 % des petits-enfants d'immigrés du Maghreb dont le parent enquêté accorde beaucoup d'importance à la religion dans sa vie portent un prénom « arabo-musulman », contre 17 % des petits-enfants d'immigrés du Maghreb dont le parent est sans religion. Mais dans ce cas, seulement 50 % des petits-enfants d'immigrés du Maghreb de la figure 3 de l'article sont d'origine musulmane, 20 % étant d'origine chrétienne, 4 % d'origine juive et 24 % sans origine religieuse.<sup>52</sup> *Une fois les petits-*

<sup>51</sup> À ceci s'ajoute ce que les auteurs notent à juste titre : « La poursuite de l'immigration en provenance du Maghreb, de Turquie et d'Afrique subsaharienne favorise la rétention culturelle car une proportion significative des descendants d'immigré sont en union avec des immigrés, ce qui est plus rarement le cas pour les personnes originaires d'Europe du Sud » (Coulmont et Simon 2019, p. 3).

<sup>52</sup> L'enquête TeO ne renseigne pas l'*affiliation* religieuse des petits-enfants d'immigrés du Maghreb ni, plus

*enfants d'immigrés du Maghreb dissociés des petits-enfants de pieds-noirs*, les résultats obtenus sont sensiblement différents : 70 % des petits-enfants d'immigrés du Maghreb dont le parent enquêté accorde beaucoup d'importance à la religion dans sa vie portent un prénom « arabo-musulman », contre 22 % des petits-enfants d'immigrés du Maghreb dont le parent enquêté est sans religion (figure 9.3b). Dans ce cas, la plupart des petits-enfants d'immigrés du Maghreb sont d'origine musulmane (67 %, contre 10 % d'origine chrétienne, 1 % d'origine juive et 21 % sans origine religieuse).

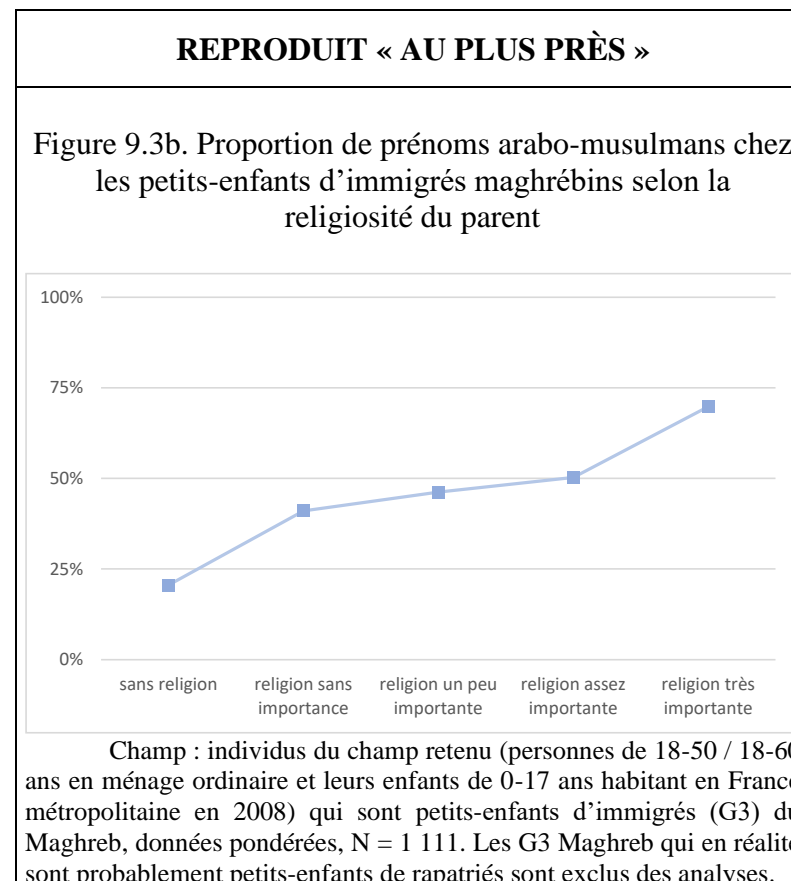
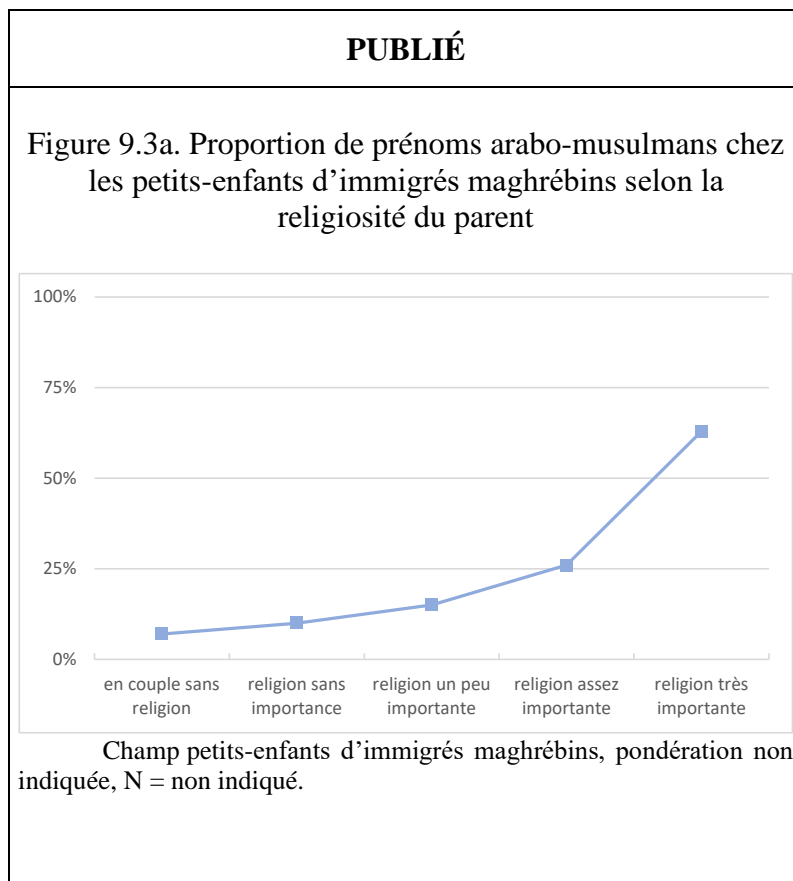
La figure 3 de Coulmont et Simon (2019, p. 4) sous-estime donc sensiblement la part de prénoms « arabo-musulmans » chez les petits-enfants d'immigrés du Maghreb, quel que soit le degré d'importance de la religion dans la vie de leur parent enquêté. Cela dit, il est avéré que plus les petits-enfants d'immigrés du Maghreb ont des parents qui accordent de l'importance à la religion dans leur vie, plus ils se sont vus attribuer des prénoms « arabo-musulmans » (figure 9.3b).

---

largement, l'affiliation religieuse des enfants des enquêtés (RELIG\_IND, variable 11 du tableau 5.1), mais elle permet de leur attribuer une *origine* religieuse (ORIG\_RELIG\_IND, variable 23 du tableau 5.1). Pour rappel, le codage de la variable d'*origine* religieuse des individus est réalisé en 4 étapes : (i) si l'individu est descendant d'un seul parent affilié à une religion (soit christianisme, soit islam, soit judaïsme, soit bouddhisme ou hindouisme, etc.), c'est-à-dire si son autre parent n'est affilié à aucune religion, on lui associe cette origine religieuse ; (ii) si l'individu est descendant de deux parents affiliés à la même religion, on lui associe cette origine religieuse ; (iii) si l'individu est descendant de deux parents affiliés à des religions différentes, on lui associe — conformément à la convention adoptée dans l'enquête TeO pour la variable d'origine (géographique) des individus (ORIG\_IND, variable 11 du tableau 5.1) — la religion de son père ; (iv) si aucun des parents de l'individu n'est affilié à une religion (ils sont tous deux sans religion ou sans réponse), on ne lui associe aucune origine religieuse.



## Comparaison entre la figure 3 publiée page 4 et la même figure reproduite « au plus près »

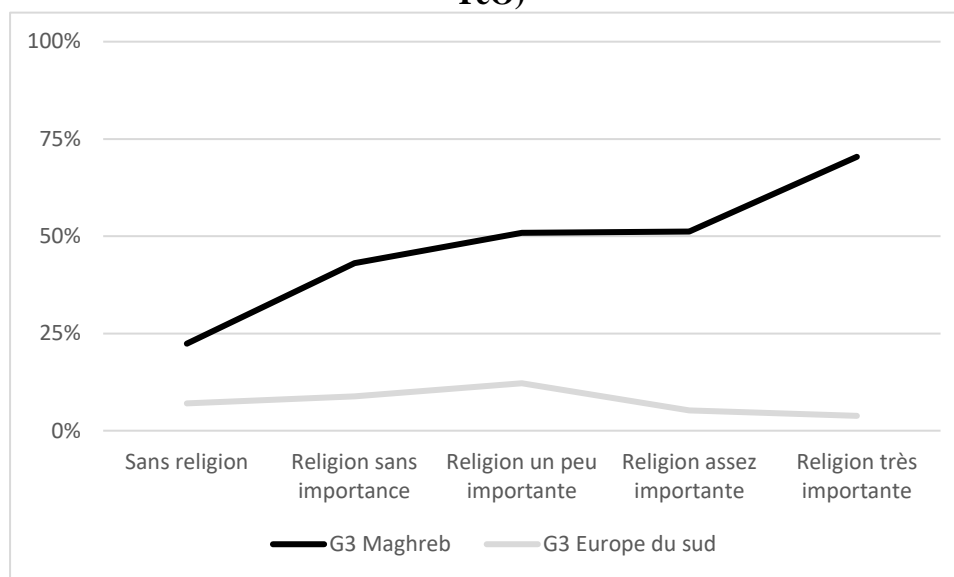


#### ❖ Analyses complémentaires autour de la figure 3 (page 4)

Les petits-enfants d’immigrés du Maghreb portent d’autant plus souvent un prénom spécifique que leur parent accorde plus d’importance à la religion dans sa vie (figure 9.3b). Mais en va-t-il de même pour les petits-enfants d’immigrés d’Europe du Sud ? Globalement, non : si chez les petits-enfants d’immigrés du Maghreb la part de prénoms spécifiques passe de 22 % à 70 % selon l’importance accordée par le parent enquêté à la religion dans sa vie, chez les petits-enfants d’immigrés d’Europe du Sud elle passe de 4 % à 12 % — et cela, de façon non monotone selon l’importance accordée à la religion (figure 9.3c). Autrement dit, si les descendants d’immigrés du Maghreb plus religieux attribuent des prénoms plus souvent « arabo-musulmans », les descendants d’immigrés d’Europe du Sud plus religieux n’attribuent pas des prénoms plus souvent « latins ». Comme on l’a vu de l’Annexe 6A (p. 267) au Chapitre 8 (p. 46), en France comme plus généralement dans le monde, les musulmans disposent d’un stock de prénoms différent de celui des chrétiens. Pas étonnant, dès lors, que plus un enfant d’immigrés du Maghreb accorde de l’importance à la religion dans sa vie, plus il attribue à ses enfants des prénoms « arabo-musulmans », notamment des prénoms coraniques ou les prénoms des proches et successeurs du prophète de l’islam. Par contraste, un enfant d’immigrés d’Europe du Sud qui accorde plus d’importance à la religion dans sa vie peut attribuer aussi bien des prénoms « français » (voire « internationaux ou anglophones ») que « latins » : les premiers ne sont pas moins chrétiens que les seconds.

La figure 9.3c révèle en creux l’écart entre les stocks de prénoms des petits-enfants d’immigrés, selon qu’ils sont originaires du Maghreb ou d’Europe du Sud. Les petits-enfants d’immigrés du Maghreb sont sensiblement plus nombreux que ceux d’Europe du Sud à porter des prénoms spécifiques, et cela reste exact *quelle que soit l’importance de la religion dans la vie de leurs parents* (figure 9.3c).

**Figure 9.3c. Proportion de prénoms spécifiques chez les petits-enfants d’immigrés du Maghreb et d’Europe du Sud, selon la religiosité du parent, en France en 2008 (enquête TeO)**



Lecture : d’après la nomenclature de l’INED, les petits-enfants d’immigrés du Maghreb portent plus souvent des prénoms spécifiques que les petits-enfants d’immigrés d’Europe du Sud, et ce quel que soit le degré de religiosité de leur parent enquêté.

Champ : individus du champ retenu (personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008) qui sont petits-enfants d’immigrés (G3) du Maghreb ou d’Europe du Sud en France métropolitaine en 2008, données pondérées, N = 3 737. Les G3 Maghreb qui en réalité sont probablement petits-enfants de rapatriés sont exclus des analyses.

Source : TeO (voir Annexe 5B, p. 183).

### ❖ Résumé

La figure 1 de l’article présente des résultats *sélectifs*, ce qui nuit à la connaissance des faits et biaise l’image globale qui s’en dégage. Si, depuis la fin des années 1940, de moins en moins de prénoms attribués en France sont « français », ce n’est pas seulement — comme le montre la figure 1 — parce que le groupe « majoritaire » attribue de moins en moins de prénoms « français » (au profit de prénoms « internationaux ou anglophones »), mais aussi — ce que cette figure 1 ne montre pas — parce que le groupe « minoritaire » attribue lui aussi de moins en moins de prénoms français (au profit de prénoms « arabo-musulmans »), et parce que ce groupe « minoritaire » devient démographiquement de plus en plus important au fil du temps. Quant à la figure 3 de l’article, elle sous-estime la part de prénoms « arabo-musulmans » chez les petits-enfants d’immigrés du Maghreb, même s’il est avéré que plus les petits-enfants d’immigrés du Maghreb ont des parents religieux, plus ils se sont vus attribuer des prénoms « arabo-musulmans ».

## 10. Une seule option : la rétractation

La reproduction « au plus près » de l'article de Coulmont et Simon (2019), dont la méthodologie est présentée dans le Chapitre 5 (p. 24) et le Chapitre 6 (p. 25), conduit à trois conclusions à propos des personnes de 18-50 / 18-60 ans en ménage ordinaire et de leurs enfants de 0-17 ans, habitant en France métropolitaine en 2008 :

- ✓ les prénoms les plus fréquents chez les petits-fils d'immigrés du Maghreb sont non pas « Yanis » et « Nicolas » (comme dans le tableau, p. 3 de l'article), mais « Karim » et « Nassim » (Chapitre 7, p. 26) ; en effet, la plupart des « Nicolas » du tableau publié ne sont pas des petits-fils d'immigrés du Maghreb, mais des petits-fils de pieds-noirs ;
- ✓ les petits-enfants d'immigrés du Maghreb ne sont pas seulement 23 % à porter un prénom « arabo-musulman » (comme dans la figure 2, p. 3 de l'article), mais 49 % (Chapitre 8, p. 46) ;
- ✓ et si les deux résultats les plus spectaculaires sont faux, c'est le cas aussi des autres résultats, erronés (figure 3, p. 4) ou sélectifs au point que cela biaise l'image globale qui s'en dégage (figure 1, p. 2) (Chapitre 9, p. 52).

Les principaux résultats de l'article de Coulmont et Simon (2019) étant faux, quelle devrait être l'attitude des auteurs et de l'INED ? Au regard [des normes et des usages de la communauté scientifique internationale](#), plusieurs solutions sont envisageables :

- l'INED pourrait publier une notice exprimant ses réserves quant à certains résultats de l'article (*expression of concern*) ; toutefois, une telle décision ne paraît pas appropriée, car les investigations sur la qualité scientifique de la publication ne sont plus en cours : elles sont désormais publiées ;
- l'INED pourrait publier un erratum corrigeant les erreurs publiées ; mais une telle décision paraîtrait insuffisante, car il ne s'agit pas d'erreurs minimales et/ou d'erreurs sur des aspects secondaires de l'article, mais d'erreurs majeures portant sur ses principaux résultats ;
- les auteurs et/ou l'INED pourraient rétracter l'article, c'est-à-dire l'invalidiser.

Les principaux résultats de l'article de Coulmont et Simon (2019) étant faux, la solution à mes yeux la plus adaptée est la rétractation. Fondamentalement, c'est bien le cœur de l'article (Chapitre 1, p. 9), repris par la plupart des médias en 2019 (Chapitre 2, p. 11), qui est faux : les descendants d'immigrés du Maghreb n'abandonnent pas leurs prénoms spécifiques presque autant que les descendants d'immigrés d'Europe du Sud. Au contraire, les descendants d'immigrés du Maghreb sont, de loin, ceux qui abandonnent *le moins* leurs prénoms spécifiques : 49 % des petits-enfants d'immigrés du Maghreb portent un prénom « arabo-musulman », alors que 8 % des petits-enfants d'immigrés d'Europe du Sud portent un prénom « latin ». L'idée selon laquelle les descendants d'immigrés du Maghreb — par rapport à ceux d'Europe du Sud — abandonnent beaucoup moins leurs prénoms spécifiques en faveur de prénoms proches de ceux du groupe « majoritaire », n'est pas simplement une « idée reçue » : c'est un fait solidement établi.

**L'article doit donc être rétracté en bonne et due forme**, ce qui implique notamment de publier une notice de rétractation mais aussi, [sur le site Internet de l'INED](#), de mentionner la rétractation sur chaque page de la version électronique (PDF) de l'article. Comme l'indique Hervé Maisonneuve, référent à l'intégrité scientifique de la faculté de médecine de l'université de Paris, [une rétractation en bonne et due forme](#) doit non seulement préciser qui est l'initiateur de la rétractation (l'éditeur ou l'auteur) et les motifs de la rétractation, comme l'indique le formulaire standardisé de rétractation qu'on trouvera [ici](#) (Maisonneuve et Decullier 2015). La rétractation doit aussi être mentionnée de façon parfaitement visible sur chaque page de la version électronique de l'article (Maisonneuve 2017), comme le font des revues scientifiques

comme [\*The Lancet\*](#),<sup>53</sup> mais aussi les revues de sciences sociales de bonne tenue comme [\*Social Science & Medicine\*](#) (tableau 10.3).

Difficile, en revanche, de reprocher à la presse d'avoir largement repris les spectaculaires résultats publiés dans la revue d'un établissement public à caractère scientifique comme l'INED (Chapitre 2, p. 11). Les journalistes, même spécialisés dans l'identification des *fake news*, n'ont ni le temps ni les compétences pour juger de la fiabilité des résultats d'un article scientifique. Et les auraient-ils, qu'ils n'auraient pas accès aux bases de données qui leur permettraient de vérifier ces résultats. Comme l'indique la personne chargée de la documentation et de la mise à disposition des enquêtes à l'INED, « Nos données ne sont pas accessibles aux particuliers ni aux journalistes » (Beauchemin et al. 2019, p. 172).

---

<sup>53</sup> Pour un exemple particulièrement médiatisé de rétractation — à l'initiative des auteurs — d'un article publié dans la revue médicale *The Lancet* (juin 2020), voir [ici](#).

### Comparaison entre le tableau publié page 3 et le tableau corrigé

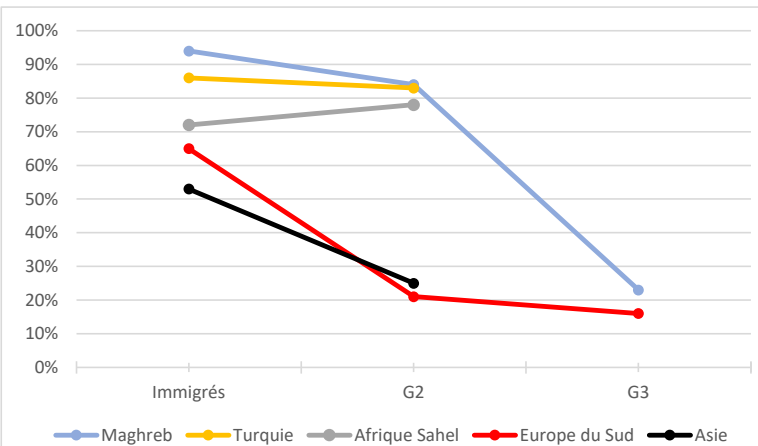
PUBLIÉ				
Tableau 10.1a. Top 3 des prénoms par sexe et générations, Europe du Sud ou Maghreb				
	Europe du Sud		Maghreb	
	Homme	Femme	Homme	Femme
<b>Immigrés</b>	José Antonio Manuel	Maria Marie Ana	Mohamed Ahmed Rachid	Fatima Fatiha Khadija
<b>Descendants d'immigrés</b>	Jean David Alexandre	Marie Sandrine Sandra	Mohamed Karim Mehdi	Sarah Nadia Myriam
<b>Petits-enfants d'immigrés</b>	Thomas Lucas Enzo	Laura Léa Camille	Yanis Nicolas Mehdi	Sarah Inès Lina
Champ non indiqué, pondération non indiquée, N = non indiqué.				

CORRIGÉ				
Tableau 10.1b. Top 3 des prénoms par sexe et générations, Europe du Sud ou Maghreb				
	Europe du Sud		Maghreb	
	Homme	Femme	Homme	Femme
<b>Immigrés</b>	José Antonio Manuel	Maria Marie Isabelle	Mohamed Ahmed Rachid	Fatima Khadija Samira
<b>Enfants d'immigrés</b>	Daniel David Philippe	Marie-Carmen Élisabeth Évelyne	Mohamed Karim Farid	Nadia Sarah Myriam
<b>Petits-enfants d'immigrés</b>	Jean-Baptiste Samuel Thomas	Anne-Laure Émilie Carinne	Karim Nassim Farès	Sarah Sophia Lina
Champ : individus du champ retenu (personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008) qui sont immigrés (G1), enfants d'immigrés (G2) ou petits-enfants d'immigrés (G3) d'Europe du Sud ou du Maghreb, données pondérées, N = 5 630 (d'après les tableaux 7.3 et 7.15). Les G1, G2 et G3 Maghreb qui en réalité sont probablement (descendants de) pieds-noirs sont exclus des analyses.				

## Comparaison entre la figure 2 publiée page 3 et la figure corrigée

### PUBLIÉ

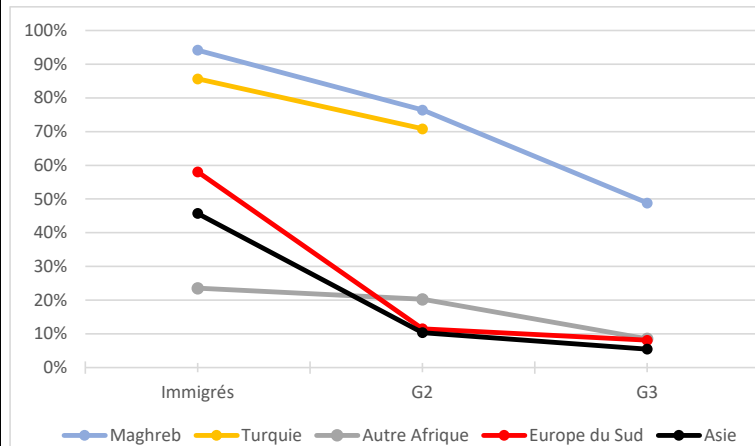
Figure 10.2a. Proportion de prénoms spécifiques par origine et génération (%) (sauf la catégorie « G2 mixtes »)



Champ non indiqué, pondération non indiquée, N = non indiqué.

### CORRIGÉ

Figure 10.2b. Proportion de prénoms spécifiques par origine et génération (%)



Champ : personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008, données pondérées, N = 42 208 (d'après le tableau 8.1a). Les G1, G2 et G3 Maghreb qui en réalité sont probablement (descendants de) rapatriés sont exclus des analyses.



**Tableau 10.3. Exemples récents de rétractations d'articles scientifiques**

Rétractation d'un article de <i>The Lancet</i> sur le traitement du COVID-19 par (hydroxy-) chloroquine (2020)	Rétractation d'un article de <i>Social Science &amp; Medicine</i> sur la mortalité des homosexuels aux États-Unis (2018)
 <p><b>Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis</b></p> <p><b>Summary</b> Background Hydroxychloroquine or chloroquine, often in combination with a second-generation macrolide, are widely used for treatment of COVID-19 despite no conclusive evidence of their benefit. Although they have been used for approved indications such as autoimmune disease or malaria, the safety and benefit of these treatment regimens are poorly evaluated in COVID-19.</p> <p><b>Methods</b> We did a multinational registry analysis of the use of hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19. The registry comprised data from 27 hospitals in 10 countries. We included patients hospitalised between Dec 28, 2019, and April 14, 2020, with a positive laboratory result for SARS-CoV-2. Patients who received one of the treatments of interest within 48 h of diagnosis were included in the four treatment groups (chloroquine alone, chloroquine with a macrolide, hydroxychloroquine alone, or hydroxychloroquine with a macrolide), and patients who received none of these treatments formed the control group. Patients for whom one of the treatments of interest was initiated more than 48 h after diagnosis, or for whom there was mechanical ventilation, as well as patients who received remdesivir, were excluded. The main outcome was in-hospital mortality, as well as the occurrence of de-novo ventricular arrhythmia. The main results were based on ventricular tachycardia or ventricular fibrillation.</p> <p><b>Findings</b> 76 812 patients (mean age 55·8 years, 46·1% women) were hospitalised during the study period and met the inclusion criteria. Of these, 19 318 patients were in the treatment groups (19 318 received chloroquine, 1753 received chloroquine with a macrolide, 16 565 received hydroxychloroquine, and 1221 received hydroxychloroquine with a macrolide) and 57 494 patients were in the control group. 10 478 (51·5%) patients died in hospital. After controlling for multiple comparisons, age, sex, race or ethnicity, body-mass index, underlying cardiovascular disease and its risk factors, underlying lung disease, smoking, immunosuppressed condition, and baseline disease severity), we observed no difference in in-hospital mortality in the control group (5·5%), hydroxychloroquine (4·8%; hazard ratio 1·10, 95% CI 1·05–1·15), chloroquine (5·1%; hazard ratio 1·12, 95% CI 1·07–1·17), hydroxychloroquine with a macrolide (5·2%; hazard ratio 1·14, 95% CI 1·09–1·19), chloroquine with a macrolide (5·2%; hazard ratio 1·14, 95% CI 1·09–1·19), hydroxychloroquine with a macrolide (5·2%; hazard ratio 1·14, 95% CI 1·09–1·19), and chloroquine with a macrolide (5·2%; hazard ratio 1·14, 95% CI 1·09–1·19). These results were consistent across all subgroups. There was no difference in the risk of de-novo ventricular arrhythmia during hospitalisation.</p> <p><b>Interpretation</b> There is no evidence of a benefit of hydroxychloroquine or chloroquine, when used alone or with a macrolide, for treatment of COVID-19. Each of these drug regimens was associated with decreased in-hospital mortality compared with the control group.</p> <p><b>Provenance</b> William H. Hargrett-Anderson, Distinguished Chair in Advanced Cardiovascular Medicine at Brigham and Women's Hospital.</p> <p><b>Copyright</b> © 2020 Elsevier Ltd. All rights reserved.</p> <p><b>Introduction</b> The absence of an effective treatment against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection has led clinicians to redirect drugs that are known to be effective for other medical conditions to the treatment of COVID-19. Key among these repurposed therapeutic agents are the antimalarial drug chloroquine and its analogue hydroxychloroquine, which is used for the treatment of autoimmune diseases, such as systemic lupus erythematosus and rheumatoid arthritis.<sup>1,2</sup> These drugs have been shown in laboratory conditions to have antiviral properties as well as immunomodulatory effects.<sup>3,4</sup> However, the use of this class of drugs for COVID-19 is based on a small number of anecdotal experiences that have shown variable responses in observational observational analyses, and small, open-label, randomised trials that have largely been inconclusive.<sup>5–7</sup> The combination of hydroxychloroquine with a second-generation macrolide, such as azithromycin (for clarithromycin), has also been advocated,</p>	 <p><b>Structural stigma and all-cause mortality in sexual minority populations</b></p> <p><b>Mark L. Hatzenbuehler<sup>a,*</sup>, Anna Bellatorre<sup>b</sup>, Yeonjin Lee<sup>c</sup>, Brian K. Fitch<sup>d</sup>, Peter Muenning<sup>e</sup>, Kevin Fiscella<sup>f</sup></b></p> <p><sup>a</sup>Department of Sociomedical Sciences, Mailman School of Public Health, Columbia University, 722 West 118th Street, New York, NY 10027, United States  <sup>b</sup>Department of Sociology, University of Nebraska-Lincoln, Lincoln, NE, United States  <sup>c</sup>Department of Sociology, University of Pennsylvania, Philadelphia, PA, United States  <sup>d</sup>ABCD-Capabilities, Santa Monica, CA, United States  <sup>e</sup>Department of Health Policy and Management, Mailman School of Public Health, Columbia University, New York, United States  <sup>f</sup>Department of Family Medicine, University of Medicine Medical Center, Rochester, NY, United States</p> <p><b>ARTICLE INFO</b> Article history: Available online 18 June 2018</p> <p><b>ABSTRACT</b> Stigma operates at multiple levels, including interpersonal appraisals (e.g., self-stigma), interpersonal events (e.g., harassment), and structural forms (e.g., community norms, institutional policies). Although prior research has examined interpersonal and interpersonal forms of stigma negatively affect the health of sexual minority populations, few studies have addressed the health consequences of exposure to structural forms of stigma. This gap, we investigated whether structural stigma—operationalized as the presence of laws that criminalize sexual activity between consenting adults—was associated with all-cause mortality. We constructed a measure capturing the average level of anti-gay laws across the United States, using data from the General Social Survey, which was then prospectively linked to mortality data via the National Death Index. Sexual minorities living in communities with high levels of anti-gay laws experienced a higher hazard of mortality than those living in communities with low levels of anti-gay laws. This result translated into a shorter life expectancy for sexual minorities living in high-prejudice communities (12 years [95% CI: 4–20 years] for sexual minorities living in high-prejudice communities). Analyses of specific causes of death revealed that suicide, homicide, violence, and cardiovascular diseases were substantially elevated among sexual minorities in high-prejudice communities. Altogether, there was an 18-year difference in average age of completed suicide between sexual minorities living in high-prejudice (age 55·7) and low-prejudice (age 73·7) communities. These results highlight the importance of examining structural forms of stigma and prejudice as social determinants of health and longevity among minority populations.</p> <p><b>Introduction</b> Stigma includes risk for deleterious mental and physical health outcomes across multiple groups, including racial/ethnic minorities (Patterson, 2006; Williams, 1999), sexual minorities (i.e., individuals who identify as lesbian, gay, or bisexual) (Meyer, 1995), individuals who are overweight/obese (Muehlen, 2008), and those with mental illness (Lusk &amp; Phelan, 2000). Stigma serves as a chronic source of psychological stress (Clark, Anderson, Clark, &amp; Williams, 1999; Link &amp; Phelan, 2006; Major &amp; O'Brien, 2005; Meyer, 2003a; Pachankis, 2007), which in turn contributes to the development of psychopathology (Brown, 1993; Dobrevend, 2000) and disrupts physiological pathways that increase vulnerability to disease (Cherkas et al., 2006; Epel et al., 2004; McEwen, 1998). As substantive evidence emerges that stigma represents an important social determinant of health (Harrison, Phelan, &amp; Link, 2013), researchers have begun to focus on the appropriate measurement and conceptualization of stigma and related constructs (Clark et al., 1999; Krueger et al., 2010; Landolt, 2006; Meyer, 2003b; Quinn &amp; Chaudry, 2009; Williams, Neighbors, &amp; Jackson, 2008). It is widely recognized that stigma operates at multiple levels, including individual (e.g., self-stigma; Mittal,</p>

## 11. Erreurs de bonne foi ou fraude scientifique ?

L'article de Coulmont et Simon (2019) doit être rétracté, pour la simple raison que ses principaux résultats sont faux (Chapitre 7, p. 26 à Chapitre 10, p. 60). Mais ces erreurs ont-elles été commises de bonne foi, ou sont-elles le fruit d'une fraude scientifique ?

### ❖ Des erreurs de bonne foi ?

Tout chercheur peut commettre des erreurs de bonne foi. C'est le cas de Coulmont et Simon eux-mêmes, qui ont notamment décidé de catégoriser les prénoms « arméniens » comme « arabo-musulmans » (message de Patrick Simon du 4 décembre 2020) (Annexe 6B, p. 285). Quoi qu'on pense de cette erreur, elle n'a pour ainsi dire aucune conséquence sur les résultats publiés et communiqués aux journalistes et au grand public, si bien qu'il y a peu de doute qu'elle a été commise *de bonne foi*. Les autres erreurs de l'article pourraient-elles avoir été commises, elles aussi, de bonne foi ?

Pourquoi des auteurs et une institution de recherche qui considèrent leurs résultats comme fiables, ou méthodologiquement défendables, empêcheraient-ils la reproductibilité de leurs résultats (Chapitre 4, p. 15) ? Pourquoi éviter de répondre aux questions posées, pourquoi dissimuler une nomenclature par de fausses excuses, et pourquoi dissimuler un programme informatique sans justification aucune ? Si Coulmont, Simon et l'INED avaient cru en la solidité des résultats qu'ils ont publiés, quel besoin auraient-ils eu de dissimuler leur méthodologie et de violer leurs engagements ainsi que la « [Charte nationale de déontologie des métiers de la recherche](#) » ? S'ils avaient été de bonne foi, pourquoi auraient-ils communiqué à la communauté scientifique une nomenclature fallacieuse et un programme informatique incomplet ? S'il est vrai que

« Le chercheur scientifique n'est pas un homme par nature plus honnête que l'ignorant. C'est quelqu'un qui s'est volontairement enfermé dans des règles telles qu'elles le condamnent, pour ainsi dire, à l'honnêteté » (Revel 1988, p. 17),

le comportement des auteurs et de l'INED offre un aperçu de ce que peut devenir la recherche, lorsqu'elle s'affranchit des règles de l'intégrité scientifique (Chapitre 4, p. 15). Vraisemblablement, les auteurs ne pouvaient ignorer que les résultats qu'ils ont publiés étaient faux ou douteux (encadré 11.1).

### Encadré 11.1. Les auteurs ont-ils publié des résultats qu'ils savaient faux ou douteux ?

Baptiste Coulmont, Professeur des universités à l'Université Paris 8 Vincennes-Saint-Denis au moment de la publication de l'article, et aujourd'hui Professeur des universités à l'École Normale Supérieure Paris-Saclay, est en France le spécialiste incontesté de la sociologie des prénoms — un sujet sur lequel il travaille, [publie](#), [blogue](#) et [tweete](#) régulièrement depuis une quinzaine d'années. Aurait-il pu sérieusement croire qu'en métropole en 2008, le prénom le plus fréquent chez les petits-fils d'immigrés du Maghreb soit, après « Yanis », « Nicolas », sans se poser de questions ni prendre le temps d'examiner les prénoms de l'entourage de la dizaine de petits-fils d'immigrés du Maghreb qui, dans la base TeO, sont prénommés « Nicolas » ? Coulmont ayant enseigné à l'Université Paris 8 Vincennes-Saint-Denis de 2004 à 2020, donc sur une période pendant laquelle les petits-fils d'immigrés du Maghreb prénommés « Nicolas » auraient eu — d'après l'enquête TeO — de 19 à 35 ans en moyenne, aurait-il eu, parmi ses étudiants qui portaient un nom de famille arabo-musulman, presque autant de « Nicolas » que de « Yanis » ? D'autres chercheurs, Araï et al. (2012), observent eux aussi, mais à partir de l'enquête *Histoire de vie* (2003), que « Nicolas » est l'un des 5 prénoms les plus fréquents chez les enfants de deux parents nés au

Maghreb (Araï et al. 2012, p. 148), mais ils n'omettent pas d'indiquer que ces résultats pourraient être dus au fait que leur enquête ne leur permet pas de distinguer les enfants d'immigrés du Maghreb des enfants de rapatriés d'Afrique du Nord (Araï et al. 2012, p. 150). Cela aurait-il pu ne pas alerter Coulmont et Simon (2019), alors que Coulmont connaît bel et bien cet article (Araï et al. 2012), puisqu'il le cite dans son ouvrage de référence sur la sociologie des prénoms (Coulmont 2014, p. 73-74) ?

Quant à Patrick Simon, Directeur de recherche à l'INED, directeur de département à l'Institut Convergences Migrations et chercheur associé à Sciences Po Paris, il a travaillé sur les migrations, les immigrés, les minorités et leurs descendants en France et publié sur ces sujets [des dizaines d'articles scientifiques depuis plus d'une vingtaine d'années](#). Aurait-il lui aussi entendu parler, au cours de sa carrière, de descendants d'immigrés du Maghreb souvent prénommés « Nicolas » ? Simon ayant abondamment travaillé sur la catégorisation statistique des immigrés du Maghreb et des rapatriés de l'ex-empire colonial français en Afrique du Nord, aussi bien avant l'enquête TeO (Simon 1998 ; Simon 2008) qu'en rapport avec cette enquête (Beauchemin et al. 2016) et même depuis cette enquête (Escafré-Dublet et al. 2018 ; Escafré-Dublet et al. 2020), et ayant aussi été [rapporteur d'une thèse](#) sur les pieds-noirs (Couto 2014, citée par Escafré-Dublet et al. 2018, p. 50), aurait-il pu ne pas envisager que ces « Nicolas » soient des descendants de pieds-noirs ? Simon étant, enfin, l'un des concepteurs de l'enquête TeO et, à ce titre, l'un de ses meilleurs connaisseurs et l'un des rares chercheurs qui soit autorisé à accéder directement à la base de données *nominative*, aurait-il pu, depuis le temps qu'il a un accès privilégié à cette base, ne pas remarquer que, bizarrement, de nombreux petits-enfants d'immigrés du Maghreb portaient des prénoms français ou européens, tout comme leur entourage ?

Enfin, les auteurs auraient-ils pu ne pas être alertés par le fait que, parmi les petits-enfants d'immigrés du Maghreb de leur figure 3, 20 % sont d'origine chrétienne, 24 % sans origine religieuse et 4 % d'origine juive, si bien que seulement 50 % sont d'origine musulmane ? On lit :

« Les musulmans gardent des prénoms spécifiques plus souvent que les personnes sans religion et les chrétiens, et le degré de religiosité a une influence quelle que soit la religion » (Coulmont et Simon 2019, p. 4).

Au moins l'un des auteurs a donc bel et bien analysé l'origine des prénoms par degré de religiosité et par affiliation religieuse, si bien qu'il peut difficilement ignorer que seulement la moitié des petits-enfants d'immigrés du Maghreb de la figure 3 de l'article sont d'origine musulmane. Les auteurs auraient-ils pu connaître cette information sans chercher à en savoir plus sur les origines religieuses si bizarres des petits-enfants d'immigrés du Maghreb ?

On peine à envisager que des chercheurs compétents, des spécialistes comme Coulmont et Simon aient pu croire en la fiabilité des résultats qu'ils ont publiés. Il est plus vraisemblable qu'ils ne savaient que trop bien que leurs spectaculaires résultats étaient faux. C'est pourquoi, dès fin avril, début mai 2019, ils ont empêché leur libre examen (Annexe 4A, p. 94).

### ❖ Une fraude scientifique ? Le passif de Patrick Simon et de l'INED

Baptiste Coulmont n'a jusqu'à ce jour jamais été mis en cause pour méconduite scientifique, et il n'y a pas de raison de penser qu'il en ait déjà commis. Patrick Simon, en revanche, a déjà été épinglé pour avoir cherché à tromper son lecteur. Dans le compte rendu que fait la démographe Michèle Tribalat (2016) de l'ouvrage de l'INED sur l'enquête TeO (Beauchemin et al. 2016), elle critique un chapitre dont Patrick Simon est co-auteur, à propos

de la religiosité des immigrés et de leurs descendants (Simon et Tiberj 2016). Comme le montrent les deux points suivants, « L'étude des affiliations religieuses frise la malhonnêteté » (Tribalat 2016, p. 701-702) :

« Enfin, les auteurs développent une analyse spéculaire de la transmission de l'islam. Alors que celle-ci s'est considérablement améliorée au fil du temps auprès des enfants d'immigré, les auteurs concluent à « une sécularisation progressive au fil des générations ». Pour cela ils ont construit un indicateur en trois postes : sécularisation, reproduction, renforcement. Seront dits en voie de sécularisation ceux qui accordent moins d'importance à la religion que n'en accordaient leurs parents dans leur éducation. Ainsi, ceux, très nombreux, dont les parents accordaient beaucoup d'importance à la religion et qui en accordent tout autant sont dans la reproduction. Ils accordent autant d'importance à la religion que leurs parents. Aucun renforcement n'est possible pour eux car il n'y a rien au-dessus de « beaucoup ». À l'opposé, ceux, peu nombreux, dont les parents n'accordaient pas d'importance à la religion, et qui déclarent en accorder un peu, sont dans le renforcement. La sécularisation concerne tout aussi bien ceux qui passent de « pas du tout » à « sans religion » que ceux qui passent de « beaucoup » à « assez ». La part de « renforcement » sera d'autant plus faible que l'on est déjà au top, côté parents ! Or, au fil des générations nées en France, les parents ont accordé de plus en plus d'importance à la religion. Cette « entourloupe » permet aux auteurs de conclure que « le regain religieux ne concerne qu'une petite minorité de jeunes musulmans » et « à une forme atténuée de sécularisation et de baisse de religiosité qui signale une transformation du rôle de la religion ». Rappelons simplement que, des enfants d'au moins un parent musulman nés en France en 1958-1964 à ceux nés en 1985-1989, la transmission est passée de 43 % à 87 % ! Considérant — comme nous l'avons fait — l'effondrement, entre MGIS 1992 et TEO 2008, de la proportion de sans religion chez les jeunes d'origine algérienne, les auteurs invoquent « une religiosité réactive » liée « au durcissement du cadre laïque » et à une « vision péjorative de l'islam ». » (Tribalat 2016, p. 702)

« Les auteurs invalident l'idée d'un communautarisme musulman. La question sur les amis rencontrés au cours des quinze derniers jours est censée renseigner sur leur sociabilité. Si l'enquêté a rencontré un ami seulement et qu'il est de même religion que lui, c'est du 100 %, et réciproquement. Les enquêtés sont déclarés religieusement homophiles s'ils ont rencontré un ou des amis dont plus de la moitié a la même religion. 2 % des musulmans seulement ne savent pas dénommer la religion du ou des amis rencontrés. Mais c'est le cas de 16 % des catholiques et de 28 % des personnes sans religion. Au lieu de prendre au sérieux leur réponse reflétant probablement une réelle indifférence à l'affiliation religieuse de leurs amis, les auteurs ont recalculé la proportion d'homophiles en sortant les réponses de ceux qui disent ne pas connaître la religion de leur(s) ami(s). Et là, miracle, ce sont les catholiques (65 %) et les sans religion (50 %) qui sont nettement plus homophiles que les musulmans (43 %). Le tour est joué : « les plus “communautaires” ne sont pas ceux qu'on croit ». Les auteurs pensent invalider ainsi « les représentations médiatiques d'une communauté repliée sur elle-même ». La possibilité d'un biais lié à la position majoritaire n'est évoquée que pour les origines. Pas pour les religions. Pourtant, si l'on retient la composition religieuse des 18-50 ans de TEO et si les religions étaient réparties uniformément sur tout le territoire sans tenir compte d'aucune affinité particulière, la probabilité théorique des catholiques de fréquenter des catholiques serait de 43 %, celle des sans religion de fréquenter des personnes sans religion de 45 % et celle des musulmans de fréquenter des musulmans de 8 %. Rien d'étonnant à ce que l'homophilie des catholiques soit plus élevée que celle des musulmans. » (Tribalat 2016, p. 702)

Récapitulons. Dans un chapitre d'ouvrage publié par l'INED en 2016, Patrick Simon et un coauteur ont publié une interprétation fallacieuse d'un indicateur, leur permettant d'accréditer la thèse selon laquelle les musulmans seraient, dans leurs choix d'amis, plus

ouverts que les catholiques. La manipulation consiste à inférer que les musulmans sont moins « communautaires » que les catholiques à partir du fait qu'il est plus rare pour les musulmans que pour les catholiques d'avoir plus de la moitié de leurs amis qui sont de la même religion qu'eux (42 % contre 54 %), alors que cela dépend avant tout du fait que le stock d'amis potentiels catholiques est beaucoup plus grand que le stock d'amis potentiels musulmans (43 % des 18-60 ans en métropole en 2008, contre 8 %) (Simon et Tiberj 2016, p. 562 et 574-576).<sup>54</sup> Dans le même chapitre, Patrick Simon et son coauteur présentent un autre indicateur trompeur permettant d'accréditer la thèse selon laquelle, au fil des « générations », les descendants d'immigrés se séculariseraient presque autant que la population « majoritaire » (Simon et Tiberj 2016, p. 571-574). La manipulation consiste ici à calculer la part d'individus qui sont moins — ou plus — religieux que leurs parents, un indicateur biaisé par le fait que de nombreux enfants de la population « majoritaire », dont les parents sont peu religieux, ne peuvent pas l'être moins que leurs parents (cet « effet plancher » conduisant à sous-estimer la baisse de la religiosité de la population « majoritaire » au fil des « générations »), tandis que de nombreux enfants d'immigrés, dont les parents sont très religieux, ne peuvent pas l'être plus que leurs parents (cet « effet plafond » conduisant à sous-estimer la hausse de la religiosité des immigrés — et plus particulièrement des musulmans — et de leurs descendants au fil des « générations »).<sup>55</sup> **Patrick Simon et l'INED ont donc un passif en matière de manipulations statistiques destinées à « démontrer » que, contrairement aux idées reçues, les immigrés du Maghreb et leurs descendants s'intègrent pleinement dans la société française.** Depuis la publication de ces manipulations — et leur dénonciation — en 2016, ces manipulations et le biais qu'elles révèlent n'ont été contestés ni par leurs auteurs, ni par l'INED. Mais elles sont restées sans suite.

### ❖ Une fraude scientifique ? Le mobile idéologique et politique

Si en mathématique et en sciences physiques la fraude scientifique semble rare car la plupart des démonstrations et des expériences sont reproductibles, en sciences de la vie et en sciences médicales les auteurs de fraude scientifique semblent surtout motivés par la perspective de publier dans une grande revue et, par-là, d'obtenir un poste, une promotion ou des moyens supplémentaires de recherche. En sciences sociales, de telles motivations sont, au moins dans le cas qui nous occupe, peu vraisemblables. Un chercheur — et a fortiori un chercheur confirmé comme Coulmont ou Simon — n'a pas d'intérêt carriériste à frauder pour pouvoir publier dans *Population et sociétés*, car une telle publication n'est pas susceptible de

<sup>54</sup> « Les taux d'homophilie religieuse calculés pour les seuls [sic] réponses exprimées montrent que des trois groupes étudiés [sans religion, chrétiens et musulmans], les musulmans sont les moins homophiles. Les athées fréquentent ainsi plus souvent des personnes sans religion comme eux. D'une certaine façon, les plus "communautaires" ne sont pas ceux que l'on croit. [...] La diversité sociale, ethnique et religieuse des cercles de relations des musulmans invalide les représentations médiatiques d'une communauté repliée sur elle-même. Par comparaison avec les catholiques et les athées, leurs amis sont plus fréquemment choisis en dehors de la communauté d'orientation religieuse » (Simon et Tiberj 2016, p. 576).

<sup>55</sup> Au fil des années de naissance, les enfants d'au moins un parent catholique sont plutôt moins nombreux à se déclarer, à leur tour, catholiques (de 68 % pour les enquêtés nés dans les années 1960 à 60 % pour ceux nés dans les années 1980). Par contraste, les enfants d'au moins un parent musulman sont de plus en plus nombreux à se déclarer, à leur tour, musulmans, qu'il s'agisse des immigrés arrivés en métropole avant 16 ans (de 76 % pour ceux nés dans les années 1960 à 93 % pour ceux nés dans les années 1980) ou a fortiori des natifs (de 43 % pour ceux nés dans les années 1960 à 87 % pour ceux nés dans les années 1980) (Tribalat 2013, p. 158 et 161). De même, au fil des années de naissance, la proportion de chrétiens qui déclarent accorder « beaucoup d'importance » à la religion dans leur vie baisse légèrement (de 12 % pour ceux nés dans les années 1950 à 10 % pour ceux nés dans les années 1990), alors que la proportion de musulmans qui déclarent accorder « beaucoup d'importance » à la religion dans leur vie continue d'augmenter (de 54 % pour ceux nés dans les années 1950 à 60 % pour ceux nés dans les années 1990). Pour une exploitation fiable de l'enquête TeO sur les affiliations religieuses, la religiosité et la pratique religieuse, voir Tribalat 2013.

favoriser sa carrière. En sciences sociales, et tout particulièrement en sociologie, et plus particulièrement encore dans les champs de recherche politiquement « brûlants » comme la sociologie de l'immigration et la démographie des migrations, la fraude peut avoir des motivations moins carriéristes qu'*idéologiques*.

En sociologie de l'immigration et en démographie des migrations, il semblerait qu'il s'agit avant tout d'invalider les « préjugés » négatifs à propos des immigrés et de leurs descendants, de rassurer les Français sur leur rythme d'assimilation à la société française, et plus généralement d'influer favorablement sur les perceptions et les attitudes du grand public à leur égard, pour améliorer leur image, réduire la prévalence des discriminations à leur encontre et ainsi faire advenir un monde meilleur. En l'occurrence, montrer que les prénoms les plus courants chez les petits-fils d'immigrés du Maghreb sont « Yanis » et « *Nicolas* », et montrer plus généralement que seulement 23 % des petits-enfants d'immigrés du Maghreb portent un prénom arabo-musulman, c'est démontrer — « déconstruire » — les clichés, les idées reçues, les préjugés et les supposés « problèmes d'intégration » qui nourrissent les discriminations, le racisme et l'extrême droite. Ce mobile idéologique permet de comprendre pourquoi, sur la figure 1 de l'article (Chapitre 9, p. 52), les auteurs présentent des résultats *sélectifs* : ils ont omis de présenter les résultats (figure 9.2) qui auraient pu souligner l'origine de plus en plus fortement maghrébine et africaine des immigrés et de leurs descendants. Le problème, ici, n'est pas que de nombreux spécialistes de l'immigration aient à cœur d'améliorer l'image de l'immigration. **Le problème décisif survient lorsque, pour améliorer l'image de l'immigration, un chercheur ou un éditeur s'autorise à biaiser les résultats qu'il publie.** Le guide du CNRS [« Pratiquer une recherche intègre et responsable »](#) suggère qu'un chercheur « doit avoir l'honnêteté de distinguer entre ses connaissances scientifiques et ses opinions personnelles auxquelles son statut de chercheur ne donne aucune légitimité particulière ».

**Pour un chercheur-militant de la mouvance idéologique « décoloniale » comme Patrick Simon, l'objectif est de montrer qu'il n'existe pas de problème d'intégration.** Quant à Baptiste Coulmont, qui n'est ni militant ni « décolonial », ce sera à lui de préciser pourquoi il a consenti à publier des résultats dont il ne pouvait ignorer qu'ils étaient faux ou douteux (encadré 11.1), avant d'en dissimuler la méthodologie. Une hypothèse possible est qu'il aurait décidé de publier des résultats aussi spectaculaires, dans une revue à forte visibilité, en raison non pas de leur valeur idéologique, mais de leur potentiel médiatique. Si cette hypothèse était correcte, son vœu semble avoir été exaucé : son article lui a donné accès non seulement à la presse écrite — dont il est familier, puisqu'il est chroniqueur au journal *Le Monde* —, mais aussi à [France Info](#), [BFM TV](#) et [RMC](#) (Chapitre 2, p. 11).

#### ❖ Une fraude scientifique à but idéologique et politique, permise par la complaisance du champ de recherche sur l'immigration

Qu'un chercheur puisse, par idéologie ou pour une autre raison, être tenté de publier des résultats qu'il sait faux, cela peut arriver. Mais comment se fait-il que rien ne le retienne ?

##### **La complaisance au sein de l'INED**

Comment le comité de rédaction (*peer review*) de la revue *Population et sociétés*, son rédacteur en chef et la directrice de la publication ont-ils pu laisser publier l'article de Coulmont et Simon (2019) sans leur poser plus de questions sur leurs résultats les plus spectaculaires ? Une fois l'article publié (Chapitre 3, p. 12), pourquoi la [vingtaine de chercheurs de l'INED qui sont spécialistes des migrations internationales et des minorités](#), et plus généralement les [plus de 70 chercheurs de l'INED](#) (et [la quarantaine de chercheurs associés à l'INED](#)), ne s'en sont-ils pas étonnés ? Une fois que j'ai alerté l'INED sur le fait que les auteurs d'une de leurs publications refusaient de communiquer leur méthodologie, la hiérarchie de l'INED a refusé d'assumer ses fonctions et n'a pas exigé des auteurs qu'ils communiquent leur méthodologie, comme en témoignent les dizaines d'échanges de messages électroniques que j'ai dû avoir, à



partir du 26 septembre 2019, non plus seulement avec les auteurs mais aussi avec le rédacteur en chef de la revue, puis avec la directrice de la publication et directrice de l'INED, puis avec le référent à l'intégrité scientifique de l'INED (Annexe 4A, p. 94). L'INED aurait-il cru pouvoir esquiver la discussion scientifique en décourageant mes investigations (Chapitre 4, p. 15) ? Subsistent enfin quelques questions troublantes. Quelle est la probabilité que l'article de Coulmont et Simon (2019), qui est le premier dont je me permets de demander la méthodologie, soit aussi le premier article publié par l'INED sur l'immigration à présenter des résultats non seulement sélectifs, mais erronés ? **Les publications de l'INED sur l'immigration ne sauraient être tenues pour fiables avant qu'elles aient été systématiquement vérifiées**, comme je l'ai fait pour l'article de Coulmont et Simon (2019).

### La complaisance des spécialistes des migrations

Comment les [plusieurs centaines de chercheurs de l'Institut Convergences Migrations](#) ont-ils pu laisser passer les résultats de Coulmont, Simon et l'INED sans rien dire ?<sup>56</sup> Je ne crois pas que ce milieu professionnel soit déconnecté du réel au point d'avoir pu croire sur parole de tels résultats. Simplement, des chercheurs avant tout soucieux de l'image qu'ont les Français des immigrés et de leurs descendants n'éprouvent que rarement le besoin d'évaluer sérieusement la fiabilité des publications scientifiques de leurs collègues, tant les résultats publiés sont, le plus souvent, idéologiquement confortables. Une sociologue, qui a accueilli Patrick Simon comme [chercheur associé](#) à l'Observatoire sociologique du changement de Sciences Po Paris plusieurs mois *après* la parution de son article avec Coulmont (2019), est une spécialiste de l'immigration : aurait-elle entendu parler de nombreux petits-fils d'immigrés du Maghreb prénommés « Nicolas », et envisagerait-elle que seulement 23 % des petits-enfants d'immigrés du Maghreb portent un prénom « arabo-musulman » ? Plus généralement Patrick Simon, qui est le chercheur français le plus fréquemment cité sur l'immigration, n'est pas moins souvent cité depuis que ses manipulations statistiques ont été exposées, en 2016 (figure 11.2).

**Figure 11.2. Nombre annuel de citations des publications scientifiques de Patrick Simon**



Source : [Google Scholar](#), consulté le 1<sup>er</sup> mai 2021.

En France actuellement, un institut est chargé de fédérer la recherche en sciences sociales sur les migrations : c'est l'Institut Convergences Migrations (ICM).<sup>57</sup> Il comprend 4

<sup>56</sup> Le fait que des chercheurs qui ne sont ni démographes ni spécialistes des migrations (Corpart 2019 ; Piketty 2019 ; De Singly 2020) puissent citer l'article de Coulmont et Simon (2019) est moins étonnant.

<sup>57</sup> Parmi les [missions de l'Institut Convergences Migrations](#), figure celle d'« éclairer le débat public » : « Soucieux de renouer le lien science/sociétés sur la question des migrations, l'IC Migrations se donne les moyens d'un décryptage critique et constructif de l'actualité par le biais de sa revue De facto. Face à des sujets de société aussi sensibles que les migrations, le scientifique ne peut rester au-dessus de la mêlée ; il doit s'y engager avec l'ambition

départements, dont l'un est sous la responsabilité de Patrick Simon ([département INTEGER](#), censé étudier les processus d'intégration des immigrés et de leurs descendants). Or, le directeur de l'ICM, François Héran, ne peut ignorer le passif de Patrick Simon : non seulement il a préfacé le livre de l'INED dont l'un des chapitres est celui qui contient les manipulations statistiques rappelées ci-dessus (Héran 2016, dans Beauchemin et al. 2016), mais en outre il a connaissance de la recension de Tribalat (2016) qui expose lesdites manipulations, puisqu'il la cite dans l'un de ses ouvrages (Héran 2017, p. 292). Malgré cela, Patrick Simon a été invité au Collège de France pour donner une présentation intitulée « [L'intégration, une responsabilité de la majorité](#) », dans un séminaire organisé en lien avec le cours de François Héran « Migrations et sociétés » (5 février 2020). François Héran, l'Institut Convergences Migrations et le Groupe international d'experts sur les migrations (GIEM), soit la crème des scientifiques français spécialistes des migrations et de l'immigration, ont même signé avec Patrick Simon une tribune du *Huffington Post* intitulée « [Comment parler d'immigration en scientifique ?](#) » (octobre 2019). Et tout cela, plusieurs années *après* qu'ont été exposées publiquement les méconduites scientifiques de Patrick Simon (Tribalat 2016), et plusieurs mois *après* la parution et la médiatisation de l'article de Coulmont et Simon (avril 2019). Un chercheur-militant dont on sait qu'il a un passif de méconduites scientifiques à visée idéologique, et qui vient de publier sur les prénoms des descendants d'immigrés un article aussi spectaculaire que politiquement opportun, peut-il encore être considéré comme fiable *sans vérification sérieuse* ?

Je soutiens que la sociologie de l'immigration et la démographie des migrations (ou des « minorités »), telles qu'elles sont globalement pratiquées en France aujourd'hui, souffrent d'un biais idéologique. Et pour cause : des chercheurs qui visent à améliorer l'image de l'immigration ont toujours mieux à faire que de contrôler la qualité des publications qui servent leur cause — même lorsqu'elles sont, d'évidence, sélectives ou simplement fausses. La complaisance idéologique d'une large part des sociologues de l'immigration et des démographes des migrations les conduit à fermer les yeux sur le manque de crédibilité scientifique des « pieux mensonges » de leurs collègues militants. C'est dans ce contexte que Patrick Simon a pu occuper — et continue d'occuper — diverses positions d'influence dans des institutions publiques françaises et internationales.<sup>58</sup>

### **La complaisance en sciences sociales**

La complaisance idéologique des spécialistes des migrations en France s'insère dans un climat d'impunité plus général en sciences sociales, dû en partie au fait que, dans notre pays, presque aucun article de sciences sociales n'est jamais rétracté. Selon la base de données [Retraction Watch](#) au 1<sup>er</sup> mai 2021, on recense dans le monde 25 954 rétractations dans des journaux scientifiques, dont 3 577 en sciences sociales, comme des articles fraudés de Diederik Stapel<sup>59</sup> et des canulars contre des revues de *cultural studies* (par exemple, [ici](#) et [là](#)). En France, on compte 25 rétractations classées en sciences sociales, mais à y regarder de près, seules 3 ont

---

d'en relever le niveau. Cela suppose de rappeler les ordres de grandeur des phénomènes, de pratiquer la critique des données existantes, d'en construire de nouvelles, d'effectuer des comparaisons dans le temps et dans l'espace, de décoder les arguments échangés, d'articuler au mieux les dimensions descriptive, explicative et normative des phénomènes observés. »

<sup>58</sup> Patrick Simon a été [expert pour le Conseil de l'Europe \(ECRI\)](#) (2007), [expert pour Eurostat](#) (2007), membre du conseil scientifique de la *Fundamental Rights Agency* de la Commission européenne (2009-2013), de nouveau [expert pour Eurostat](#) (2011) et [expert pour l'ONU \(UNESCO\)](#) (2019). Il a aussi récemment été membre du jury du concours « [No\(s\) Cliché\(s\) : les réfugiés, une chance pour la France](#) » (2019), il a participé à une enquête du [Défenseur des droits](#) sur les discriminations (Défenseur des droits 2020), et il est à ce jour membre de l'[Observatoire de la diversité du Conseil supérieur de l'audiovisuel \(CSA\)](#).

<sup>59</sup> L'[affaire Stapel](#) (2011-2012), du nom d'un chercheur néerlandais en psychologie sociale qui a falsifié les données de plusieurs dizaines d'articles — rétractés depuis — a en partie décrédibilisé sa discipline : collègues peu regardants sur la rigueur méthodologique, évaluateurs de revues scientifiques prompts à permettre la publication de résultats un peu « trop beaux pour être vrais », etc.

effectivement été publiées dans une revue francophone,<sup>60</sup> et aucune d'entre elles ne relève vraiment des sciences sociales. Concrètement, à part quelques salutaires articles-canulars publiés par des revues farfelues et qu'elles ont finalement dû rétracter ([cet article](#), qui était un canular de [Quinon et Saint-Martin \(2015\)](#) destiné à démontrer le manque de sérieux de la [revue Sociétés](#), et [cet article](#), qui était un canular de [Barberousse et Huneman \(2016\)](#) contre la revue *Badiou studies*), les revues françaises de sciences sociales ne rétractent presque *jamais* d'articles. Des chercheurs français en sciences sociales ont bien vu certains de leurs articles rétractés de revues anglo-saxonnes (par exemple, [ici](#) et [là](#)), mais pas de revues françaises. Ainsi *Actes de la recherche en sciences sociales*, une revue de sociologie fondée par Pierre Bourdieu, n'a toujours pas rétracté [un article](#) dont [Bavouzet et Weisz-Patrault \(2018\)](#) ont démontré qu'il « ne satisfait pas aux standards de la méthode scientifique ».

Alors, que faire ? Ce sera à la collectivité des chercheurs en sciences sociales d'en décider, peut-être discipline par discipline. D'ici là, il serait utile que les revues exigent de leurs auteurs qu'ils publient ou s'engagent à communiquer les programmes informatiques et tout autre élément méthodologique légalement communicable et nécessaire à la reproduction à l'identique des résultats. J'invite les chercheurs à ne plus hésiter à demander à leurs collègues leur méthodologie, notamment lorsqu'elle produit des résultats douteux ou biaisés. J'invite les institutions de recherche à consacrer une partie de leur budget à des travaux de reproduction ou de réplication de résultats déjà publiés. J'invite, enfin, la communauté des chercheurs en sciences sociales à discuter d'un secret de polichinelle : aujourd'hui en France, certains champs de recherche « politiquement brûlants », comme la sociologie de l'immigration et la démographie des migrations, sont dominés par des « chercheurs » qui sont aussi des militants — en l'occurrence, « décoloniaux ». Ce contexte militant n'étant pas particulièrement propice au progrès des connaissances scientifiques (Heinich 2021), comment le changer ?

### ❖ Résumé

Pourquoi des auteurs et une institution de recherche qui considèrent leurs résultats comme fiables, ou méthodologiquement défendables, empêcheraient-ils de reproduire leurs résultats ? Quel besoin auraient-ils de communiquer une nomenclature fallacieuse et un programme informatique incomplet, et de violer leurs engagements ainsi que la « [Charte nationale de déontologie des métiers de la recherche](#) » ? Deux chercheurs compétents — l'un spécialiste des prénoms, l'autre spécialiste des immigrés et fin connaisseur des catégorisations statistiques des immigrés du Maghreb et des pieds-noirs — auraient-ils pu croire qu'en métropole en 2008, les prénoms les plus fréquents chez les petits-fils d'immigrés du Maghreb étaient « Yanis » et « *Nicolas* », sans se poser de questions ni prendre le temps d'examiner les prénoms de l'entourage des 11 « Nicolas » concernés ? Auraient-ils pu ne pas être alertés par le fait que, parmi les petits-enfants d'immigrés du Maghreb de leur figure 3, seulement 50 % sont d'origine musulmane ? Vraisemblablement, les auteurs ne savaient que trop bien que leurs spectaculaires résultats étaient faux ou douteux : c'est pourquoi, dès fin avril, début mai 2019, ils ont empêché leur libre examen. Patrick Simon et l'INED ont d'ailleurs un passif en matière de manipulations statistiques destinées à « démontrer » que, contrairement aux idées reçues, les immigrés du Maghreb et leurs descendants s'intègrent pleinement dans la société française. Tout indique que l'objectif de Patrick Simon — chercheur-militant de la mouvance idéologique « décoloniale », Directeur de recherche à l'INED, directeur de département à l'Institut Convergences Migrations et chercheur associé à Sciences Po Paris — est d'influer favorablement sur les attitudes du grand public à l'égard des immigrés et de l'immigration. Il est regrettable que ces méconduites scientifiques bénéficient de la complaisance du milieu de la recherche, non seulement au sein de l'INED mais plus globalement en sociologie et en démographie de l'immigration.

---

<sup>60</sup> Il s'agit de [cet article](#) de 2004 de *Douleurs : Evaluation — diagnostic — traitement*, de [cet article](#) de 2011 des *Annales de dermatologie et de vénéréologie* et de [cet article](#) de 2014 de *L'Evolution psychiatrique*.

## Conclusion

En publiant ce travail, je vise plusieurs objectifs :

- ✓ faire **rétracter** l'article de Coulmont et Simon (2019), et rétablir la vérité au sujet des prénoms des descendants d'immigrés en France ;
- ✓ alerter la communauté des chercheurs en sciences sociales sur une **imposture scientifique** : l'INED, qui est l'institut de recherche censé informer les Français et les pouvoirs publics sur les questions démographiques, non content d'avoir publié — et diffusé aux journalistes — des énormités, a violé les règles de l'intégrité scientifique et m'a empêché de vérifier ses résultats ;
- ✓ alerter les journalistes et le grand public sur une **fraude scientifique** dont l'objectif est manifestement de biaiser l'image de l'immigration aux yeux des Français ;
- ✓ alerter sur le contexte académique qui a permis tout cela : la **complaisance** du milieu de la recherche en sociologie et en démographie de l'immigration, et le manque de fiabilité scientifique de ce champ de recherche, qui souffre d'un biais idéologique volontiers « décolonial ».

En décrédibilisant la recherche sur l'immigration *aux yeux des chercheurs* de champs de recherche connexes (comme l'auteur de ces lignes), la fraude empêche l'accumulation des connaissances. En décrédibilisant la recherche en sciences sociales *aux yeux du grand public et des élus*, les méconduites scientifiques contribuent aussi à nourrir les approches irrationnelles du monde social. Comme l'indique le [réfèrent à l'intégrité scientifique du CNRS](#),

« L'intégrité scientifique est une nécessité absolue au regard de la notion de confiance : confiance entre scientifiques, pour faire progresser la connaissance, mais aussi celle du grand public. La société doit avoir confiance dans le travail de chercheurs, condition pour le succès d'approches rationnelles sur des questions sociétales. »

Enfin, en biaisant les faits publiés et communiqués aux journalistes, au grand public et aux pouvoirs publics, les méconduites scientifiques dégradent le débat démocratique. Si, comme je le crois, la recherche en sciences sociales doit aider à éclairer le débat public et à lutter contre certaines idées colportées par des militants de tous bords, c'est en apportant aux citoyens des arguments *factuels*, fondés sur une méthode scientifique dénuée d'idéologie.

Je termine en répondant à quatre questions qu'un lecteur pourrait se poser.

- **Pourquoi avoir choisi de prendre du temps pour reproduire les résultats d'un article déjà publié, plutôt que d'allouer ce temps à des recherches nouvelles ? Pourquoi avoir choisi de reproduire cet article-là, plutôt qu'un autre ?**

Depuis quelques années j'ai commencé à douter de l'impartialité et de la fiabilité scientifiques de beaucoup de publications de sociologie et même de démographie de l'immigration, et pas seulement celles de Patrick Simon. À la lecture de l'article fortement médiatisé de Coulmont et Simon (2019), ses résultats me sont apparus comme douteux (Chapitre 3, p. 12). Je me suis alors dit que cet article pourrait constituer un exemple concret à partir duquel mettre mes doutes à l'épreuve.

- **Pourquoi avoir choisi de publier ce travail, plutôt que de le garder pour moi ?**

En publiant ce travail, je souhaite *alerter* la communauté scientifique, les journalistes, le grand public et les pouvoirs publics sur le manque d'impartialité et de fiabilité scientifiques des publications de sociologie de l'immigration et de démographie des migrations. Je soutiens que pour que les sciences sociales deviennent intègres et fiables, même sur les sujets politiquement

« brûlants », et pour qu'elles puissent contribuer au débat démocratique en lui apportant des faits solidement établis, nos pratiques de chercheurs doivent changer.

Les auteurs et l'INED ayant refusé de respecter les règles de l'intégrité scientifique, je dois — pour faire rétracter leur article — prendre à témoin la communauté scientifique, ainsi que les journalistes et le grand public, qui ont été trompés.

➤ **Pourquoi avoir choisi de publier cet essai de reproduction maintenant ?**

Je n'ai pas pu publier ce travail plus tôt, parce que je n'ai pu commencer l'analyse statistique qu'au printemps 2020, et parce que pour réaliser le travail de reproduction ici présenté, il m'a fallu plus d'une année. C'est pourquoi je viens seulement de terminer de rédiger (et de relire) le présent document.

➤ **Est-ce que je ne voudrais pas régler des comptes contre les auteurs ou contre l'INED ?**

Je n'ai jamais rencontré Patrick Simon. Deux mois avant la publication de l'article, en février 2019, j'avais croisé Baptiste Coulmont (car lui et moi travaillions dans le même bâtiment, à un étage d'écart) et la discussion, de moins d'un quart d'heure, avait été cordiale et suivie d'un échange de messages électroniques à propos de textes récents sur les prénoms des immigrés et de leurs descendants (Coulmont 2017a, 2017b et 2017c), et à propos des résultats de l'article à paraître deux mois plus tard. Depuis la publication de l'article, j'ai rencontré Baptiste Coulmont deux fois pour discuter de l'article avec lui (le 17 avril et le 13 mai 2019), puis une troisième fois (le 16 juin 2020) je l'ai rencontré par hasard dans notre bâtiment. N'ayant à ce jour jamais publié sur l'immigration ni sur les prénoms, ni même participé à des séminaires ou des colloques sur ces sujets, je n'ai pas eu plus de contacts avec ces chercheurs et n'ai aucun motif personnel d'hostilité à leur égard. Quant à l'INED, j'ai de l'estime intellectuelle pour la plupart de ses publications sur la plupart des sujets chers aux démographes : fécondité, mortalité, démographie mondiale ou démographie historique. Mais je crains qu'à l'image de la sociologie des immigrés, la démographie des migrations et des « minorités » souffre d'un biais idéologique volontiers « décolonial ». C'est sur ce biais que je souhaite alerter.



## Références

### ❖ Article reproduit

COULMONT Baptiste, SIMON Patrick. Quels prénoms les immigrés donnent-ils à leurs enfants en France ? *Population et sociétés*, 2019, 565, p. 1-4. [https://www.ined.fr/fichier/s\\_rubrique/29081/565.population.societes.avril2019.immigres.prenoms.france.fr.pdf](https://www.ined.fr/fichier/s_rubrique/29081/565.population.societes.avril2019.immigres.prenoms.france.fr.pdf)

### ❖ Autres références

ALFORD, Richard D. *Naming and identity: A cross-cultural study of personal naming practices*. New Haven: HRAF, 1988.

ALGAN Yann, MAYER Thierry, THOENIG Mathias. The Economic Incentives of Cultural Transmission: Spatial Evidence from Naming Patterns across France. *CEPII Working Paper*, 2013. [http://www.cepii.fr/PDF\\_PUB/wp/2013/wp2013-25.pdf](http://www.cepii.fr/PDF_PUB/wp/2013/wp2013-25.pdf)

ALGAVA Élisabeth, LHOMMEAU Bertrand. À l'origine de l'enquête TeO. Enjeux de l'échantillonnage, collecte et pondérations de l'enquête. *Document de travail n°F1304 de la Direction des statistiques démographiques et sociales*, 2013. <https://www.insee.fr/fr/statistiques/fichier/1381084/F1304.pdf>

ALLEGRA, Luciano. La famiglia ebraica torinese nell'Ottocento. Le spie di un'integrazione sociale. *Il matrimonio ebraico. Le ketubbot dell'Archivio Terracini* / dir. par Micaela VITALE. Torino: Silvio Zamorani editore, 1997, p. 67-111.

AMARA, Muhammad Hasan. Palestinian Given Names in Israel and the West Bank (Palestine): Political Boundaries and Socio-cultural Reflections. 2014, p. 214-230. [https://www.researchgate.net/publication/312219409\\_Palestinian\\_Given\\_Names\\_in\\_Israel\\_and\\_the\\_West\\_Bank\\_Palestine](https://www.researchgate.net/publication/312219409_Palestinian_Given_Names_in_Israel_and_the_West_Bank_Palestine)

ARAÏ Mahmood, BESANCENOT Damien, HUYNH Kim, SKALLI Ali. Children's First Names, Religiosity and Immigration Background in France. *International Migration*, 2015, 53(6), p. 145-152. <https://onlinelibrary.wiley.com/doi/abs/10.1111/imig.12010>

ASHDOWN, Nick. What's in a Name? The fraught and funny world of Turkish nomenclature. *Newlines Magazine*, 2021. <https://newlinesmag.com/reportage/whats-in-a-name/>

ASLAN, Senem. Incoherent State: The Controversy over Kurdish Naming in Turkey. *European Journal of Turkish Studies*, 2009, 10, p. 1-20. <https://journals.openedition.org/ejts/pdf/4142>

AUGER Nathalie, MOÏSE Claudine. Ancrages énonciatifs et identitaires : le choix du prénom des enfants dans les couples mixtes. 2000. <https://hal.archives-ouvertes.fr/hal-02512189/document>

AUGER Nathalie, MOÏSE Claudine. Relation à l'altérité ans le choix des prénoms des enfants de couples mixtes franco-arabes. 2007. <https://hal.archives-ouvertes.fr/hal-02498309/document>

BAHLOUL, Joëlle. Noms et prénoms juifs nord-africains. *Terrain*, 1985, 4, p. 62-69. <https://journals.openedition.org/terrain/2872>

BARBEROUSSE Anouk, HUNEMAN Philippe. Un « philosophe français » label rouge. Relecture tripodienne d'Alain Badiou. *Carnet Zilsel*, 1er avril 2016, <http://zilsel.hypotheses.org/2548>

BARRY III Herbert, HARPER Aylene S. Three Last Letters Identify Most Female First Names. *Psychological Reports*, 2000, 87(1), p. 48-54. <https://journals.sagepub.com/doi/abs/10.2466/pr0.2000.87.1.48?journalCode=prxa>

BARRY, Laurent. *La Parenté recomposée. Figures peul de l'alliance sur les hauts plateaux de l'Adamaoua (Nord Cameroun)*. Thèse de doctorat soutenue en 1996.



BAVOUZET Julia, WEISZ-PATRAULT Daniel. Revue critique de l'article "l'enracinement social de la mortalité routière". 2018. <https://hal.archives-ouvertes.fr/hal-01881517/document>

BEAUCHEMIN Cris, BRILHAULT Gwennaëlle, MORISSET Amandine, SIMON Patrick, GALONNIER Juliette, LE COURANT Stefan, NOÛS Camille. Le travail des données. Entretien autour du service des enquêtes de l'INED. *Tracés. Revue de sciences humaines*, 2019, p. 163-181. <https://journals.openedition.org/traces/10918>

BEAUCHEMIN Cris, HAMEL Christelle, SIMON Patrick (dir.). *Trajectoires et origines. Enquête sur la diversité des populations en France*. Paris : INED éditions, 2016. <https://www.ined.fr/fr/publications/editions/grandes-enquetes/trajec-toires-et-origines/>

BEAUCHEMIN Cris, LHOMMEAU Bertrand, SIMON Patrick. Histoires migratoires et profils socioéconomiques. *Trajectoires et origines. Enquête sur la diversité des populations en France* / dir. par Cris BEAUCHEMIN, Christelle HAMEL et Patrick SIMON. Paris: INED éditions, 2016, p. 33-60. Annexes [ici](#).

BEESTON, A. F. L. *Arabic Nomenclature: A summary guide for beginners*. Oxford: Oxford University Press, 1971.

BEHAGHEL Luc, CRÉPON Bruno, LE BARBANCHON Thomas. Unintended Effects of Anonymous Résumés. *American Economic Journal: Applied Economics*, 2015, 7(3), p. 1-27. <https://pubs.aeaweb.org/doi/pdfplus/10.1257/app.20140185>

BEIDER, Alexander. Perspectives of Jewish Onomastics. 2014. <http://www.iiig.org/wp-content/uploads/2014/02/beider.pdf>

BEIDER, Alexander. Discontinuity of Jewish Naming Traditions. *Avotaynu*, 2012, XXVIII(2), p. 43-53. [http://www.academia.edu/download/48250186/Avotaynu\\_2012\\_Summer\\_DISCONTINUITY.pdf](http://www.academia.edu/download/48250186/Avotaynu_2012_Summer_DISCONTINUITY.pdf)

BESNARD Philippe, GRANGE Cyril. La fin de la diffusion verticale des goûts ? (Prénoms de l'élite et du vulgum). *L'Année sociologique*, 1993, 43, p. 269-294. <https://www.jstor.org/stable/27890201>

BESNARD, Philippe. Le choix d'un prénom. Actualité de la méthode durkheimienne. *Recherches sociologiques*, 1991, 3, p. 53-60. <https://sharepoint.uclouvain.be/sites/ras/Revue/1991-XXII-3.pdf#page=54>

BLÉVIS, Laure. *Sociologie d'un droit colonial : citoyenneté et nationalité en Algérie (1865-1947) : une exception républicaine ?* Thèse de doctorat en science politique soutenue en 2004.

BLÉVIS, Laure. Les avatars de la citoyenneté en Algérie coloniale ou les paradoxes d'une catégorisation. *Droit et société*, 2001, 48(2), p. 557-581. <https://www.cairn.info/revue-droit-et-societe1-2001-2-page-557.htm>

BONNET, Doris. Le sens avoué du prénom en Afrique subsaharienne. *Vues d'enfance*, 1996, 5/6, p. 39-43.

BOUQUET, Olivier. *Onomasticon Ottomonicum* : identification administrative et désignation sociale dans l'État ottoman du XIX<sup>e</sup> siècle. *Revue des mondes musulmans et de la Méditerranée*, 2010, 127, p. 213-235. <https://journals.openedition.org/remmm/6743>

BOUQUET Olivier, FLICHE Benoît. La réforme des noms propres en Turquie : introduction. *Revue d'histoire moderne et contemporaine*, 2013, 60(2), p. 7-17. <https://www.cairn.info/revue-d-histoire-moderne-et-contemporaine-2013-2-page-7.htm>

BOURIN, Monique (dir.). *Genèse médiévale de l'anthroponymie moderne. Tome I*. Tours : Presses universitaires François-Rabelais, 1990.

BOURIN Monique, CHAREILLE Pascal (dir.). *Genèse médiévale de l'anthroponymie moderne. Tome II-1 : Persistances du nom unique. Le cas de la Bretagne. L'anthroponymie des clercs*. Tours : Presses universitaires François-Rabelais, 1992.

BOURIN Monique, CHAREILLE Pascal (dir.). *Genèse médiévale de l'anthroponymie moderne. Tome II-2 : Persistances du nom unique. Désignation et anthroponymie des femmes. Méthodes statistiques pour l'anthroponymie*. Tours : Presses universitaires François-Rabelais, 1992.

BOURIN Monique, CHAREILLE Pascal (dir.). *Genèse médiévale de l'anthroponymie moderne. Tome III : Enquêtes généalogiques et données prosopographiques*. Tours : Presses universitaires François-Rabelais, 1995.

BECK, Patrice (dir.). *Genèse médiévale de l'anthroponymie moderne. Tome IV : Discours sur le nom. Normes, usages, imaginaire (VIe-XVIe siècles)*. Tours : Presses universitaires François-Rabelais, 1997.

BOURIN Monique, CHAREILLE Pascal (dir.). *Genèse médiévale de l'anthroponymie moderne. Tome V-1 : Intégration et exclusion sociale, lectures anthroponymiques. Serfs et dépendants au Moyen Âge (VIIIe-XIIe siècle)*. Tours : Presses universitaires François-Rabelais, 2002.

BOURIN Monique, CHAREILLE Pascal (dir.). *Genèse médiévale de l'anthroponymie moderne. Tome V-2 : Intégration et exclusion sociale, lectures anthroponymiques. Serfs et dépendants au Moyen Âge (Le « nouveau servage »)*. Tours: Presses universitaires François-Rabelais, 2002.

BOYD-BOWMAN, Peter. Los nombres de pila en México desde 1540 hasta 1950. *Nueva Revista de Filología Hispánica*, 1970, 19(1), p. 12-48. <https://nrfh.colmex.mx/index.php/nrfh/article/view/437/437>

BOYER, Denise. La prénomination en France du XVII<sup>e</sup> siècle à nos jours : aspects diachroniques, diatopiques, diastratiques. *Onomastica*, 2016, p. 11-54. <https://www.raco.cat/index.php/Onomastica/article/download/369717/463465/>

BOZON, Michel. Histoire et sociologie d'un bien symbolique, le prénom. *Population*, 1987, 42(1), p. 83-98. [https://www.persee.fr/docAsPDF/pop\\_0032-4663\\_1987\\_num\\_42\\_1\\_16886.pdf](https://www.persee.fr/docAsPDF/pop_0032-4663_1987_num_42_1_16886.pdf)

BRAMWELL, Ellen S. Naming and transplanted traditions: Change and continuity in Glasgow's Pakistani Muslim community. *Onoma*, 2011, 46, p. 29-51. [http://poj.peeters-leuven.be/secure/POJ/downloadpdf.php?ticket\\_id=5c8f11e983c53](http://poj.peeters-leuven.be/secure/POJ/downloadpdf.php?ticket_id=5c8f11e983c53)

BRUNET, Étienne. Les prénoms et les noms de la Bible. Étude quantitative à travers quatre siècles de littérature. *Onomastique biblique* / dir. par Yannick LE BOULICAUT. Angers, 1988, p. 119-147. <https://hal.archives-ouvertes.fr/hal-01544825/document>

BRYLLA, Eva. Bynames and Nicknames. *The Oxford Handbook of Names and Naming* / dir. par Carole HOUGH. Oxford: Oxford University Press, 2016, p. 237-250.

BULLIET, Richard W. *Conversion to Islam in the Medieval Period. An Essay in Quantitative History*. Cambridge: Harvard University Press, 1979.

CARNEIRO Pedro, LEE Sokbae, REIS Hugo. Please call me John: Name choice and the assimilation of immigrants in the United States, 1900-1930. *Labour Economics*, 2020, 62. <https://www.sciencedirect.com/science/article/pii/S0927537119301149>

CHAREILLE, Pascal (dir.). *Genèse médiévale de l'anthroponymie moderne. Tome VI : Le nom, histoire et statistiques. Quelles méthodes quantitatives pour une étude de l'anthroponymie médiévale ?* Tours : Presses universitaires François-Rabelais, 2008.

CHEBIRA, Abderrazak. *L'indispensable manuel orthographique des prénoms français d'origine arabe et musulmane*. Alger: APIC, 2005.

CHOO Sungjae, KANG Peter, LAUDER Multamia R. M. T. Asian Onomastics — An introduction. *Onoma*, 2016, 51, p. 7-12. <https://onomajournal.org/wp-content/uploads/2019/09/Onoma-51-01.-Introduction-final-25.08.2019.pdf>

CIOFFARI, Gerardo. Saint Nicolas en Espagne, sur le chemin de Saint-Jacques et dans la Reconquista. *En Orient et en Occident : le culte de Saint Nicolas en Europe (X<sup>e</sup>-XXI<sup>e</sup> siècle)*

/ dir. par Véronique GAZEAU, Catherine GUYON et Catherine VINCENT. Paris : Cerf, 2015, p. 169-191.

COLLET, Beate. Prénommer son enfant dans les couples mixtes. Stratégies d'ajustements interculturels et logiques de genre. *Recherches familiales*, 2019, 16(1), p. 155-167. <https://www.cairn.info/revue-recherches-familiales-2019-1-page-155.htm>

COLONNA Fanny, TARAUD Christelle. La minorité européenne d'Algérie (1830-1962) : inégalités entre « nationalités », résistances à la francisation et conséquences sur les relations avec la « majorité musulmane ». Colloque *Pour une histoire critique et citoyenne. Le cas de l'histoire franco-algérienne*, 2006. [http://colloque-algerie.ens-lyon.fr/communication.php3?id\\_article=249](http://colloque-algerie.ens-lyon.fr/communication.php3?id_article=249)

CORPART, Isabelle. L'étrange nom de l'étranger. *Revue juridique personnes et famille*, 2019, 12.

CORVOL, Pierre. Garantir l'intégrité scientifique : la science face à la défiance. 2019. <https://www.canalacademie.com/ida12115-garantir-integrite-scientifique-la-science-face-a-la-defiance.html>

COULMONT, Baptiste. Pouvoir nommer. Congrès de l'Association française de sociologie à Amiens. 2017, p. 1-8. a

COULMONT, Baptiste. From Fatima to Lina. Congrès de la Société internationale de sociologie des religions à Lausanne. 2017, p. 1-8. b

COULMONT, Baptiste. Fashion Backwards: Assimilationist Naming and Temporal Lag. Conférence du European Consortium for Sociological Research à Milan. 2017, p. 1-14. c

COULMONT, Baptiste. *Sociologie des prénoms, nouvelle édition*. Paris : La Découverte, 2014.

COUTO, Marie-Paule. *Les paradoxes de la citoyenneté. Le devenir des pieds-noirs en France métropolitaine*. Paris : Thèse de doctorat soutenue en 2014.

COUTO, Marie-Paule. L'intégration socio-économique des pieds-noirs en France métropolitaine : le lien de citoyenneté à l'épreuve. *Revue européenne des migrations internationales*, 2013, 29(3), p. 93-119. <https://journals.openedition.org/remi/6548>

CREAMER, Thomas B. I. Chinese Place and Personal Names. *Name studies: an international handbook of onomastics* / dir. par Ernst EICHLER, Gerold HILTY, Heinrich LÖFFLER, Hugo STEGER, Herbert Ernst WIEGAND. Berlin: De Gruyter, 1995, p. 906-912.

DE SINGLY, François. *Le questionnaire. 5<sup>e</sup> édition*. Paris : Colin, 2020.

DECHARNE Marie-Noëlle, LIEDTS Eric. Porter un prénom arabe ou musulman est-il discriminant dans l'enseignement supérieur ? Orientation et poursuite d'études dans l'enseignement supérieur du Nord-Pas de Calais. Observatoire régional des études supérieures - ORES, 2007. <https://www.crefe38.fr/IMG/pdf/pulnpdc.pdf>

DÉFENSEUR DES DROITS. *Inégalités d'accès aux droits et discriminations en France. Contributions des chercheurs à l'enquête du Défenseur des droits*. Paris : La documentation française, 2020. [https://juridique.defenseurdesdroits.fr/doc\\_num.php?explnum\\_id=19415](https://juridique.defenseurdesdroits.fr/doc_num.php?explnum_id=19415)

DENÉCHÈRE, Yves. *Des enfants venus de loin. Histoire de l'adoption internationale en France*. Paris : Colin, 2011.

DESPLANQUES, Guy. Les enfants de Michel et Martine Dupont s'appellent Nicolas et Céline. *Économie et statistique*, 1986, 184, p. 63-83. [https://www.persee.fr/docAsPDF/estat\\_0336-1454\\_1986\\_num\\_184\\_1\\_2421.pdf](https://www.persee.fr/docAsPDF/estat_0336-1454_1986_num_184_1_2421.pdf)

DUPÂQUIER Jacques, PÉLISSIER Pierre, RÉBAUDO Danièle. *Le temps des Jules. Les prénoms en France au XIX<sup>e</sup> siècle*. Paris : Éditions Christian, 1987.

DUPÂQUIER Jacques, BIDEAU Alain, DUCREUX Marie-Elizabeth. *Le prénom. Mode et histoire*. Paris : EHESS, 1984.

EICHLER Ernst, HILTY Gerold, LÖFFLER Heinrich, STEGER Hugo, WIEGAND Herbert Ernst (dir.). *Name studies: an international handbook of onomastics*. Berlin: De Gruyter, 1995.

EL KAROUI Hakim, JARDIN Antoine. *Un islam français est possible*. Paris : Institut Montaigne, 2016. <https://www.institutmontaigne.org/publications/un-islam-francais-est-possible>

ESCAFRÉ-DUBLET Angéline, KESZTENBAUM Lionel, SIMON Patrick. Quand le recensement comptait les Français musulmans. *Population et sociétés*, 2020, 538, p. 1-4. [https://www.ined.fr/fichier/s\\_rubrique/30717/583.population.societe.novembre.2020.recense-ment.musulmans.france.fr.pdf](https://www.ined.fr/fichier/s_rubrique/30717/583.population.societe.novembre.2020.recense-ment.musulmans.france.fr.pdf)

ESCAFRÉ-DUBLET Angéline, KESZTENBAUM Lionel, SIMON Patrick. La greffe coloniale en métropole. Les Français musulmans dans le recensement de 1954. *Sociétés contemporaines*, 2018, 110(2), p. 35-59. <https://www.cairn.info/journal-societes-contemporaines-2018-2-page-35.htm>

ESKOLA Ksenia, HÄMÄLÄINEN Lasse. Given names of Russian-speaking children born in Finland between 2000-2018. *Onoma*, 2019, 54, p. 197-217. <https://onomajournal.org/wp-content/uploads/2020/06/Onoma-54-1.10-Eskola-H%C3%A4mäläinen-final-1.pdf>

FABBIANO, Giulia. Mixité postcoloniale. Les unions de descendants d'émigrés algériens à l'épreuve de l'expérience migratoire parentale. *Diasporas. Histoire et sociétés*, 2009, 15, p. 97-110. [https://www.persee.fr/doc/diasp\\_1637-5823\\_2009\\_num\\_15\\_1\\_1202](https://www.persee.fr/doc/diasp_1637-5823_2009_num_15_1_1202)

FABRE, Paul. Les noms propres en France. *Name studies: an international handbook of onomastics* / dir. par Ernst EICHLER, Gerold HILTY, Heinrich LÖFFLER, Hugo STEGER, Herbert Ernst WIEGAND. Berlin: De Gruyter, 1995, p. 1099-1106.

FOURQUET Jérôme, MANTERNACH Sylvain. *L'archipel français. Naissance d'une nation multiple et divisée*. Paris : Seuil, 2019. a

FOURQUET Jérôme, MANTERNACH Sylvain. Cent ans d'immigration racontés par les prénoms. *Hérodote*, 2019, 174(3), p. 113-140. b <https://www.cairn.info/revue-herodote-2019-3-page-113.html>

FOURQUET, Jérôme. Le vote pied-noir : mythe ou réalité ? *IFOP Focus*, 2014, 107, p. 1-8. [https://www.ifop.com/wp-content/uploads/2018/03/697-1-document\\_file.pdf](https://www.ifop.com/wp-content/uploads/2018/03/697-1-document_file.pdf)

FRYER Roland G., LEVITT Steven D. The Causes and Consequences of Distinctively Black Names. *Quarterly Journal of Economics*, 2004, 99(3), p. 767-805. <https://scholar.harvard.edu/files/fryer/files/the-causes-and-consequences-of-distinctively-black-names.pdf>

GASPERONI, Michaël. Les noms de famille juifs à Rome au XVIII<sup>e</sup> siècle : le ghetto entre onomastique et histoire sociale. *Revue des études juives*, 2018, 177(1-2), p. 135-172. <https://societedesetudesjuives.org/article-rej/les-noms-de-familles-juifs-a-rome-au-xviii-siecle-le-ghetto-entre-onomastique-et-histoire-sociale/>

GASPERONI, Michaël. I cognomi degli ebrei marchigiani. *Marca / Marche*, 2014, 3, p. 69-80.

GAZEAU Véronique, GUYON Catherine, VINCENT Catherine (dir.). *En Orient et en Occident : le culte de Saint Nicolas en Europe (X<sup>e</sup>-XXI<sup>e</sup> siècle)*. Paris : Cerf, 2015

GAZEAU. Conclusion. *En Orient et en Occident : le culte de Saint Nicolas en Europe (X<sup>e</sup>-XXI<sup>e</sup> siècle)* / dir. par Véronique GAZEAU, Catherine GUYON et Catherine VINCENT. Paris: Cerf, 2015, p. 489-496.

GERHARDS Jürgen, HANS Silke. From Hasan to Herbert: Name-Giving Patterns of Immigrant Parents between Acculturation and Ethnic Maintenance. *American Journal of Sociology*, 2009, 114(4), p. 1102-1128. <https://www.journals.uchicago.edu/doi/abs/10.1086/595944>



GERRITZEN, Doreen. Naming Children in a Globalizing World. *Acta onomastica*, 2006, XLVII, p. 177-184. <https://pure.knaw.nl/ws/portalfiles/portal/466584/DGActaOnomastica.pdf>

GOMEZ-LE CHEVANTON, Corinne. *Naissance d'une population et construction des identités en contexte colonial. Les Européens en Algérie, 1830-1962. Guide des sources*. Paris, 2021.

GRANGAUD Isabelle, MICHEL Nicolas. Introduction. *Revue des mondes musulmans et de la Méditerranée*, 2010, 127, p. 13-27. <https://journals.openedition.org/remmm/6571>

GRANGE, Cyril. Nommer : enjeux symboliques, sociaux et politiques. *Annales de démographie historique*, 2016, 1, p. 7-11. a <https://www.cairn.info/revue-annales-de-demographie-historique-2016-1-page-5.htm>

GRANGE, Cyril. Choix du prénom et acculturation : l'exemple de la bourgeoisie juive parisienne, 1800-1920. *Annales de démographie historique*, 2016, 1, p. 65-95. b <https://www.cairn.info/revue-annales-de-demographie-historique-2016-1-page-65.htm>

HABIBI, Nader. Popularity of Islamic and Persian Names in Iran before and after the Islamic Revolution. *International Journal of Middle East Studies*, 1992, 24(2), p. 253-260. <https://www.cambridge.org/core/journals/international-journal-of-middle-east-studies/article/abs/popularity-of-islamic-and-persian-names-in-iran-before-and-after-the-islamic-revolution/6C238FEE0AAE0861E88CACBF3D66B1FE>

HANKS Patrick, PARKIN Harry. Family Names. *The Oxford Handbook of Names and Naming* / dir. par Carole HOUGH. Oxford: Oxford University Press, 2016, p. 214-236.

HASSOUN, Jean-Pierre. Le choix du prénom chez les Hmong au Laos puis en France. Diversité, complexification et processus d'individuation. *Revue française de sociologie*, 1995, 36(2), p. 241-271. [https://www.persee.fr/doc/AsPDF/rfsoc\\_0035-2969\\_1995\\_num\\_36\\_2\\_4402.pdf](https://www.persee.fr/doc/AsPDF/rfsoc_0035-2969_1995_num_36_2_4402.pdf)

HEAD Keith, MAYER Thierry. Detection of Local Interactions from the Spatial Pattern of Names in France. *Journal of Regional Science*, 2008, 48(1), p. 67-95. <https://hal.archives-ouvertes.fr/hal-00266554/document>

HEINICH, Nathalie. *Ce que le militantisme fait à la recherche*. Paris : Gallimard, 2021.

HÉRAN, François. L'intégration des immigrés : débats et constats. *La vie des idées*, 2020. <https://laviedesidees.fr/L-integration-des-immigres-debats-et-constats.html>

HÉRAN, François. *Avec l'immigration. Mesurer, débattre, agir*. Paris : La Découverte, 2017.

HÉRAN, François. La science par dérogation, ou comment l'enquête TeO a rempli sa mission. *Trajectoires et origines. Enquête sur la diversité des populations en France* / dir. par Cris BEAUCHEMIN, Christelle HAMEL et Patrick SIMON. Paris: INED Éditions, 2016, p. 7-16.

HÉRAN, François. Les mots de la démographie des origines à nos jours : une exploration numérique. *Population*, 2015, 70(3), p. 525-566. <https://www.cairn.info/revue-population-2015-3-page-525.htm>

HERBERT, Robert K. The Dynamics of Personal Names and Naming Practices in Africa. *Name studies: an international handbook of onomastics* / dir. par Ernst EICHLER, Gerold HILTY, Heinrich LÖFFLER, Hugo STEGER, Herbert Ernst WIEGAND. Berlin: De Gruyter, 1995, p. 1222-1227.

HOUGH, Carole (dir.). *The Oxford Handbook of Names and Naming*. Oxford: Oxford University Press, 2016.

HUMPERT Andreas, SCHNEIDERHEINZE Klaus. Linguistic analysis of personal names based on the "onomastic process". 2013.

HUMPERT Andreas, SCHNEIDERHEINZE Klaus. Stichprobenziehung für telefonische Zuwandererumfragen: Einsatzmöglichkeiten der Namenforschung. *ZUMA Nachrichten*, 2000, 24(47), 36-63. <https://nbn-resolving.org/urn:nbn:de:0168-ss0ar-208052>

INSEE. Guide d'utilisation et dictionnaire des codes TeO. Mars 2011. [https://teo1.site.ined.fr/fichier/s\\_rubrique/20242/guide.utilisation.et.dictionnaire.des.codes.teo.fr.pdf](https://teo1.site.ined.fr/fichier/s_rubrique/20242/guide.utilisation.et.dictionnaire.des.codes.teo.fr.pdf)

INSEE. Trajectoires et Origines. Questionnaire. 2008. [https://teo1.site.ined.fr/fichier/s\\_rubrique/20242/questionnaire.teo.fr.pdf](https://teo1.site.ined.fr/fichier/s_rubrique/20242/questionnaire.teo.fr.pdf)

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Nomes mais populares. 2010. <https://censo2010.ibge.gov.br/nomes/#/ranking>

JONES Russell, PHILLIPS Nigel G. Personal Names in Malaysia and Indonesia. *Name studies: an international handbook of onomastics* / dir. par Ernst EICHLER, Gerold HILTY, Heinrich LÖFFLER, Hugo STEGER, Herbert Ernst WIEGAND. Berlin: De Gruyter, 1995, p. 902-904.

JOUNIAUX, Léo. *Les 20 000 plus beaux prénoms du monde*. Paris : Hachette Pratique, 2001.

KATEB Kamel, BRUNET Guy. Les familles israélites d'Algérie au XIXe siècle : de l'indigénat à l'assimilation. *Familles en transformation. Quand les modes de construction familiale se réinventent* / dir. par Jean-Paul SANDERSON et Michel ORIS. Paris : Association internationale des démographes de langue française, 2018.

KATEB, Kamel. *Européens, « indigènes » et juifs en Algérie (1830-1962). Représentations et réalités des populations*. Paris : INED, 2001.

KUIPERS Joel C., ASKURI. Islamization and Identity in Indonesia: The Case of Arabic Names in Java. *Indonesia*, 2017, 103, p. 25-49. <http://digilib2.unisayogya.ac.id/xmlui/bitstream/handle/123456789/2567/repository%203.pdf?sequence=1&isAllowed=y>

KYDJIAN, Maïlys. Mémoires croisées. Retour sur l'expérience coloniale et la guerre d'indépendance à travers trois générations d'« Algériens », « Harkis » et « Pieds-Noirs ». 2016. <https://tel.archives-ouvertes.fr/tel-01871444/document>,

KOÇ, Claire. *Claire, le prénom de la honte*. Paris : Albin Michel, 2021.

LAFFORT, Bruno. *Les couples mixtes chez les enfants de l'immigration algérienne*. Paris: L'Harmattan, 2003.

LANDMAN, Shlomit. Onomastics in a Melting Pot Society with Common Roots: Israeli Jews in the Second Half of the Twentieth Century. *Annales de démographie historique*, 2016, 1, p. 131-149. <https://www.cairn.info/revue-annales-de-demographie-historique-2016-1-page-131.htm>

LARCHER, Pierre. Le nom propre dans la tradition grammaticale arabe. *Les non-dits du nom. Onomastique et documents en terre d'Islam. Mélanges offerts à Jacqueline Sublet* / dir. par Christian MÜLLER et Muriel ROILAND-ROUABAH. Beyrouth: Presses de l'Ifpo, 2013, p. 303-318. <https://books.openedition.org/ifpo/5722?lang=fr>

LAWSON, Edwin D. Personal Naming Systems. *The Oxford Handbook of Names and Naming* / dir. par Carole HOUGH. Oxford: Oxford University Press, 2016, p. 167-198.

LAWSON, Edwin D. Personal Name Stereotypes. *Name studies: an international handbook of onomastics* / dir. par Ernst EICHLER, Gerold HILTY, Heinrich LÖFFLER, Hugo STEGER, Herbert Ernst WIEGAND. Berlin: De Gruyter, 1995, p. 1744-1747.

LE MINEZ, Sylvie. Oui, la statistique publique produit des statistiques ethniques. Panorama d'une pratique ancienne, encadrée et évolutive. *Le blog de l'Insee*, 2020. <https://blog.insee.fr/statistique-publique-produit-des-statistiques-ethniques/>

LEIBRING, Katharina. Given Names in European Naming Systems. *The Oxford Handbook of Names and Naming* / dir. par Carole HOUGH. Oxford: Oxford University Press, 2016, p. 199-213.

LIEBAU Elisabeth, HUMPERT Andreas, SCHNEIDERHEINZE Klaus. Wie gut funktioniert das Onomastik-Verfahren? Ein Test am Beispiel des SOEP-Datensatzes,



SOEPpapers on Multidisciplinary Panel Data Research, 2018, No. 976, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin. <http://hdl.handle.net/10419/181235>.

LLEWELLYN, Aisyah. What's in a name in Indonesia? *Asia Times*, 2018. <https://asiatimes.com/2018/02/whats-name-indonesia/>

MACKENZIE, Laurel. What's in a name? Teaching linguistics using onomastic data. *Language*, 2018, 94(4), p. 293-310. [http://laurelmackenzie.com/papers/MacKenzie\\_2018\\_Language.pdf](http://laurelmackenzie.com/papers/MacKenzie_2018_Language.pdf)

MAISONNEUVE, Hervé. La rétractation des publications scientifiques. Association française pour l'information scientifique. 2017. <https://www.pseudo-sciences.org/La-retractation-des-publications-scientifiques>

MAISONNEUVE Hervé, DECULLIER Evelyne. Explaining Retractions. *The Scientist*, 2015. <https://www.the-scientist.com/critic-at-large/explaining-retractions-34428>

MASSICOT, Simone. Effets sur la nationalité française de l'accession à l'indépendance de territoires ayant été sous la souveraineté française. *Population*, 1986, 41(3), p. 533-546. [https://www.persee.fr/docAsPDF/pop\\_0032-4663\\_1986\\_num\\_41\\_3\\_17644.pdf](https://www.persee.fr/docAsPDF/pop_0032-4663_1986_num_41_3_17644.pdf)

MATEOS Pablo, TUCKER Ken. Forenames and Surnames in Spain in 2004. *Names. A Journal of Onomastics*, 2008, 56(3), p. 165-184. <https://discovery.ucl.ac.uk/id/eprint/43541/1/43541.pdf>

MAZIÈRES Antoine, ROTH Camille. Large-Scale Diversity Estimation Through Surname Origin Inference. *Bulletin de méthodologie sociologique*, 2018, 139, p. 59-73. <https://hal.archives-ouvertes.fr/hal-01766665/document>

MENSAH Eyo, ROWAN Kirsty. African anthroponyms: Sociolinguistic currents and anthropological reflections. *Sociolinguistic Studies*, 2019, 13(2/4), p. 157-170. <https://journals.equinoxpub.com/SS/article/download/37819/37703>

MIGNOT, Jean-François. L'écart d'âge entre conjoints. *Revue française de sociologie*, 2010, 51(2), p. 281-320. <https://www.cairn.info/revue-francaise-de-sociologie-1-2010-2-page-281.htm>

MOREAU, Marie-Louise. Le marquage des identités ethniques dans le choix des prénoms en Casamance (Sénégal). *Cahiers d'études africaines*, 2001, 163-164, p. 541-556. <http://journals.openedition.org/etudesaficaines/108>

MULON, Marianne. La recherche onomastique en France. *Name studies: an international handbook of onomastics* / dir. par Ernst EICHLER, Gerold HILTY, Heinrich LÖFFLER, Hugo STEGER, Herbert Ernst WIEGAND. Berlin: De Gruyter, 1995, p. 148-152.

NICOLAISEN, Wilhelm FH. Onomastics. *International Encyclopedia of the Social and Behavioral Sciences, Second Edition* / ed. par Neil J. SMELSER, Paul B. BALTES. Amsterdam: Elsevier, 2015, vol. 17, p. 223-226.

ODASSO, Laura. *Mixités conjugales. Discrédits, résistances et créativité dans les familles avec un partenaire arabe*. Rennes : Presses universitaires de Rennes, 2016.

OFFICE FOR NATIONAL STATISTICS. Baby Names in England and Wales: 2019. 2020.

<https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/bulletins/babynamesenglandandwales/2019>

OFFICE FOR NATIONAL STATISTICS. Top 100 baby names in England and Wales: historical data. 2014.

<https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/datasets/babynamesenglandandwalestop100babynameshistoricaldata>

OLIVIERI, Lucia M. M. Saint Nicolas et l'Irlande. *En Orient et en Occident : le culte de Saint Nicolas en Europe (X<sup>e</sup>-XXI<sup>e</sup> siècle)* / dir. par Véronique GAZEAU, Catherine GUYON et Catherine VINCENT. Paris : Cerf, 2015, p. 311-321.

PARIETTI Giuseppe, TOURNEUX Henry. *Dictionnaire fulfulde-français / français-fulfulde (Dialect[e] peul [du] Diamaré, Cameroun)*. Pessano con Bornago : Mimep-Docete, 2018.

PIKETTY, Thomas. *Capital et idéologie*. Paris : Seuil, 2019.

PINA-CABRAL, João de. Names and Naming. *International Encyclopedia of the Social and Behavioral Sciences, Second Edition* / ed. par Neil J. SMELSER, Paul B. BALTES. Amsterdam: Elsevier, 2015, vol. 16, p. 183-187.

PITCHER Benjamin J., MESOUDI Alex, MCELLIGOTT Alan G. Sex-Biased Sound Symbolism in English-Language First Names. *PLoS ONE*, 2013, 8(6), p. 1-6. <https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0064825&type=printable>

POUZAT Christophe, LEGRAND Arnaud, HINSEN Konrad. Recherche reproductible : principes méthodologiques pour une science transparente. 2020. <https://www.fun-mooc.fr/courses/course-v1:inria+41016+self-paced/about>

PUZENAT, Amélie. Le vécu de la mixité conjugale chez les couples franco-maghrébins et la transmission identitaire aux enfants. *Diversité urbaine*, 2008, 8(1), p. 113-128. <https://www.erudit.org/fr/revues/du/2008-v8-n1-du2347/018619ar/>

QUINON Manuel, SAINT-MARTIN Arnaud. Le maffesolisme, une « sociologie » en roue libre. Démonstration par l'absurde. *Carnet Zilsel*, 7 mars 2015, <http://zilsel.hypotheses.org/1713>

RAHMAN, Tariq. Personal Names of Pakistani Muslims: An Essay on Onomastics. *Pakistan Perspectives*, 2013, 18(1), p. 33-57. [https://www.researchgate.net/publication/274570421\\_Personal\\_names\\_of\\_Pakistani\\_Muslims\\_an\\_essay\\_on\\_Onomastics](https://www.researchgate.net/publication/274570421_Personal_names_of_Pakistani_Muslims_an_essay_on_Onomastics)

RAHMAN, Tariq. Names as Traps: Onomastic Destigmatization Strategies in Pakistan. *Pakistan Perspectives*, 2014, 19(1), p. 9-25. [https://www.researchgate.net/publication/274570292\\_NAMES\\_AS\\_TRAPS\\_Onomastic\\_Destigmatization\\_Strategies\\_in\\_Pakistan](https://www.researchgate.net/publication/274570292_NAMES_AS_TRAPS_Onomastic_Destigmatization_Strategies_in_Pakistan)

REVEL, Jean-François. *La connaissance inutile*. Paris : Grasset, 1988.

ROBUSTOVA, Veronika V. Anthroponymic system of Malaysia: name popularity and culture shift. *Onoma*, 2016, 51, p. 207-221. <https://onomajournal.org/anthroponymic-system-of-malaysia-name-popularity-and-culture-shift/>

SAARELMA, Minna. African and European Anthroponymic Systems: Similarities and Differences. *Onoma*, 2009, 44, p. 191-217.

SABUNCU Elifsu, COULMONT Baptiste. Des prénoms à la mode en Turquie ? 2011. <http://coulmont.com/blog/2011/09/20/prenoms-mode-turquie/>

SIDHU David M., PEXMAN Penny M. What's in a Name? Sound Symbolism and Gender in First Names. *PLoS ONE*, 2015, 10(5), p. 1-22. <https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0126809&type=printable>

SIMON Patrick, TIBERJ Vincent. Sécularisation ou regain religieux : la religiosité des immigrés et de leurs descendants. *Trajectoires et origines. Enquête sur la diversité des populations en France* / dir. par Cris BEAUCHEMIN, Christelle HAMEL et Patrick SIMON. Paris : INED éditions, 2016, p. 559-582. [https://www.ined.fr/fichier/s\\_rubrique/19585/document\\_travail\\_2013\\_196\\_religion.fr.pdf](https://www.ined.fr/fichier/s_rubrique/19585/document_travail_2013_196_religion.fr.pdf)

SIMON, Patrick. Nationalité et origine dans la statistique française. Les catégories ambiguës. *Population*, 1998, 53(3), p. 541-567. <https://www.cairn.info/revue-population-1998-3-page-541.htm>

SIMON, Patrick. Les statistiques, les sciences sociales françaises et les rapports sociaux ethniques et de « race ». *Revue française de sociologie*, 2008, 49(1), p. 153-162. <https://www.cairn.info/revue-francaise-de-sociologie-1-2008-1-page-153.htm>

SOCIAL SECURITY. Popular baby names. 2020. <https://www.ssa.gov/oact/babynames/index.html>

STREIFF-FÉNART, Jocelyne. La nomination de l'enfant dans les familles franco-maghrébines. *Sociétés contemporaines*, 1990, 4, p. 5-18. [https://www.persee.fr/doc/socco\\_1150-1944\\_1990\\_num\\_4\\_1\\_972](https://www.persee.fr/doc/socco_1150-1944_1990_num_4_1_972)

STRIPE, Nick. What's in a name? There's more than you might think to the ONS baby names list. 2019. <https://blog.ons.gov.uk/2019/08/28/whats-in-a-name-theres-more-than-you-might-think-to-the-ons-baby-names-list/>

SUE Christina A., TELLES Edward E. Assimilation and Gender in Naming. *American Journal of Sociology*, 2007, 112(5), p. 1383-1415. <https://www.journals.uchicago.edu/doi/abs/10.1086/511801>

SUSEWIND, Raphael. What's in a name? Probabilistic inference of religious community from South Asian names. *Field Methods*, 2015, 27(4), p. 319-332. <https://journals.sagepub.com/doi/10.1177/1525822X14564275>

TAGGART, Emma. Fascinating Map Reveals the Most Common Surnames in Every Country. *My Modern Met*, 2020. <https://mymodernmet.com/most-common-surnames-map/>

THURSTON, Alex. Muslim First Names in West Africa. *Sahel Blog*, 2013. <https://sahelblog.wordpress.com/2013/03/12/muslim-first-names-in-west-africa/>

TILMATINE, Mohand. Arabization and linguistic domination: Berber and Arabic in the North of Africa. *Language Empires in Comparative Perspective* / dir. par Christel STOLZ. Berlin: De Gruyter, 2015, p. 1-16.

TRIBALAT, Michèle. *Assimilation. La fin du modèle français*. Paris : Éditions du Toucan, 2013.

TRIBALAT, Michèle. La diversité des populations en France. *Commentaire*, 2016, 155(3), p. 698-703. <https://www.commentaire.fr/boutique/achat-d-articles/la-diversite-des-populations-en-france-11076>

TRIBALAT Michèle, SIMON Patrick, RIANDEY Benoît. *De l'immigration à l'assimilation. Enquête sur les populations d'origine étrangère en France*. Paris : INED / La Découverte, 1996.

TUPPAT Julia, GERHARDS Jürgen. Immigrants' First Names and Perceived Discrimination: A Contribution to Understanding the Integration Paradox. *European Sociological Review*, 2021, 37(1), p. 121-136. <https://academic.oup.com/esr/article-abstract/37/1/121/5907976?redirectedFrom=fulltext>

TURKSTAT. Statistics on Child. 2019. <https://turkstatweb.tuik.gov.tr/PreHaberBultenleri.do?id=33733>

TWENGE Jean M., ABEBE Emodish M., CAMPBELL W. Keith. Fitting In or Standing Out: Trends in American Parents' Choices for Children's Names, 1880-2007. *Social Psychological and Personality Science*, 2010, 19(1), p. 19-25. <https://journals.sagepub.com/doi/10.1177/1948550609349515>

VALETAS Marie-France, BRINGÉ Arnaud. Prénoms des enfants d'immigrés en France : une pratique différente selon le sexe ? *Du genre et de l'Afrique. Hommage à Thérèse Locoh* / dir. par Jacques VALLIN. Paris : INED, 2009, p. 57-65.

VARRO, Gabrielle. Analyse de contenu et analyse de discours : à propos du prénom. *Sociétés contemporaines*, 1994, 18/19, p. 121-144. [https://www.persee.fr/doc/socco\\_1150-1944\\_1994\\_num\\_18\\_1\\_1167](https://www.persee.fr/doc/socco_1150-1944_1994_num_18_1_1167)

VARRO, Gabrielle. Enfants et adolescents mixtes : une identité spécifique ? *Enfance*, 1990, 43(3), p. 303-322. [https://www.persee.fr/doc/enfan\\_0013-7545\\_1990\\_num\\_43\\_3\\_1950](https://www.persee.fr/doc/enfan_0013-7545_1990_num_43_3_1950)

VARRO Gabrielle, LESBET Djaffar. Le prénom révélateur. *Généralisations issues de l'immigration : « Mémoires et devenir »* / dir. par Georges ABOU-SADA et Hélène MILET. Paris: Arcantère Éditions, 1986, p. 139-153.

WHISSELL, Cynthia. Sound and Emotion in Given Names. *Names. A Journal of Onomastics*, 2001, 49(2), p. 97-120.  
<https://www.tandfonline.com/doi/abs/10.1179/nam.2001.49.2.97?journalCode=ynam20>

YERMÈCHE, Ouerdia. L'anthroponymie algérienne : entre rupture et continuité ? *Le nom propre maghrébin de l'homme, de l'habitat, du relief et de l'eau* / dir. par Ouerdia YERMÈCHE et Farid BENRAMDANE. Oran: CRASC, 2013, p. 47-62.  
<https://ouvrages.crasc.dz/pdfs/2013-nom-propre-maghrebin-fr-yermeche-sadat-ouerdia.pdf>

ZGUSTA, Ladislav. Name. *Encyclopedia Britannica*. 2014.  
<https://www.britannica.com/topic/name>

ZGUSTA, Ladislav. Names and their Study. *Name studies: an international handbook of onomastics* / dir. par Ernst EICHLER, Gerold HILTY, Heinrich LÖFFLER, Hugo STEGER, Herbert Ernst WIEGAND. Berlin: De Gruyter, 1995, p. 1876-1890.

## **Annexe 1A : Copie de l'article reproduit**





# Population & Sociétés

English  
version

## Quels prénoms les immigrés donnent-ils à leurs enfants en France ?

Baptiste Coulmont\* et Patrick Simon\*\*

Les prénoms donnés par les immigrés à leurs enfants, et ceux que reçoivent à la génération suivante leurs petits-enfants, sont-ils puisés dans le registre des prénoms des pays d'origine, ou sont-ils au contraire similaires à ceux donnés par la population majoritaire ? Analysant les données de l'enquête *Trajectoires et Origines*, Baptiste Coulmont et Patrick Simon expliquent que la situation varie selon les régions et cultures d'origine.

Le choix d'un prénom est affaire de goût parental, d'enorme sociale et de références culturelles [1]. Mais qu'en est-il pour les immigrés<sup>(1)</sup> et leurs enfants ? Dans leur cas, le choix d'un prénom est un marqueur culturel. Beaucoup d'études interprètent comme une mesure de l'assimilation le degré de proximité de ces prénoms avec ceux de la population majoritaire (les personnes qui n'ont pas d'ascendance immigrée sur deux générations) [2]. Il peut aussi refléter un souci d'épargner des difficultés futures à l'enfant au cours de sa vie. Les enquêtes de *fasting* ont bien montré l'implication négative que pouvaient produire des prénoms maghébiniens, africains ou asiatiques sur des recruteurs, des agences immobilières ou des banques [3]. L'attribution d'un prénom « majoritaire » assurerait alors une invisibilité partielle des descendants d'immigrés et pourrait les protéger de certaines discriminations. Examinons ici comment se transmettent les prénoms dans les familles d'immigrés de différentes origines vivant en France en analysant les données de l'enquête *Trajectoires et Origines* (TeO) (voir encadré).

### Dis-moi comment tu t'appelles, je te dirai d'où tu viens

La correspondance entre le prénom et le contexte culturel de naissance a longtemps été la règle. Cette correspondance est assurée non seulement par les usages

sociaux, mais également par des politiques d'unification onomastique dans de nombreux pays. En France, les règles administratives en vigueur jusqu'au début des années 1990 favorisaient, lors de la déclaration des nouveau-nés à l'état civil, la sélection de prénoms « en usage dans les différents calendriers ». Cette règle a été levée et le choix des prénoms s'est diversifié pour la population majoritaire comme pour les immigrés.

Cette ouverture se traduit par un élargissement des choix de prénoms donnés aux nouveau-nés après 1990 pour la population majoritaire, et donc à une diminution des prénoms « français » (voir encadré). On peut la mesurer tout d'abord par le nombre de prénoms attribués : 80 prénoms couvraient la moitié des naissances en 1986, il en faut plus de 100 en 1998 et plus de 200 en 2008 [7]. Elle se traduit également par l'intégration de prénoms « exotiques », extérieurs au répertoire usuel, soit qu'il s'agisse de prénoms inventés ou tirés de la littérature, de films, de séries, ou qu'ils soient repris de prénoms en vigueur dans d'autres pays, soit par l'influence de l'immigration en France, soit enfin par le choix de prénoms de stars internationales (acteurs et actrices, sportifs et sportives, hommes et femmes politiques, etc.) (figure 1). Cependant, cette « exotisation » est relativement sélective et ne concerne pas n'importe quels prénoms, ou du moins ne puise pas dans tous les registres culturels. En 2005, 50 % des enfants de la population majoritaire ont reçu un prénom qui n'est pas typiquement « français » (figure 1), mais cette diversification ne se traduit pas pour autant par une

\* Université Paris 8 et CRESPA

\*\* Institut national d'études démographiques et Institut des Migrations

(1) Personnes nées étrangères à l'étranger et résidents en France

## Quels prénoms les immigrés donnent-ils à leurs enfants en France ?

### Encadré. Source et méthodes

#### L'enquête *Trajectoires et Origines* (TeO) (2008, Ined-Insee)

L'enquête *Trajectoires et Origines* (TeO), réalisée en 2008-2009 par l'Ined et l'Insee, vise à décrire et analyser les conditions de vie et les trajectoires sociales des individus en fonction de leurs origines sociales et de leur lien à la migration. Elle a interrogé un échantillon de 22 000 personnes en France métropolitaine. Elle a recueilli les prénoms de tous les membres du ménage ainsi que ceux de leurs enfants non résidents. Il est ainsi possible d'étudier trois générations : celle des immigrés qui, nés étrangers à l'étranger, ont été nommés selon les normes des sociétés d'origine ; celle des descendants d'immigrés (seconde génération ou G2) qui, nés en France d'un ou de deux parents immigrés, reçoivent des prénoms selon des déterminants combinant les héritages familiaux (sociaux, culturels et religieux) et les normes sociales de leur milieu ; celle des petits-enfants d'immigrés (G3) qui n'ont pas été directement enquêtés mais sont identifiés à travers les choix de prénoms de leur parents, descendants d'immigrés.

#### Le classement des prénoms

Il y a plusieurs façons de classer les prénoms selon la culture à laquelle ils se réfèrent. La première suppose l'existence de listes de prénoms par culture, et on procède alors par classement manuel ou automatique (voir la référence [4] pour une étude en Allemagne et la référence [5] pour une recherche systématique des prénoms d'origine arabe ou musulmane enregistrés dans l'état civil en France). Une deuxième méthode consiste à relier chaque prénom à une catégorie linguistique et culturelle à partir d'une classification probabiliste des référents linguistiques des syllabes. Cette méthode automatique peut également être appliquée aux noms de famille, qui sont plus stables que les prénoms [6]. Le fait que des schémas linguistiques soient spécifiques à des aires culturelles se retrouve dans les prénoms et autorise cette attribution. Une troisième approche consiste à qualifier les prénoms en fonction de leur fréquence dans les groupes de population étudiés, ceci indépendamment de leurs signifiants linguistiques ou culturels. Un indice est construit en le rapportant à sa diffusion relative dans la population n'ayant pas d'ascendance migratoire sur deux générations [2].

Nous utilisons ici une classification linguistique probabiliste qui catégorise les prénoms en « groupes culturels » :

**Français** : prénoms dont les consonances sont caractéristiques des schémas linguistiques français. Ce ne sont pas nécessairement les prénoms portés par des personnes de nationalité ou d'origine française.

**Arabo-musulman** : prénoms circulant dans le monde arabe (Maghreb et Moyen-Orient) ou de tradition musulmane. Y compris leur déclinaison turques.

**Latin** : prénoms donnés en Europe du Sud (principalement en Italie, Espagne et Portugal) et dans le monde hispanophone et lusophone.

**Africain** : ensemble des prénoms des aires linguistiques au sud du Sahara, et pour cette raison relativement hétérogène.

**Asiatique** : ensemble des aires linguistiques du continent asiatique, principalement Asie du Sud-Est, sous-continent indien et Chine.

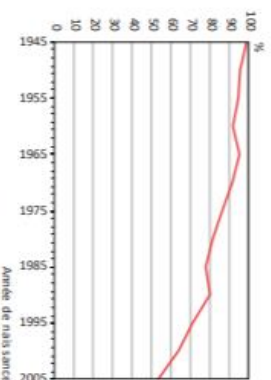
**Internationaux ou anglophones** : prénoms donnés dans les aires anglophones et qui se sont diffusés dans l'ensemble des aires culturelles.

#### Autre groupe

**Inclassable** : lorsque l'analyse des syllabes n'a pas permis d'associer le prénom à une catégorie connue.

Nous associons aussi à chaque prénom un score basé sur le lieu de naissance des parents : ce score vaut 100 s'ils sont nés en France, 0 s'ils ne le sont pas. Ce score mesure la distance entre le lieu de naissance des parents et le lieu de naissance de l'enfant. Il est appelé « score de distance ».

Figure 1. Proportion de nouveau-nés recevant un prénom « français » \* par année de naissance – France, population majoritaire (%)



B. Coulinmont, P. Simon, *Population & Sociétés* n° 565, INED, avril 2019.

Champ : Enfants de la population majoritaire.

Source : Enquête *Trajectoires et Origines* (TeO), 2008-2009 (Ined et Insee).

Note \* : voir encadré.

convergence avec les prénoms portés par les immigrés ou leurs enfants. Ceci est mesurable par le score de distance (voir encadré) qui reste stable (autour de 75) tout au long des 60 dernières années. Si la population majoritaire sélectionne des prénoms dans des registres culturels variés (prénoms internationaux, prénoms latins...), elle ne pioche pas dans le stock de prénoms portés par les immigrés non européens des années 1990. Par exemple, la population majoritaire donne comme prénoms à ses enfants Enzo et Clara, mais pas Mohammed ou Farida.

Les immigrés arrivent en France avec un prénom souvent caractéristique de l'aire culturelle de leur pays d'origine. On observe ainsi une forte prédictibilité de l'origine : parmi les immigrés venus du Maghreb, 94 % portent un prénom classé comme arabo-musulman, 82 % des immigrés d'Afrique sahéloenne portent un prénom « africain » ou arabo-musulman, et 92 % des immigrés de Turquie ont un prénom turc. C'est moins le cas des immigrés originaires d'Asie du Sud-Est qui ne sont que 58 % à garder un prénom aux consonances asiatiques (35 % ont un prénom français) et les originaires d'Afrique centrale dont les deux tiers ont des prénoms français. Nous avons tendance spontanément à attribuer des origines à partir des prénoms. Mais l'attribution est-elle juste ? Un prénom arabo-musulman désigne effectivement dans 82 % des cas une personne originaire du Maghreb (immigré ou descendant d'immigrés), un prénom « africain » est porté dans 95 % des cas par un ou une originaire d'Afrique subsaharienne, mais un prénom latin ne concerne un originaire d'Europe du Sud que dans 56 % des cas. Ces prénoms latins sont largement repris par la population majoritaire, avec une nette différence de genre : 18 % des hommes porteurs d'un prénom latin sont de la population majoritaire, et 36 % des femmes.



Quels prénoms les immigrés donnent-ils à leurs enfants en France ?

Tableau. Top 3 des prénoms par sexe et générations, Europe du Sud ou Maghreb

	Europe du Sud		Maghreb	
	Homme	Femme	Homme	Femme
Immigrés	José	Maria	Mohamed	Fatima
	Antonio	Maria	Ahmed	Fatima
	Manuel	Ana	Rachid	Khadja
Descendants d'immigrés	Jean	Maria	Mohamed	Sarah
	David	Sandrine	Karim	Nadia
	Alexandre	Sandra	Mehdi	Myriam
Petits-enfants d'immigrés	Thomas	Laura	Yanis	Sarah
	Lucas	Léa	Nicolas	Inès
	Enzo	Camille	Mehdi	Lina

Source : Enquête Trajectoires et Origines (TeO), 2008-2009 (Ined et Insee).

### Le prénom comme marqueur de transformation culturelle

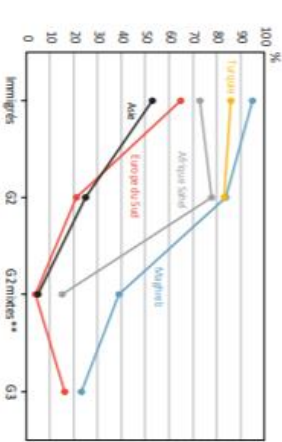
Les prénoms donnés par les immigrés à leurs enfants nés en France reflètent l'héritage culturel ainsi que les normes dominantes en France, sachant que ces dernières évoluent : s'appeler Lina, Mila ou Inès pour les filles, Adam, Liam ou Ethan pour les garçons, autant de prénoms dans le top 20 en 2017, est une marque d'adoption de goûts dominants. Pourtant, ces prénoms sont quasiment inexistant avant l'an 2000 et peuvent difficilement passer pour « typiquement français ».

La transmission des prénoms sur trois « générations » ne peut être étudiée avec l'enquête TeO que pour les Européens du Sud et les Africains du Nord en raison de la taille de l'échantillon. L'attribution de « prénoms spécifiques », c'est-à-dire de prénoms arabo-musulmans pour les personnes d'origine maghrébine, de prénoms latins pour celles originaires d'Europe du Sud (Espagne, Italie, Portugal), etc., suivent deux logiques différentes de transmission.

Environ 65 % des Européens du Sud arrivés en France avant 1980 portent des prénoms latins (Marta, José, Antonio)(tableau). Ces prénoms sont alors peu donnés en France. L'abandon de ces prénoms se fait dès la première génération née en France : seuls 16 % des descendants portent un prénom latin (figure 2). À la génération des petits-enfants, le score de distance montre que les prénoms portés ne se distinguent plus des prénoms de la population majoritaire. On remarque cependant un « retour » depuis le milieu des années 1990 aux prénoms latins (comme Enzo, ou Laura), mais ils ne renvoient quasiment plus alors un signal d'altérité.

Les immigrés du Maghreb arrivent en France avec des prénoms très éloignés de ceux de la population majoritaire (Mohamed, Fatima). La part des prénoms arabo-

Figure 2. Proportion de prénoms spécifiques\* par origine et génération (%)



B. Coulmont, P. Simon, Population & Sociétés n° 565, INED, avril 2019.

Source : Enquête Trajectoires et Origines (TeO), 2008-2009 (Ined et Insee).

Notes : \* voir encadré, \*\* enfants dont un seul parent est immigré et l'autre natif.

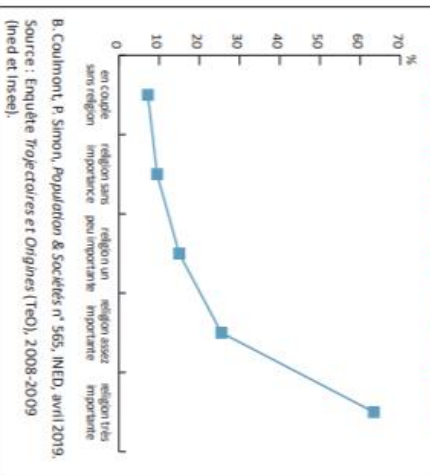
musulmans est supérieure à 90 %. À la génération suivante, encore près des deux tiers des enfants d'immigrés reçoivent un prénom arabo-musulman, mais leur registre culturel est plus ambigu (Nadia, Myriam). Les prénoms que reçoivent les petits-enfants sont, en 2008, proches de ceux que la population majoritaire donne à ses enfants. Là aussi, on remarque, comme pour les petits-enfants des Européens du Sud, le droit croissant de prénoms appartenant au registre des prénoms maintenant considérés comme communs (Inès, Sarah).

La trajectoire suivie par les originaires du Maghreb mène au même point d'arrivée que celle suivie par les Européens du Sud, mais de manière différée. Une partie de cette différence tient à la proximité des prénoms latins et français, et une autre à l'abandon plus fréquent des prénoms spécifiques par les couples mixtes. Les descendants dont les parents étaient en couple mixte, quelles que soient leurs origines, ont moins souvent un prénom spécifique (figure 2). Or, les descendants d'immigrés d'origine sud-européenne ont grandi dans une famille mixte dans 56 % des cas, ceux d'origine maghrébine dans 32 % des cas.

On peut aussi comparer le début des trajectoires suivies par les autres groupes. Les personnes d'origine turque ou d'Afrique sahélienne ajustent leurs comportements sur le modèle suivi par celles d'origine maghrébine, alors que les personnes d'origine asiatique collent à l'évolution des immigrés d'Europe du Sud et de leurs descendants. La poursuite de l'immigration en provenance du Maghreb, de Turquie et d'Afrique subsaharienne favorise la réintégration culturelle car une proportion significative des descendants d'immigrés sont en union avec des immigrés, ce qui est plus rarement le cas pour les personnes originaires d'Europe du Sud.

## Quels prénoms les immigrés donnent-ils à leurs enfants en France ?

**Figure 3. Proportion de prénoms arabo-musulmans chez les petits-enfants d'immigrés maghrébins selon la religiosité du parent**



### La religion a-t-elle une influence ?

Les musulmans gardent des prénoms spécifiques plus souvent que les personnes sans religion<sup>(2)</sup> et les chrétiens, et le degré de religiosité a une influence quelle que soit la religion. Si, en moyenne, 23 % des petits-enfants d'immigrés maghrébins portent un prénom arabo-musulman, la proportion varie selon la religiosité des parents (figure 3). Alors que des parents sans religion choisissent dans 7 % des cas un prénom arabo-musulman pour leur enfant, ce sont 63 % de ceux à forte religiosité (qui considèrent la religion comme très importante dans leur vie) qui le font. Ce n'est que dans ce cas que l'on constate une rétention culturelle à la troisième génération. On ne constate pas une dynamique similaire à la troisième génération pour les familles chrétiennes d'origine immigrée, car les prénoms qui leur seraient spécifiques sont communs avec ceux des familles majoritaires chrétiennes.

### Conclusion

La convergence entre population majoritaire et descendants d'immigrés ne se fait pas autour de prénoms typiquement « Français », mais de prénoms internationaux auxquels tous et toutes peuvent s'identifier. Il existe

pourant une transmission significative de prénoms associée à des origines maghrébines qui suivent deux registres différents. Tout d'abord une transmission dans la filiation des prénoms arabo-musulmans, en particulier dans les familles à forte religiosité ; mais également sous une forme d'innovation culturelle avec l'introduction de nouveaux prénoms (Yanis, Rayane ou Lina) qui sont rapidement perçus comme d'origine maghrébine par la population majoritaire. En d'autres termes, pour la population d'origine maghrébine, l'insémination des connotations culturelles des prénoms n'est pas complètement réalisée : Yanis n'est pas encore vu comme Enzo.

### Références

- [1] Bernard P. et Desplanches G., 1986, *Un prénom pour toujours : la cote des prénoms hier, aujourd'hui et demain*, Paris, Balland.
- [2] Goldstein J. et Stecklov G., 2015, « From Patrick to John F.: Ethnic names and occupational success in the last era of mass migration », *American Sociological Review*, 81(1), p. 85-106.
- [3] Feroni F., Ruault M. et Valat E., 2016, « Discrimination à l'embauche selon « l'origine » : que nous apprend le testing auprès des grandes entreprises ? », *Dares Analyses*, n° 76.
- [4] Gerhards J. et Hans S., 2009, « From Hasan to Herbert: Naming patterns of immigrant parents between acculturation and ethnic maintenance », *American Journal of Sociology*, 114(4), p. 1102-1128.
- [5] Decharne M.-N. et Liedts E., 2007, « Porter un prénom arabe ou musulman est-il discriminant dans l'enseignement supérieur ? Orientation et poursuite d'études dans l'enseignement supérieur du Nord-Pas de Calais », *Observatoire régional des études supérieures*, ORES Population.
- [6] Mazères A. et Roth C., 2018, « Large-scale diversity estimation through surname origin inference », *Bulletin of Sociological Methodology*, 139, p. 59-73.
- [7] Coulmont B., 2011, *Sociologie des prénoms*, Paris, La Découverte.

### Résumé

La convergence entre population majoritaire et descendants d'immigrés ne se fait pas autour de prénoms typiquement « Français », mais de prénoms internationaux auxquels tous et toutes peuvent s'identifier. Pour les enfants ou petits-enfants d'immigrés d'origine maghrébine, des prénoms arabo-musulmans continuent en partie à être donnés ainsi que de nouveaux prénoms (Yanis, Rayane ou Lina) perçus comme d'origine maghrébine par la population majoritaire.

### Mots-clés

Prénoms, immigrés, des descendants d'immigrés, France, Maghreb, Afrique, Europe du Sud.

(2) Personnes qui dans l'enquête ont déclaré ne pas avoir de religion.

## **Annexe 1B : L'enquête Trajectoires et origines (TeO)**

L'enquête TeO est une enquête aléatoire réalisée en 2008 par l'INED et l'INSEE auprès de 21 761 personnes d'âge actif vivant en ménage ordinaire en France métropolitaine en 2008. Elle distingue 5 sous-échantillons :

- i. les immigrés, nés étrangers à l'étranger entre 1948 et 1990 (âgés de 18 à 60 ans en 2008) ;
- ii. les personnes nées dans un département d'Outre-Mer (DOM) entre 1948 et 1990 (âgées de 18 à 60 ans) ;
- iii. les enfants d'au moins un parent immigré, c'est-à-dire nés en France métropolitaine entre 1958 et 1990 d'au moins une personne née étrangère à l'étranger (âgés de 18 à 50 ans) ;
- iv. les enfants d'au moins un parent né dans un DOM, c'est-à-dire nés en France métropolitaine entre 1958 et 1990 (âgés de 18 à 50 ans) ;
- v. les autres personnes, nées entre 1948 et 1990 (âgées de 18 à 60 ans), n'appartenant à aucun des groupes précédents, et aussi appelées population « majoritaire » ; ce groupe inclut des personnes qui ont aussi un lien à la migration vers la France métropolitaine : les Français nés français à l'étranger et leurs enfants, les rapatriés de l'ex-empire colonial français et leurs enfants, les personnes originaires des COM et leurs enfants, et les personnes qui ont une ascendance immigrée lointaine.

Pour plus de précisions, voir le [Guide d'utilisation et dictionnaire des codes TeO](#).



## Annexe 2 : Les reprises médiatiques de l'article

Cette Annexe 2, associée au Chapitre 2, p. 11, comprend une liste non exhaustive des reprises de presse de l'article de Coulmont et Simon (2019), entre avril et juillet 2019. Elle n'inclut pas les enseignements qui utilisent cet article, comme ce [cours de sociologie de la migration à Sciences Po Lille](#).

	Média et lien	Date	Titre
Presse nationale	<a href="#">Le Monde</a>	10/04	Yanis, Sarah ou Inès : quand l'intégration se lit dans les prénoms des petits-enfants d'immigrés
	<a href="#">Le Figaro</a>	10/04	Ce que les prénoms révèlent de l'intégration en France
	<a href="#">Libération</a>	10/04	Patrick Simon et Baptiste Coulmont : « On ne peut pas juger la volonté d'assimilation en ne se fondant que sur des prénoms »
	<a href="#">La Croix</a>	10/04	Des immigrés plus ouverts dans le choix des prénoms
	<a href="#">20 Minutes</a>	10/04	Quels prénoms les immigrés donnent-ils à leurs descendants ?
	<a href="#">L'Obs</a>	10/04	« Après deux générations, les enfants d'immigrés abandonnent les prénoms du pays de leurs parents »
	<a href="#">L'Express</a>	10/04	Les immigrés privilégient les prénoms "internationaux" pour leurs enfants
	<a href="#">Valeurs actuelles</a>	10/04	Inès, Mehdi... Quels prénoms portent les descendants d'immigrés ?
	<a href="#">Sciences Humaines</a>	01/07	Les prénoms des enfants d'immigrés
Presse régionale	<a href="#">Le Parisien</a>	10/04	Que révèlent les prénoms des enfants d'immigrés en France ?
	<a href="#">Ouest France</a>	10/04	Enzo, Nicolas, Inès... Quels prénoms portent les enfants d'immigrés ?
	<a href="#">Le Dauphiné</a>	10/04	Quels prénoms sont donnés aux enfants d'immigrés ? 4 infos à retenir
	<a href="#">Le Progrès</a>	10/04	Quels prénoms sont donnés aux enfants d'immigrés ? 4 tendances à retenir
	<a href="#">L'Est républicain</a>	10/04	Quels prénoms sont donnés aux enfants d'immigrés ? 4 infos à retenir
	<a href="#">Les dernières nouvelles d'Alsace</a>	10/04	Quels prénoms sont donnés aux enfants d'immigrés ? 4 infos à retenir
	<a href="#">Le Républicain lorrain</a>	11/04	Les prénoms aussi s'internationalisent
	<a href="#">La Provence, Corse-Matin</a>	23/04	Garance, Kevin, Inès : ce que nos prénoms disent de nous
Presse étrangère	<a href="#">El Watan</a>	16/04	Une enquête sur les prénoms des enfants de l'immigration
	<a href="#">The Economist</a>	9/05	France: When Mary met Yanis. What French baby names say about the country
	<a href="#">Inside Over</a>	4/06	France's Baby Names: a Reflection of the Political Climate?

	Média et lien	Date	Titre
Radio	<a href="#">France Info</a>	10/04	Choix des prénoms : "Les petits-enfants d'immigrés ont peu souvent des prénoms en référence au pays d'origine de leurs grands-parents"
	<a href="#">RTL</a>	10/04	VIDÉO - Quels prénoms les immigrés choisissent-ils pour leurs enfants ?
	<a href="#">Europe 1</a>	10/04	Enzo, Léa, Inès : les petits-enfants d'immigrés ont les mêmes prénoms que la population majoritaire
	<a href="#">RFI</a>	10/04	France: ce que révèle le choix des prénoms chez les descendants d'immigrés
	<a href="#">France Inter</a>	11/04	Comment les immigrés nomment leurs enfants ?
	<a href="#">RMC</a>	11/04	La GG du jour : Les prénoms, marqueurs d'intégration ?
Télévision	<a href="#">France 2</a>	10/04	Les prénoms d'enfants d'immigrés changent au fil des générations
	<a href="#">M6</a>	10/04	VIDÉO - Quels prénoms les immigrés choisissent-ils pour leurs enfants ?
	<a href="#">LCI</a>	10/04	Yanis, Inès et Lina : comment ces prénoms nous en disent long sur l'intégration des immigrés en France
	<a href="#">Africa news</a>	10/04	En France, les immigrés choisissent des prénoms "internationaux" pour leurs progénitures
	<a href="#">CNEWS</a>	10/04	Ce que révèlent les prénoms d'enfants d'immigrés
	<a href="#">BFM TV</a>	10/04	Le prénom, marqueur d'intégration ?
	<a href="#">RMC</a>	18/04	Les significations des prénoms changent-elles avec le temps ?
Sites d'information	<a href="#">Huffington Post</a>	10/04	D'Afrique du Nord ou d'Europe du Sud, comment les immigrés en France nomment leurs enfants
	<a href="#">France Soir</a>	10/04	Prénoms, immigration et intégration : quelle évolution ?
	<a href="#">Yabiladi</a>	10/04	France : De plus en plus de Maghrébins prénomment leurs enfants Yanis ou Nicolas
	<a href="#">Le café pédagogique</a>	10/04	Quels prénoms les immigrés donnent-ils à leurs enfants ?
	<a href="#">Le journal des femmes</a>	10/04	Quels prénoms les enfants d'immigrés portent-ils en France ?
	<a href="#">Oumma.com</a>	11/04	Seuls 23 % des enfants de la troisième génération portent un prénom arabo-musulman, selon une étude
	<a href="#">les étudiants du CFJ</a>	12/04	Immigration et prénom à Paris : une histoire d'intégration
	<a href="#">Ressources en sciences économiques et sociales</a>	12/04	Quels prénoms les immigrés donnent-ils à leurs enfants en France ? INED. Avril 2019.
	<a href="#">Le petit journal</a>	15/04	L'intégration en France passe-t-elle par le choix d'un prénom ?
	<a href="#">vie-publique</a>	19/04	Le choix du prénom est-il différent selon les origines ?

#### **Annexe 4A : Les échanges de messages électroniques avec les auteurs et l'INED**

Cette Annexe 4A, associée au Chapitre 4, p. 15, comprend la totalité des 50 messages électroniques que j'ai échangés avec les auteurs de l'article et l'INED, du 14 avril 2019 au 18 décembre 2020, pour connaître la méthodologie de l'article de Coulmont et Simon (2019).

J'ai échangé successivement :

- avec les seuls auteurs de l'article, Baptiste Coulmont (11 messages du 14 avril 2019 au 13 mai 2019) puis Baptiste Coulmont et Patrick Simon (15 messages du 14 mai 2019 au 26 septembre 2019) ;
- avec les auteurs et le rédacteur en chef de la revue *Population et sociétés*, Gilles Pison (4 messages du 26 au 30 septembre 2019) ;
- avec les auteurs et la directrice de la publication de *Population et sociétés* et directrice de l'INED, Magda Tomasini, ainsi qu'avec le référent à l'intégrité scientifique de l'INED, Henri Leridon (5 messages du 30 septembre au 14 novembre 2019) ;
- avec les auteurs, la directrice de l'INED, le référent à l'intégrité scientifique de l'INED et la Mission à l'intégrité scientifique du CNRS (9 messages du 17 février 2020 au 25 mai 2020) ;
- avec les auteurs, la directrice de l'INED, le référent à l'intégrité scientifique de l'INED, la Mission à l'intégrité scientifique du CNRS et le médiateur intégrité scientifique de l'ENS Paris-Saclay (5 messages du 1<sup>er</sup> décembre 2020 au 5 décembre 2020) ;
- avec les auteurs, la directrice de l'INED, le référent à l'intégrité scientifique de l'INED et le médiateur intégrité scientifique de l'ENS Paris-Saclay (1 message le 18 décembre 2020), l'investigation de la Mission à l'intégrité scientifique du CNRS ayant pris fin le 17 décembre 2020.

❖ **Échanges avec les auteurs de l'article, Baptiste Coulmont et Patrick Simon**

<b>Intitulé :</b> Quels prénoms les immigrés donnent-ils à leurs enfants en France ?		
<b>Date :</b> 14 avril 2019		
<b>De :</b> Jean-François Mignot	<b>À :</b> Baptiste Coulmont	<b>Copie :</b> -
<p>Cher collègue,</p> <p>J'ai lu avec grand intérêt et avec attention votre Pop &amp; Soc, que je souhaiterais citer dans mes travaux, mais j'aurais plusieurs questions, notamment méthodologiques, à vous poser pour comprendre vos résultats (ci-dessous). Accepteriez-vous d'en discuter ?</p> <p>Je me permets de vous indiquer que je serais disponible pour déjeuner à Pouchet tous les jours de cette semaine (sauf vendredi) ainsi que jeudi 25 et vendredi 26. Si jamais il était plus pratique pour vous d'en discuter ailleurs ou à d'autres dates ou horaires, je saurai me rendre disponible.</p> <p>Bien cordialement,</p> <p>Jean-François Mignot</p> <p><b>Questions sur la méthodo</b></p> <p>Comment fonctionne précisément la classification linguistique des prénoms que vous utilisez, à partir de leur structure syllabique (« classification probabiliste des référents linguistiques des syllabes ») ?</p> <p>Comment cette classification fait-elle émerger un groupe « asiatique » regroupant ensemble les prénoms indiens, chinois et d'Asie du sud-est, alors que les langues d'Inde (indo-européennes ou dravidiennes), de Chine (sino-tibétaines) et d'Asie du sud-est (austroasiatiques) ont des consonances si différentes ?</p> <p>L'enfant d'un immigré malien, qui se prénommerait Mamadou, Ibrahim ou Fatoumata, serait-il catégorisé comme ayant un prénom « africain » ou « arabo-musulman » ?</p> <p>Les prénoms appelés dans l'encadré « internationaux ou anglophones » sont-ils les mêmes que les prénoms appelés « internationaux » en conclusion et dans le résumé ?</p> <p>Dans votre encadré méthodo, vous indiquez qu'il existe trois façons de catégoriser les prénoms selon leur origine culturelle : sauriez-vous dans quelle mesure vos résultats seraient affectés par le choix d'une de ces méthodes plutôt que l'une des deux autres ?</p> <p><b>Questions sur les résultats</b></p> <p>Quelles proportions de l'ensemble des prénoms de votre base de données sont classifiés comme « français », « arabo-musulmans », « latins », « africains », « asiatiques », « internationaux ou anglophones », « autre groupe » et « inclassables » ?</p> <p>Comment les fréquences relatives de ces catégories ont-elles évolué au fil des cohortes de naissance ?</p>		

<b>Intitulé :</b> Re: Quels prénoms les immigrés donnent-ils à leurs enfants en France ?		
<b>Date :</b> 16 avril 2019		
<b>De :</b> Baptiste Coulmont	<b>À :</b> Jean-François Mignot	<b>Copie :</b> -
<p>Bien reçu.</p> <p>J'ai quelques trucs en retard, et je vais manger très rapidement aujourd'hui.</p> <p>Pourquoi pas un café sur la terrasse demain mercredi (s'il fait beau ça sera plus agréable que la cantine) ?</p> <p>Bien cordialement,</p> <p>B. Coulmont</p>		



<b>Intitulé :</b> Re: Quels prénoms les immigrés donnent-ils à leurs enfants en France ?		
<b>Date :</b> 16 avril 2019		
<b>De :</b> Jean-François Mignot	<b>À :</b> Baptiste Coulmont	<b>Copie :</b> -
Super, merci de votre réponse. Vous avez raison, c'est plus sympa sur la terrasse. Pourrions-nous dire RV à 13h00 ? Bien cordialement, et en vous souhaitant une excellente journée, Jean-François		

<b>Intitulé :</b> Re: Quels prénoms les immigrés donnent-ils à leurs enfants en France ?		
<b>Date :</b> 17 avril 2019		
<b>De :</b> Baptiste Coulmont	<b>À :</b> Jean-François Mignot	<b>Copie :</b> -
Ca marche. 13h sur la terrasse. -- Baptiste Coulmont		

<b>Intitulé :</b> Re: Quels prénoms les immigrés donnent-ils à leurs enfants en France ?		
<b>Date :</b> 17 avril 2019		
<b>De :</b> Jean-François Mignot	<b>À :</b> Baptiste Coulmont	<b>Copie :</b> -
Merci, à tout à l'heure, Jean-François		

[17 avril 2019 : Discussion orale avec Baptiste Coulmont]

<b>Intitulé :</b> Quels prénoms les immigrés donnent-ils à leurs enfants en France ?		
<b>Date :</b> 18 avril 2019		
<b>De :</b> Jean-François Mignot	<b>À :</b> Baptiste Coulmont	<b>Copie :</b> -
<p>Bonjour Baptiste,</p> <p>Je vous remercie d'avoir accepté d'échanger hier, ce qui me permet d'y voir déjà beaucoup plus clair dans l'article. J'espère ne pas avoir été trop pénible, et dans le cas contraire je vous prie de m'en excuser. Pour information, la classe linguistique 441 « Iwrit », dont nous nous demandions ce qu'elle pouvait bien signifier, semble désigner l'hébreu moderne.</p> <p>Concernant mes motifs de méfiance quant à l'honnêteté intellectuelle de votre coauteur - qui sont une raison majeure de ma curiosité méthodologique sur ce Pop&amp;Soc -, je tiens à vous les présenter sans fard, tels qu'exprimés par Tribalat (<a href="https://www.cairn.info/revue-commentaire-2016-3-page-698.htm?contenu=resume">https://www.cairn.info/revue-commentaire-2016-3-page-698.htm?contenu=resume</a>) à propos du chapitre du livre TeO sur les religions.</p> <p>Premier motif :</p> <p>L'étude des affiliations religieuses frise la malhonnêteté. Tout d'abord, il est d'usage de citer les travaux parus avant les siens, surtout quand on s'en est inspiré. Or j'ai publié, notamment dans <i>Commentaire</i>, une estimation de la population musulmane (dont la méthode a été reprise dans ce livre) et un grand nombre d'analyses à partir de TEO qui sont ignorées, mais parfois dupliquées. [Tribalat ne mentionne pas qu'elle a aussi été omise de la longue bibliographie du site de l'enquête TeO : <a href="https://teo.site.ined.fr/fr/donnees_et_resultats/publications/">https://teo.site.ined.fr/fr/donnees_et_resultats/publications/</a>] Enfin, les auteurs développent une analyse spéculative de la transmission de l'islam [elle fait référence au graphique qui, depuis, a été reproduit sur le site de l'Ined : <a href="https://www.ined.fr/fr/tout-savoir-population/memos-demo/focus/dynamique-des-religions-immigres-et-descendants-en-france/">https://www.ined.fr/fr/tout-savoir-population/memos-demo/focus/dynamique-des-religions-immigres-et-descendants-en-france/</a>]. Alors que celle-ci s'est considérablement améliorée au fil du temps auprès des enfants d'immigrés, les auteurs concluent à « une sécularisation progressive au fil des générations ». Pour cela ils ont construit un indicateur en trois postes : sécularisation, reproduction, renforcement. Seront dits en voie de sécularisation ceux qui accordent moins d'importance à la religion que n'en accordaient leurs parents dans leur éducation. Ainsi, ceux, très nombreux, dont les parents accordaient beaucoup d'importance à la religion et qui en accordent tout autant sont dans la reproduction. Ils accordent autant d'importance à la religion que leurs parents. Aucun renforcement n'est possible pour eux car il n'y a rien au-</p>		

dessus de « beaucoup ». À l’opposé, ceux, peu nombreux, dont les parents n’accordaient pas d’importance à la religion, et qui déclarent en accorder un peu, sont dans le renforcement. La sécularisation concerne tout aussi bien ceux qui passent de « pas du tout » à « sans religion » que ceux qui passent de « beaucoup » à « assez ». La part de « renforcement » sera d’autant plus faible que l’on est déjà au top, côté parents ! Or, au fil des générations nées en France, les parents ont accordé de plus en plus d’importance à la religion. Cette « entourloupe » permet aux auteurs de conclure que « le regain religieux ne concerne qu’une petite minorité de jeunes musulmans » et « à une forme atténuée de sécularisation et de baisse de religiosité qui signale une transformation du rôle de la religion ». Rappelons simplement que, des enfants d’au moins un parent musulman nés en France en 1958-1964 à ceux nés en 1985-1989, la transmission est passée de 43 % à 87 % ! Considérant — comme nous l’avions fait — l’effondrement, entre MGIS 1992 et TEO 2008, de la proportion de sans religion chez les jeunes d’origine algérienne, les auteurs invoquent « une religiosité réactive » liée « au durcissement du cadre laïque » et à une « vision péjorative de l’islam ».

#### Second motif :

Les auteurs invalident l’idée d’un communautarisme musulman. La question sur les amis rencontrés au cours des quinze derniers jours est censée renseigner sur leur sociabilité. Si l’enquêté a rencontré un ami seulement et qu’il est de même religion que lui, c’est du 100 %, et réciproquement. Les enquêtés sont déclarés religieusement homophiles s’ils ont rencontré un ou des amis dont plus de la moitié a la même religion. 2 % des musulmans seulement ne savent pas dénommer la religion du ou des amis rencontrés. Mais c’est le cas de 16 % des catholiques et de 28 % des personnes sans religion. Au lieu de prendre au sérieux leur réponse reflétant probablement une réelle indifférence à l’affiliation religieuse de leurs amis, les auteurs ont recalculé la proportion d’homophiles en sortant les réponses de ceux qui disent ne pas connaître la religion de leur(s) ami(s). Et là, miracle, ce sont les catholiques (65 %) et les sans religion (50 %) qui sont nettement plus homophiles que les musulmans (43 %). Le tour est joué : « les plus “communautaires” ne sont pas ceux qu’on croit ». Les auteurs pensent invalider ainsi « les représentations médiatiques d’une communauté repliée sur elle-même ». La possibilité d’un biais lié à la position majoritaire n’est évoquée que pour les origines. Pas pour les religions. Pourtant, si l’on retient la composition religieuse des 18-50 ans de TEO et si les religions étaient réparties uniformément sur tout le territoire sans tenir compte d’aucune affinité particulière, la probabilité théorique des catholiques de fréquenter des catholiques serait de 43 %, celle des sans religion de fréquenter des personnes sans religion de 45 % et celle des musulmans de fréquenter des musulmans de 8 %. Rien d’étonnant à ce que l’homophilie des catholiques soit plus élevée que celle des musulmans.

Les auteurs du chapitre doublement mis en cause par Tribalat n’ont à ma connaissance rien répondu. Dans ce contexte, vous comprendrez peut-être que je ne puis me fier aux résultats publiés par votre co-auteur *sans vérifications supplémentaires*, même si l’autre co-auteur du papier — vous — est placé à l’autre extrême du spectre de l’honnêteté intellectuelle, comme en témoigne abondamment votre blog. C’est d’ailleurs pourquoi je me permets de vous contacter *vous*.

Comme je n’ai pas de visibilité sur (i) la catégorisation linguistique des prénoms par Humpert et Schneiderheinze ni sur (ii) vos regroupements de leurs catégories linguistiques, je ne peux pas en l’état juger du degré de fiabilité de la classification des prénoms en 8 origines linguistiques et culturelles. Or, la force des résultats de votre article me contraint à le discuter dans le livre que je suis en train de rédiger. Par conséquent, accepteriez-vous de me communiquer la table Excel de structure suivante ?

**Tableau de données non ré-identifiables**

	Effectif pondéré TeO	Catégorie linguistique (Humpert & Schneiderheinze 2000)	Regroupement des catégories linguistiques (Coulmont et Simon 2019)
Prénom 1			
Prénom 2			
...			
Prénom 18 163			

Cette table me permettra de me convaincre que la méthode de l'article, qui permet d'associer à chaque prénom une origine linguistique, ne pose pas de problème interne majeur. Il me permettra aussi de répliquer cette méthodologie sur le Fichier Insee des prénoms, afin d'évaluer dans quelle mesure elle produit des résultats globalement similaires à ceux de Fourquet (2019). Répliquer votre méthode sur le Fichier Insee permettrait par ailleurs de connaître quelles sont les proportions de prénoms « français », « latins », « arabo-musulmans », « africains », « asiatiques », « internationaux ou anglophones » et « inclassables » qui sont donnés chaque année en France aux nouveau-nés depuis 1900, ce qui ne serait pas inintéressant.

Si vous souhaitiez que je m'engage explicitement par écrit à ne rien publier à partir des données de ce tableau (pas même ma réplique de votre méthode sur le Fichier Insee des prénoms), ni bien entendu à le communiquer à quiconque, je le ferai volontiers : mon seul objectif est de mettre un terme aux doutes qui m'assaillent dans toute publication de votre co-auteur sur le sujet, a fortiori peut-être quand la source de sa méthodologie (Humpert & Schneiderheinze 2000) n'est pas référencée et que la référence (Mazières et Roth 2018) ne semble pas utilisée. Par ailleurs, ces références méthodologiques portent soit sur des noms de famille, soit sur des combinaisons de prénom et de nom de famille, plutôt que sur des seuls prénoms. De mon point de vue, une vérification soigneuse de la fiabilité de la méthodologie s'impose avant d'admettre les résultats. Merci d'avance de me communiquer le tableau susmentionné, qui me permettra d'être convaincu de la fiabilité des résultats publiés. Je reste disponible pour discuter de tout cela par oral, si jamais cela était plus pratique pour vous.

Bien cordialement,

Jean-François

**Intitulé :** Re: Quels prénoms les immigrés donnent-ils à leurs enfants en France ?

**Date :** 19 avril 2019

**De :** Baptiste Coulmont

**À :** Jean-François Mignot

**Copie :** -

Bonjour,

Je prends quelques jours de vacances, et je me consacre aujourd'hui à la finalisation d'un papier. Je ferai les calculs avant le premier mai.

Sur les chiffres de Fourquet (garçons, sans les prénoms rares) : quelle que soit la liste de prénoms arabes, musulmans, etc... que vous prendrez (que ce soit la "liste des beaux prénoms arabes" ou d'autres...) vous trouverez dans le "Fichier des prénoms" les mêmes proportions, à quelques % près. Idem avec les prénoms de TeO (avec le problème que de nombreux prénoms du Fichier des prénoms ne se trouvent pas dans TeO).

Sur la "surprise" de trouver peu de prénoms "arabes" en 3e génération, ce document

[https://www.ined.fr/fichier/s\\_rubrique/19578/document\\_travail\\_2013\\_189\\_rencontre\\_conjoint.fr.pdf](https://www.ined.fr/fichier/s_rubrique/19578/document_travail_2013_189_rencontre_conjoint.fr.pdf) (tableau 11 p.26)

indique qu'un tiers environ des descendants d'immigrés du Maghreb (en couple) ont 1 seul parent immigré (ils ont souvent un autre parent descendant d'immigré, mais les descendants d'immigrés n'ont pas les mêmes goûts ni les mêmes prénoms, déjà, que les immigrés), et environ 40/45 % de ces descendants d'immigrés du Maghreb ont un conjoint "majoritaire".

[quand, au lieu de s'intéresser simplement aux descendants d'immigrés en couple, mais aux descendants d'immigrés qui ont donné naissance à au moins un enfant en France, et en s'intéressant donc aussi au couple parental et non plus seulement à la situation conjugale en 2008, on a des chiffres légèrement différents]

Ceci fait que la situation modale pour les \*petits enfants\* d'immigrés est d'avoir 2 \*grands parents\* originaires du Maghreb (et, par construction, ils n'ont aucun parent immigré).

Reprenons la discussion début mai !

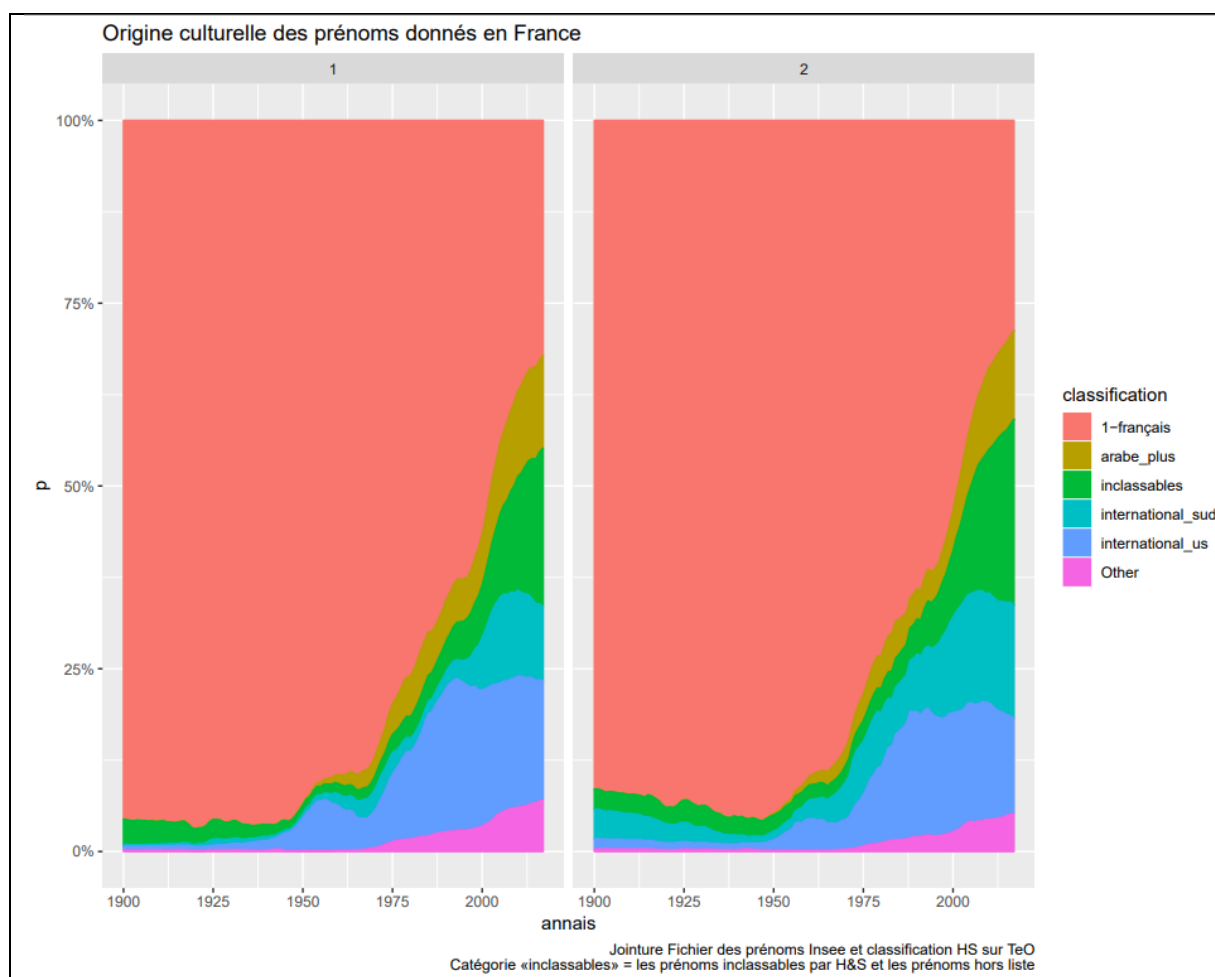
Bien cordialement,

B. Coulmont

<b>Intitulé</b> : Re: Quels prénoms les immigrés donnent-ils à leurs enfants en France ?		
<b>Date</b> : 19 avril 2019		
<b>De</b> : Jean-François Mignot	<b>À</b> : Baptiste Coulmont	<b>Copie</b> : -
<p>Bonjour,</p> <p>Merci de votre réponse, et d'accepter de me communiquer la table d'ici le 1<sup>er</sup> mai. Merci aussi de ces précisions fort utiles, même si je trouve le tableau auquel vous faites référence inhabituellement complexe. Peut-être ai-je moi aussi besoin de vacances...</p> <p>Très bonnes vacances à vous,</p> <p>Jean-François</p>		

<b>Intitulé</b> : n'oubliez pas...		
<b>Date</b> : 6 mai 2019		
<b>De</b> : Jean-François Mignot	<b>À</b> : Baptiste Coulmont	<b>Copie</b> : -
<p>... ma demande d'il y a plus de quinze jours ! Cordialement,</p> <p>Jean-François</p>		

<b>Intitulé</b> : Re: n'oubliez pas...		
<b>Date</b> : 13 mai 2019		
<b>De</b> : Baptiste Coulmont	<b>À</b> : Jean-François Mignot	<b>Copie</b> : -
<p>Bonjour,</p> <p>Avec mes excuses pour le retard : fin d'année à Paris 8, corrections de copies, correction de l'agrégation de SES, articles en retard, vacances plus que nécessaires, et j'en passe.</p> <p>En document attaché, un graphique réalisé sur le Fichier des prénoms à partir de la classification H&amp;S (avec regroupements). Ca ne donne rien de bien intéressant étant donné le grand nombre de prénoms du "Fichier des prénoms" qui ne sont pas dans TeO.</p> <p>Un autre document attaché, une sélection d'une cinquantaine de prénoms fréquents dans TeO (bases Ego et bases THL+Enfants) avec la classification H&amp;S pour ces prénoms.</p> <p>Pour une réplique complète : il faudra passer par une demande d'accès aux données nominatives via le Comité du secret statistique.</p> <p>Bien cordialement,</p> <p>B. Coulmont</p>		
<b>Pièces attachées</b> :		



PRENOM	classification	Description
CEDRIC	international_us	Great Britain (English/Welsh)
SABRINA	international_sud	Italy
OCEANE	1-français	France
NADIA	international_sud	First names common in many countries (not only European), some very widely used (Maria), some less (Alina)
MYRIAM	1-français	France
MARIE	1-français	France
JACQUES	1-français	France
LEA	international_sud	First names common in many countries (not only European), some very widely used (Maria), some less (Alina)
ROMAIN	1-français	France
FABIEN	1-français	France
KEVIN	international_us	Great Britain (English/Welsh)
JEAN	1-français	France
OMAR	arabe_plus	Maghreb (Morocco, Algeria, Tunisia)
JOSE	international_sud	Spain
MAXIME	1-français	France
MICKAEL	1-français	France
XAVIER	1-français	France

YANIS	arabe_plus	Arabic speaking States (Libya, Egypt, Jordan, Syria, Iraq, Palestine, Arabian Peninsula)
LAURENCE	1-français	France
AMINATA	afrique	Africa except for Northern Africa (Maghreb, Libya, Egypt)
WILLIAM	international_us	Great Britain (English/Welsh)
ALEXANDRE	1-français	France
EMMA	international_us	Great Britain (English/Welsh)
THIERRY	1-français	France
MARTINE	1-français	France
LUCAS	1-français	France
CLEMENT	1-français	France
JADE	international_us	Great Britain (English/Welsh)
VANESSA	international_us	Great Britain (English/Welsh)
CHANTAL	1-français	France
BENJAMIN	1-français	France
JEREMY	international_us	Great Britain (English/Welsh)
CLAUDE	1-français	France
LUCIE	1-français	France
JENNIFER	international_us	Great Britain (English/Welsh)
RACHID	arabe_plus	Maghreb (Morocco, Algeria, Tunisia)
CAROLINE	1-français	France
ADAM	1-français	France
FABRICE	1-français	France
NATHALIE	1-français	France
BEATRICE	1-français	France
SANDRINE	1-français	France
ADRIEN	1-français	France
LOUIS	1-français	France
MELISSA	international_sud	First names common in many countries (not only European), some very widely used (Maria), some less (Alina)
AXEL	international_us	Scandinavia
LINDA	international_us	Great Britain (English/Welsh)
ANDRE	1-français	France
LAETITIA	1-français	France
ERIC	1-français	France
ENZO	international_sud	Italy
STEPHANE	1-français	France
MEHDI	arabe_plus	Muslim without any further classification (names like Ahmed, Mohammed)
ANISSA	arabe_plus	Maghreb (Morocco, Algeria, Tunisia)
HERVE	1-français	France

<b>Intitulé</b> : Re: n'oubliez pas...		
<b>Date</b> : 13 mai 2019		
<b>De</b> : Jean-François Mignot	<b>À</b> : Baptiste Coulmont	<b>Copie</b> : -
<p>Bonsoir,</p> <p>Pas de souci pour le retard : je sais bien que vous n'avez pas que cela à faire.</p> <p>Simplement, je ne vous avais pas demandé d'analyse complémentaire ni d'extrait de votre liste de prénoms : je vous avais demandé une table contenant (i) la liste des 18 163 prénoms TeO, (ii) la liste des origines linguistiques détaillées associées à chacun de ces prénoms selon la classification de H&amp;S, et (iii) la liste des origines linguistiques regroupées par vous dans votre article (cf. ci-dessous, « Tableau de données non ré-identifiables »). Ces informations ne sont-elles pas nécessaires pour répliquer les résultats de votre article ? N'êtes-vous pas autorisé à communiquer cette table à un collègue ?</p> <p>Si vous ne souhaitez pas me communiquer ces données, dites-le moi franchement : je me permettrai alors de les demander à Patrick Simon, et ne vous contacterai plus. Si lui non plus ne souhaite pas me les communiquer, je me verrai contraint de les demander au rédacteur en chef de la revue où vous avez publié l'article dont vous n'autorisez pas la réplique.</p> <p>Si vous préférez discuter de tout cela par oral plutôt que par écrit, je suis disponible cette semaine et les suivantes.</p> <p>Respectueusement,</p> <p>Jean-François Mignot</p>		

[13 mai 2019 : Discussion orale avec Baptiste Coulmont]

<b>Intitulé</b> : Quels prénoms les immigrés donnent-ils à leurs enfants en France ?		
<b>Date</b> : 14 mai 2019		
<b>De</b> : Jean-François Mignot	<b>À</b> : Patrick Simon	<b>Copie</b> : Baptiste Coulmont
<p>Bonjour Monsieur Simon,</p> <p>Démographe au CNRS, j'ai lu avec intérêt et avec attention votre Pop &amp; Soc, que je souhaiterais citer dans mes travaux, mais j'aurais plusieurs questions méthodologiques à vous poser pour comprendre vos résultats, puis pour pouvoir les répliquer (ci-dessous). Accepteriez-vous d'en discuter ? J'en ai déjà discuté avec votre coauteur, Baptiste Coulmont, mais comme il n'a pas jugé utile de répondre favorablement à mes demandes, je me permets de me tourner vers vous.</p> <p>Je suis disponible dans les jours qui viennent, et je peux me rendre à l'Ined si cela vous convient.</p> <p>Dans l'attente de vous rencontrer pour échanger,</p> <p>Jean-François Mignot</p> <p><b>Questions</b></p> <p>Comment fonctionne précisément la classification linguistique des prénoms que vous utilisez, à partir de leur structure syllabique (« classification probabiliste des référents linguistiques des syllabes ») ?</p> <p>Comment avez-vous classifié puis regroupé les origines linguistiques de chacun des prénoms présents dans votre base TeO ?</p>		

<b>Intitulé</b> : Re: Quels prénoms les immigrés donnent-ils à leurs enfants en France ?		
<b>Date</b> : 14 mai 2019		
<b>De</b> : Baptiste Coulmont	<b>À</b> : Jean-François Mignot	<b>Copie</b> : Patrick Simon
<p>Bonjour,</p> <p>J'ai quand même jugé utile d'écrire ceci en réponse aux questions :</p> <p>1- Humpert &amp; Schneiderheinze : <a href="https://www.ssoar.info/ssoar/handle/document/20805">https://www.ssoar.info/ssoar/handle/document/20805</a></p> <p>2- Asiatiques = regroupement de catégories comme 432 Vietnamese</p>		



436 Indian  
437 Indonesian  
439 Iranian  
441 Iwrit  
442 Japanese  
446 Khmer  
449 Laotian  
457 Mongolian  
467 Korean  
476 Thai  
479 Chinese

3- Mamadou : 623 -- West Africa (Benin, Burkina Faso, Ivory Coast, Gambia, Ghana, Guinea, Guinea-Bissau, Liberia, Mali, Niger, Nigeria, Senegal, Sierra Leone, Togo)

3- Ibrahim : 662 Muslim without any further classification (names like Ahmed, Mohammed) - 641 Arabic speaking States -- 163 Turkey

3- Fatoumata (ou Fatou) : 623

4- Non (plus large en conclusion)

5- Résultats affectés par la méthode ?

- 18163 "prénoms" différents (105 000 "individus") : donc classement manuel impossible.

- 2 autres méthodes : "score de distance" (French name index / foreign name index) et classification syllabique.

- 2 interprétations différentes : \*\*le score de distance\*\* estime la proximité des comportement des (parents des)immigrés/descendants avec la population des natifs (sachant que les descendants d'immigrés sont des natifs). Il permet de montrer que les immigrés ne donnent pas "leurs prénoms" à leurs enfants, et que les descendants d'immigrés ne donnent pas les "prénoms de leurs parents" à leurs enfants. Les petits-enfants des immigrés ont des prénoms qui ont le même score que celui des \*\*enfants de la population majoritaire\*\*. Les \*\*catégories de prénoms\*\* sont plus directement parlantes.

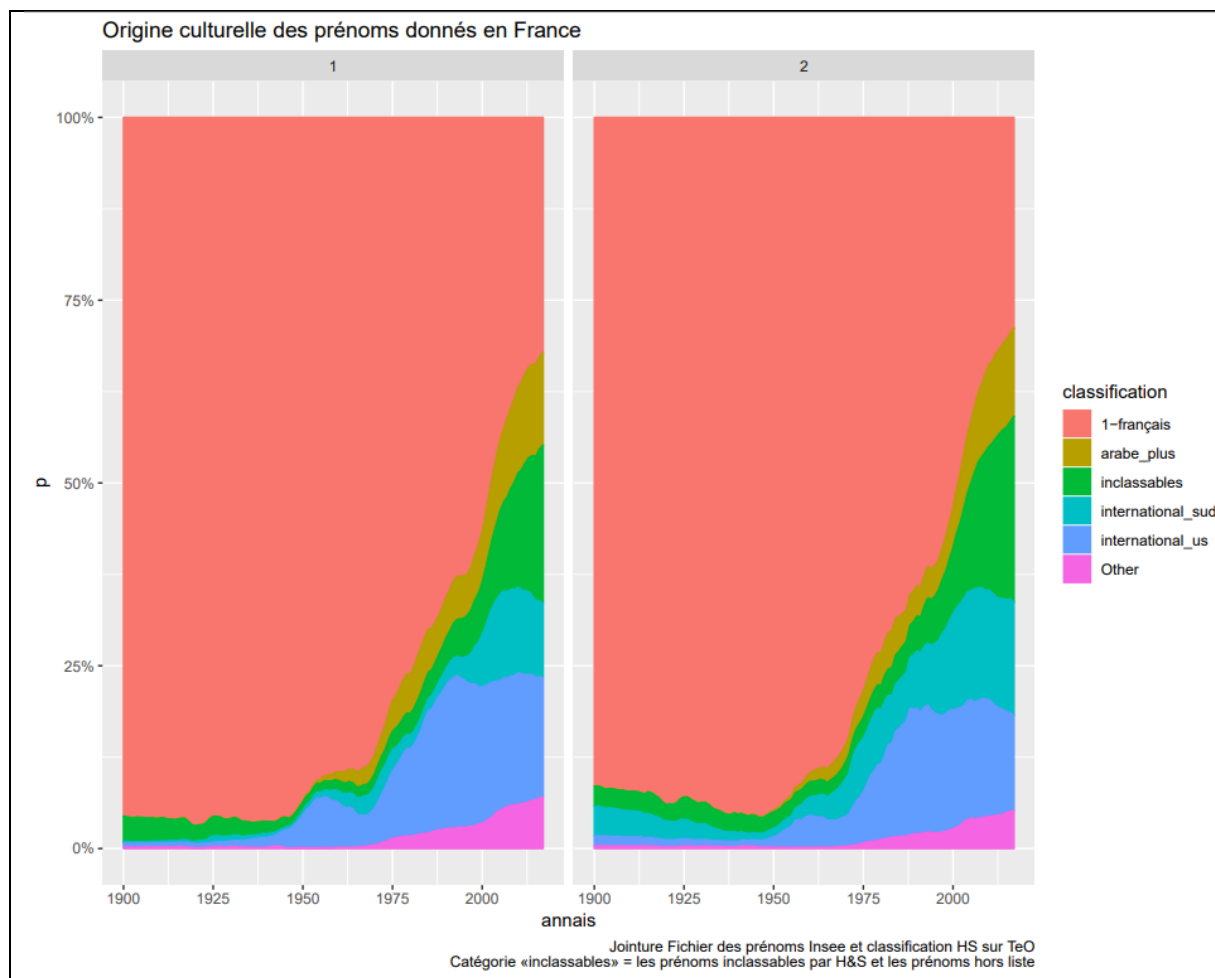
6- << proportions de l'ensemble des prénoms de votre base de données sont classifiés comme « français » >> : base THL, \*\*sans pondération\*\* : 41 % prénoms français, 20 % prénoms arabes, 9 % prénoms "international sud", 9 % "international anglo", 6 % inclassables, 5,5 % turqui, 4,5 % afrique...

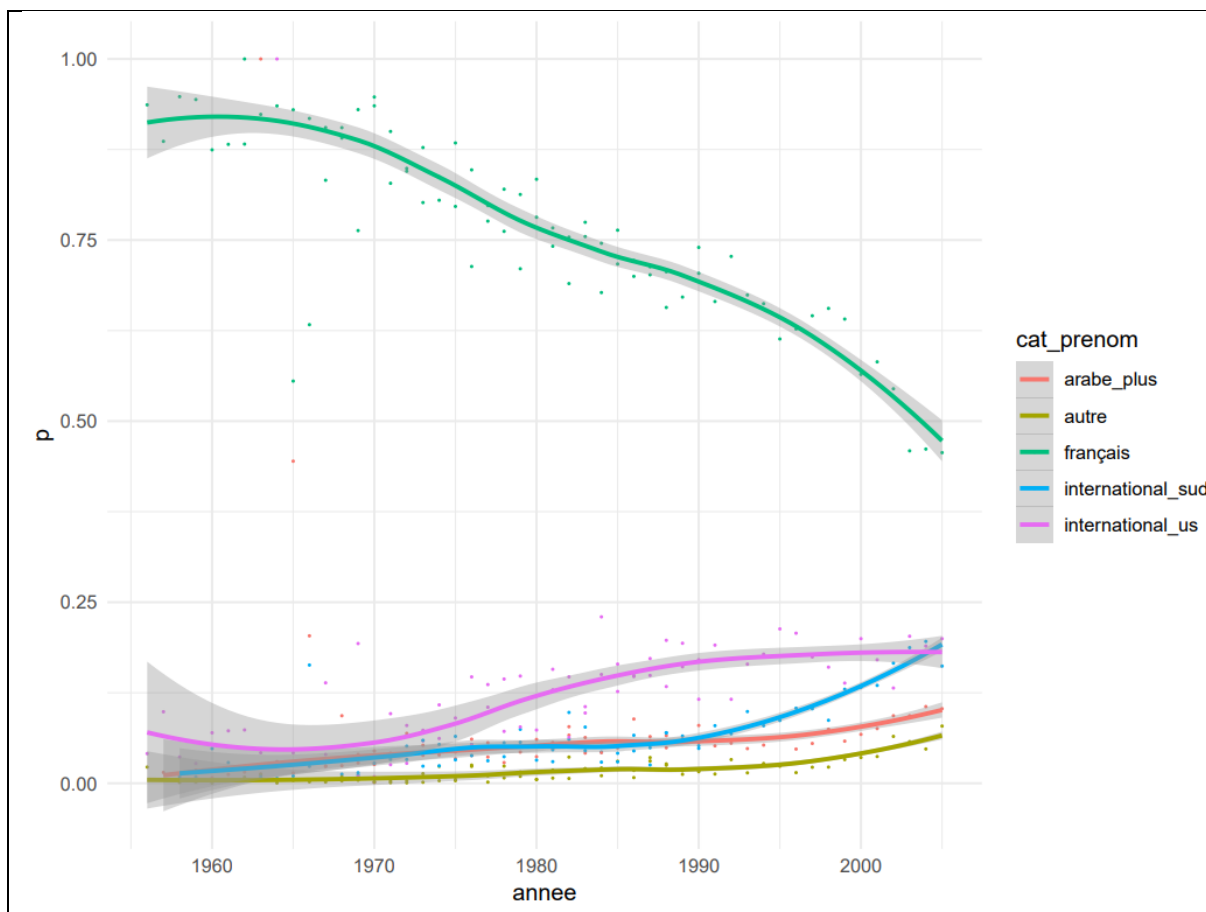
Et aussi à produire ces graphiques (en document attaché).

Mais j'estime -- et peut-être suis-je trop strict -- ne pas avoir l'autorisation de vous transmettre la liste complète des prénoms des enquêtés de TeO et de leurs enfants.

Bien cordialement,  
B. Coulmont

**Pièces attachées :**





**Intitulé :** Re: Quels prénoms les immigrés donnent-ils à leurs enfants en France ?

**Date :** 14 mai 2019

**De :** Patrick Simon

**À :** Jean-François Mignot

**Copie :** Baptiste Coulmont

Cher collègue,

Merci de votre intérêt pour notre travail. Il me semble que Baptiste vous a éclairé sur vos questions. La méthode probabiliste fondée sur des appariements linguistiques de H&S est bien décrite dans leurs publications. Je n'ai rien de plus à en dire. Les principes de regroupement sont ceux des zones linguistico-culturelles, rien de très inventif là dedans.

Je ne sais pas ce que vous voulez reproduire, mais si c'est sur la base des prénoms dans TeO, elle n'est disponible que via le comité du secret, et Baptiste a raison de ne pas la diffuser.

Je ne suis pas disponible avant début juillet, mais nous pouvons convenir d'un rdv si vous souhaitez toujours en parler.

Bien à vous,

Patrick Simon

**Intitulé :** Re: Quels prénoms les immigrés donnent-ils à leurs enfants en France ?

**Date :** 16 mai 2019

**De :** Jean-François Mignot

**À :** Patrick Simon

**Copie :** Baptiste Coulmont

Bonjour,

Merci de votre réponse. Je vais effectivement soumettre au Comité du secret statistique une demande de communication de l'enquête TeO enrichie de la variable prénom.

Mais pour pouvoir prolonger vos analyses, tout en utilisant la même méthodologie d'attribution d'une origine linguistique probable aux prénoms, j'aurais aussi besoin de la table de passage qui permet (i) de passer de chaque prénom à son origine linguistique détaillée (selon la méthode de

Humpert et Schneiderheinze, qui semble produire plusieurs dizaines de catégories linguistiques), et (ii) de regrouper ces origines linguistiques détaillées en 7 ou 8 grandes catégories (« français », « latins », « arabo-musulmans », « africains », « asiatiques », « internationaux ou anglophones » et « inclassables »). D'après ce que je peux comprendre, cette table de passage pourrait avoir la structure suivante :

**Table de passage du prénom à son origine linguistique probable**

	Catégorie linguistique détaillée (Humpert & Schneiderheinze 2000)	Catégorie linguistique regroupée (Coulmont et Simon 2019)
Prénom 1		
Prénom 2		
...		
Prénom 18 163		

Accepteriez-vous de me communiquer cette table ou tout autre élément qui me permettrait de prolonger vos analyses tout en réutilisant la même méthodologie pour attribuer une origine linguistique aux prénoms présents dans TeO ? Une telle table de passage du prénom à son origine linguistique probable ne contient des informations que sur des prénoms (et pas sur des individus, a fortiori des individus enquêtés dans TeO), si bien qu'elle n'est pas susceptible d'être utilisée de façon qui puisse porter tort à des enquêtés ; en conséquence, elle n'est pas soumise au secret statistique.

Merci d'avance et bien à vous,  
Jean-François Mignot

**Intitulé :** Re: Quels prénoms les immigrés donnent-ils à leurs enfants en France ?

**Date :** 23 mai 2019

**De :** Jean-François Mignot

**À :** Patrick Simon

**Copie :** Baptiste Coulmont

Bonjour,

Je n'ai pas reçu de réponse à ma demande de la semaine dernière. Dois-je comprendre que vous ne souhaitez pas y répondre favorablement ?

Cordialement,

Jean-François Mignot

**Intitulé :** Re: Quels prénoms les immigrés donnent-ils à leurs enfants en France ?

**Date :** 23 mai 2019

**De :** Patrick Simon

**À :** Jean-François Mignot

**Copie :** Baptiste Coulmont

Cher collègue,

Je n'ai pas eu le temps de considérer votre question et ses implications. Ce que vous demandez est le partage d'un outil de recherche alors même que vous ne participez pas à un programme de recherche commun. La mise à disposition des données ne comprend pas les constructions de variables que nous avons opérées à des fins d'analyse, ce qui est normal dans le cadre de grandes enquêtes quantitatives.

Nous partageons nos outils personnels de recherche dans le cadre d'un travail commun, or vous ne vous inscrivez pas dans ce cadre. Donc votre demande mérite réflexion. Donnez moi un bon argument pour que je vous épargne la tâche de faire votre propre classification onomastique, alors même que je n'ai aucune connaissance de votre agenda de recherche ?

Bien cordialement,

Patrick Simon

PS : et dans le cadre d'une demande d'un service, il est préférable d'éviter le ton comminatoire.

<b>Intitulé</b> : Re: Quels prénoms les immigrés donnent-ils à leurs enfants en France ?		
<b>Date</b> : 23 mai 2019		
<b>De</b> : Jean-François Mignot	<b>À</b> : Patrick Simon	<b>Copie</b> : Baptiste Coulmont
<p>Cher collègue,</p> <p>Merci de votre réponse. Si mon message a pu vous paraître comminatoire, je vous présente mes excuses.</p> <p>Ma demande de communication de la classification onomastique utilisée dans l'article ne correspond pas à la demande d'un service, ni a fortiori - dans mon esprit - à une menace, mais au cours normal de la recherche scientifique. Cette table de passage étant nécessaire pour répliquer la méthodologie de l'article – et nécessaire aussi pour la comparer à d'autres méthodes (« liste de prénoms par origine » et « score de distance ») et pour prolonger les acquis de l'article –, sa communication à un collègue chercheur fait partie du cours normal de l'activité scientifique. Ainsi lorsqu'un collègue (d'ailleurs associé à l'Ined) m'a demandé, le 18 mai 2017, de lui communiquer les éléments nécessaires à la réplique d'un article auquel il n'avait nullement participé (« L'écart d'âge entre conjoints », <i>Revue française de sociologie</i>, 2010), je lui ai communiqué ces éléments de façon inconditionnelle et dans la journée même : rien que de très normal dans le cadre de l'activité scientifique.</p> <p>Quant à mon agenda de recherche, que j'ai joint à ma demande d'accès aux données du Comité du secret statistique, il porte sur « Les déterminants du choix du prénom » dans la France actuelle et vient d'obtenir l'accord de l'Insee pour transmission au secrétariat du comité.</p> <p>Bien à vous, et en vous renouvelant mes excuses si le ton de mon message précédent a pu vous paraître inapproprié,</p> <p>Jean-François Mignot</p>		

<b>Intitulé</b> : Re: Quels prénoms les immigrés donnent-ils à leurs enfants en France ?		
<b>Date</b> : 3 juin 2019		
<b>De</b> : Jean-François Mignot	<b>À</b> : Patrick Simon	<b>Copie</b> : Baptiste Coulmont
<p>Bonjour,</p> <p>En l'absence de réponse de votre part, je me permets de vous recontacter pour savoir si vous acceptez de me communiquer la double table de passage qui conditionne la réutilisation de votre méthodologie de classification des prénoms par origine linguistique probable. Merci d'avance de votre réponse,</p> <p>Jean-François Mignot</p>		

<b>Intitulé</b> : Re: Quels prénoms les immigrés donnent-ils à leurs enfants en France ?		
<b>Date</b> : 3 juin 2019		
<b>De</b> : Patrick Simon	<b>À</b> : Jean-François Mignot	<b>Copie</b> : Baptiste Coulmont
<p>Bonjour</p> <p>Deux réponses rapides :</p> <p>1) Mon agenda est ultra bouclé jusqu'à la mi juillet et je suis ravi de constater que vous avez du temps libre mais ce n'est pas mon cas.</p> <p>2) Je ne crois pas avoir reçu de projet de recherche de votre part m'informant de ce que vous voulez faire. Commençons par là.</p> <p>Bien à vous,</p> <p>Patrick Simon</p>		

<b>Intitulé</b> : Re: Quels prénoms les immigrés donnent-ils à leurs enfants en France ?		
<b>Date</b> : 3 juin 2019		
<b>De</b> : Jean-François Mignot	<b>À</b> : Patrick Simon	<b>Copie</b> : Baptiste Coulmont
<p>Le projet que j'ai soumis au Comité du secret figure ci-dessous. Merci d'avance de votre réponse,</p> <p>Jean-François Mignot</p>		

[Web page](#)

### Les déterminants du choix du prénom en France

En avril 2019 est paru un article scientifique important sur les prénoms donnés aux descendants d'immigrés en France (Baptiste Coulmont et Patrick Simon, « Quels prénoms les immigrés donnent-ils à leurs enfants en France ? », *Population et sociétés*, avril 2019, n°565). Cet article montre notamment qu'en France en 2008, les prénoms des descendants d'immigrés (comparés à ceux de leurs parents immigrés) tendent à converger avec ceux de la population majoritaire, ce qui peut être interprété comme un signe de volonté d'intégration — ou de volonté d'évitement des discriminations fondées sur le prénom — de la part des immigrés et de leurs descendants.

Ces résultats sont forts, mais le format court de l'article (4 pages) appelle des analyses complémentaires. Pour cela, j'ai besoin du même fichier auquel ont eu accès les auteurs, par le biais du Comité du secret. Il s'agit des données de l'enquête « Trajectoires et origines » (TeO) enrichies d'une variable qui pourrait permettre des réidentifications : le prénom (précisément : le prénom de chacun des membres du ménage des enquêtés TeO ainsi que celui de leurs enfants non corésidents). Ce projet se donne plusieurs objectifs :

- Evaluer la robustesse des observations publiées : parmi les trois méthodes qui permettent d'attribuer une origine géolinguistique à un prénom (« listes de prénoms par origines », « méthode syllabique » et « score de distance »), l'article n'utilise que la deuxième, ce qui, dans un 4-pages, est inévitable ; l'usage des autres méthodes produirait-il des résultats globalement similaires ?
- Réaliser des analyses complémentaires : *au fil des cohortes de naissance*, les immigrés et les enfants d'immigré donnent-ils à leurs enfants de plus en plus — ou de moins en moins — de prénoms qui leur sont spécifiques ? Par ailleurs, les personnes qui s'autodéclarent chrétiennes, juives et musulmanes donnent-elles à leurs enfants des prénoms de plus en plus — ou de moins en moins — proches de ceux que donnent les parents sans religion ?
- Obtenir une vision d'ensemble des déterminants du choix du prénom : dans un modèle multivarié, quels seraient les liens statistiques observés entre d'une part le type de prénom choisi (« français », « latin », « international ou anglophone », « africain », « asiatique », « arabo-musulman », « autre groupe » et « inclassable »), et d'autre part la position sociale, l'origine migratoire et l'affiliation religieuse des parents, voire leur sentiment d'avoir été discriminé ? Comment s'expliquent les évolutions du choix du prénom observées au fil des générations récentes ?

Je souhaiterais publier les résultats de ces recherches dans une revue académique.

Le terminal CASD sera dans un placard fermé d'un bureau fermé d'un lieu surveillé (mon lieu de travail, rue Pouchet).

**Intitulé :** Re: Quels prénoms les immigrés donnent-ils à leurs enfants en France ?

**Date :** 5 juin 2019

**De :** Patrick Simon

**À :** Jean-François Mignot

**Copie :** Baptiste Coulmont

Bonjour,

Votre programme de recherche correspond assez bien à ce que nous faisons avec Baptiste, puisque vous l'avez compris le 4 pages est le prélude à des publications beaucoup plus conséquentes. Aussi je vous propose d'attendre que vous ayez obtenu l'accord du comité du secret et que nous ayons soumis nos publications (à horizon de la rentrée en septembre) pour que je vous transmette une table avec un identifiant des enquêtés et la position des prénoms dans les classifications que nous avons opérées. Vous citerez bien entendu la source de ces classifications dans les publications qui seront issues de vos travaux.

Bien à vous,

Patrick Simon



<b>Intitulé</b> : Re: Quels prénoms les immigrés donnent-ils à leurs enfants en France ?		
<b>Date</b> : 6 juin 2019		
<b>De</b> : Jean-François Mignot	<b>À</b> : Patrick Simon	<b>Copie</b> : -
<p>Bonjour,</p> <p>Merci de votre réponse. J'attendrai septembre sans problème, et il va sans dire que si dans une publication j'utilise votre méthodologie syllabique de classification des prénoms je citerai non seulement votre Pop&amp;Soc (classification regroupée) mais aussi Humpert et Schneiderheinze (classification détaillée). Je vous recontacterai donc courant septembre.</p> <p>Bien cordialement,</p> <p>Jean-François Mignot</p>		

<b>Intitulé</b> : Quels prénoms les immigrés donnent-ils à leurs enfants en France ?		
<b>Date</b> : 23 septembre 2019		
<b>De</b> : Jean-François Mignot	<b>À</b> : Patrick Simon	<b>Copie</b> : -
<p>Bonjour,</p> <p>Comme convenu en juin, je me permets de vous contacter pour que vous me communiquiez la double table de passage permettant de passer de chaque prénom TeO à son origine linguistique détaillée (Humpert &amp; Schneiderheinze 2000), et de cette origine linguistique détaillée à son origine linguistique regroupée (Coulmont et Simon 2019).</p> <p>Merci d'avance et bonne rentrée à vous,</p> <p>Jean-François Mignot</p>		

<b>Intitulé</b> : Re: Quels prénoms les immigrés donnent-ils à leurs enfants en France ?		
<b>Date</b> : 25 septembre 2019		
<b>De</b> : Patrick Simon	<b>À</b> : Jean-François Mignot	<b>Copie</b> : -
<p>Cher collègue,</p> <p>Nous avons pris du retard dans la rédaction de l'article. Je ne manquerai de vous tenir informé de la mise à disposition des variables quand nous aurons terminé.</p> <p>Bien à vous,</p> <p>Patrick Simon</p>		

<b>Intitulé</b> : Re: Quels prénoms les immigrés donnent-ils à leurs enfants en France ?		
<b>Date</b> : 26 septembre 2019		
<b>De</b> : Jean-François Mignot	<b>À</b> : Patrick Simon	<b>Copie</b> : -
<p>Bonjour,</p> <p>Cela fait plusieurs mois que je demande la communication de cette table, et j'ai aujourd'hui le regret de constater que vous ne vous sentez pas tenu de respecter votre engagement à me les communiquer ce mois-ci.</p> <p>Je vais donc demander au directeur de <i>Pop &amp; Soc</i> si les contributeurs à sa revue sont dispensés de faire le nécessaire pour permettre la réplique de leurs résultats.</p> <p>Jean-François Mignot</p>		

❖ Échanges avec le rédacteur en chef de la revue *Population et sociétés*, Gilles Pison

<b>Intitulé :</b> Quels prénoms les immigrés donnent-ils à leurs enfants en France ?		
<b>Date :</b> 26 septembre 2019		
<b>De :</b> Jean-François Mignot	<b>À :</b> Gilles Pison	<b>Copie :</b> -
<p>Bonjour Monsieur Pison,</p> <p>J'espère que vous allez bien et que votre rentrée se passe bien.</p> <p>Je me permets de vous contacter parce que, dans le cadre d'un projet de recherche que je mène sur les déterminants du choix du prénom en France, cela fait maintenant plus de cinq mois que j'ai demandé aux auteurs du <i>Pop &amp; Soc</i> d'avril 2019 de bien vouloir me communiquer les données permettant de répliquer et prolonger leurs résultats. Ils m'avaient finalement indiqué, en juin, qu'ils me les communiqueraient en septembre, mais ils n'ont pas tenu leur engagement.</p> <p>Précisément, je disposerai bientôt (par le biais du Comité du secret statistique) de l'enquête TeO enrichie de la variable prénom, mais malgré cela je ne pourrai pas répliquer les analyses du <i>Pop &amp; Soc</i> puisque ses auteurs refusent de me communiquer la double table de passage qui permet (i) de passer du prénom à son origine linguistique détaillée (selon la méthodologie de Humpert &amp; Schneiderheinze 2000), et (ii) de cette origine linguistique détaillée à l'origine linguistique regroupée telle que présentée dans leur article (Coulmont et Simon 2019).</p> <p>En tant que rédacteur en chef du bulletin d'information scientifique qu'est <i>Pop &amp; Soc</i>, vous n'ignorez pas que les auteurs d'une publication scientifique <i>doivent</i> déontologiquement communiquer tout ce qu'ils <i>peuvent</i> légalement communiquer pour permettre de répliquer et prolonger les analyses publiées. Ainsi, lorsqu'un collègue (d'ailleurs affilié à l'Ined) m'a demandé, le 18 mai 2017, de lui communiquer le programme informatique qui lui permettrait de répliquer les résultats d'un de mes articles (« L'écart d'âge entre conjoints », <i>Revue française de sociologie</i>, 2010), je lui ai communiqué ce programme dans la journée même : rien que de très normal.</p> <p>C'est pourquoi je me permets de me tourner vers vous : accepteriez-vous de demander aux auteurs de ce <i>Pop &amp; Soc</i> de me communiquer la double table de passage et le programme informatique qui sont nécessaires pour répliquer leur étude, dans le cadre du déroulement normal de l'activité scientifique ?</p> <p>Merci d'avance et bien cordialement,</p> <p>Jean-François Mignot</p>		

<b>Intitulé :</b> Re: Quels prénoms les immigrés donnent-ils à leurs enfants en France ?		
<b>Date :</b> 26 septembre 2019		
<b>De :</b> Gilles Pison	<b>À :</b> Jean-François Mignot	<b>Copie :</b> Patrick Simon, Baptiste Coulmont
<p>Cher Jean-François Mignot,</p> <p>Merci pour votre message. J'en fait part aux auteurs de l'article en question en les mettant en copie. Je vous souhaite de trouver avec eux une formule permettant de prolonger leur étude.</p> <p>Bien cordialement,</p> <p>Gilles Pison</p>		

<b>Intitulé :</b> Re: Quels prénoms les immigrés donnent-ils à leurs enfants en France ?		
<b>Date :</b> 27 septembre 2019		
<b>De :</b> Patrick Simon	<b>À :</b> Jean-François Mignot	<b>Copie :</b> Gilles Pison, Baptiste Coulmont
<p>Cher collègues,</p> <p>J'ai pris attache avec la direction de l'Ined, notre chargé de déontologie et le service juridique pour faire le point sur votre demande et son cadre. Il apparaît que les classifications que nous avons réalisées relèvent de la propriété scientifique (un peu du même ordre qu'un brevet) et que nous devons établir une licence d'exploitation pour vous autoriser à les utiliser.</p>		

Aux fins d'établir cette licence qui définira les conditions d'exploitation, vous devez faire votre demande à la directrice, Magda Tomasini ([magda.tomasini@ined.fr](mailto:magda.tomasini@ined.fr)) qui fera suivre au service juridique.

Je suis désolé de cette procédure un peu plus compliquée que ce que j'imaginai.

Bien à vous,

Patrick Simon

**Intitulé :** Re: Quels prénoms les immigrés donnent-ils à leurs enfants en France ?

**Date :** 30 septembre 2019

**De :** Jean-François Mignot

**À :** Patrick Simon

**Copie :** Gilles Pison, Baptiste Coulmont

Bonjour,

Merci de votre message: je vais écrire à Madame Tomasini au plus vite. Cordialement,

Jean-François Mignot

❖ **Échanges avec la directrice de la publication de *Population et sociétés* et directrice de l'INED, Magda Tomasini, et avec le référent à l'intégrité scientifique de l'INED, Henri Leridon**

<b>Intitulé :</b> Quels prénoms les immigrés donnent-ils à leurs enfants en France ?		
<b>Date :</b> 30 septembre 2019		
<b>De :</b> Jean-François Mignot	<b>À :</b> Magda Tomasini	<b>Copie :</b> Gilles Pison, Patrick Simon, Baptiste Coulmont
<p>Madame la Directrice de l'Ined,</p> <p>Démographe au CNRS, je me permets de vous contacter pour obtenir les éléments d'information nécessaires à la réplique de l'article intitulé « Quels prénoms les immigrés donnent-ils à leurs enfants en France ? », publié dans <i>Pop&amp;Soc</i> en avril 2019.</p> <p>Dans le cadre d'un projet de recherche sur les déterminants du choix du prénom en France, je disposerai bientôt de la base de données de l'enquête TeO enrichie de la variable prénom, par le biais du Comité du secret statistique. Mais la réplique de l'article susmentionné nécessite au moins deux autres éléments :</p> <ul style="list-style-type: none"> <li>- d'une part, la double table de passage qui permet (i) de passer de chaque prénom de l'enquête TeO à son origine linguistique détaillée (selon la méthodologie de Humpert &amp; Schneiderheinze 2000), et (ii) de cette origine linguistique détaillée à l'origine linguistique regroupée telle que présentée dans l'article ;</li> <li>- d'autre part, le programme informatique qui a permis aux auteurs de produire leurs résultats (choisir et empiler les bases, choisir et recoder les variables, définir les champs et pondérations, et réaliser les analyses).</li> </ul> <p>Accepteriez-vous de me communiquer les éléments nécessaires à la réplique de cet article, dans le cadre du déroulement normal de l'activité scientifique ?</p> <p>Merci d'avance et bien cordialement,</p> <p>Jean-François Mignot</p>		

<b>Intitulé :</b> Re: Quels prénoms les immigrés donnent-ils à leurs enfants en France ?		
<b>Date :</b> 3 octobre 2019		
<b>De :</b> Magda Tomasini	<b>À :</b> Jean-François Mignot	<b>Copie :</b> Gilles Pison, Patrick Simon, Baptiste Coulmont
<p>Monsieur,</p> <p>vous souhaitez accéder à une base élaborée par deux chercheurs ainsi qu'à leurs programmes informatiques à des fins de répliquabilité de leur résultat, qui selon vous entre dans le cadre d'une activité scientifique normale.</p> <p>Dans un souci d'ouverture des données de la recherche, l'Ined est membre de la TGIR PROGEDO et met à disposition ses données d'enquête sur le réseau Quetelet. Cette procédure demande un travail conséquent de documentation des bases de la part des chercheurs mais vise à éviter des sollicitations intempestives des chercheurs qui, comme vous le savez, sont sur-sollicités. Elle permet également de documenter la dissémination des données produites par la recherche. La codification réalisée par MM. Coulmont et Simon entre dans cette procédure. Elle viendra enrichir la base TeO d'ici 1 an, période communément admise en SHS pour la mise à disposition des données.</p> <p>Bien cordialement</p> <p>Magda Tomasini</p>		

<b>Intitulé</b> : Re: Quels prénoms les immigrés donnent-ils à leurs enfants en France ?		
<b>Date</b> : 7 octobre 2019		
<b>De</b> : Jean-François Mignot	<b>À</b> : Magda Tomasini	<b>Copie</b> : Gilles Pison, Patrick Simon, Baptiste Coulmont
<p>Madame,</p> <p>Merci de votre message et de l'attention prêtée à ma demande. Je me félicite comme vous du souci que l'INED peut avoir quant à l'ouverture des données de recherche.</p> <p>Je me permets de préciser quelques faits.</p> <p>Tout d'abord, je ne souhaite nullement accéder à la base de données TeO proprement dite mais à la méthodologie du <i>Pop&amp;Soc</i> n°565, c'est-à-dire à une double table de passage et à un programme informatique. A ma connaissance, seules les données collectées (comme celles de l'enquête TeO) peuvent être soumises à un délai de mise à disposition : pas la méthodologie d'une publication. Je rappelle à ce propos que dans un message électronique du 5 juin 2019, Monsieur Simon m'avait écrit qu'il me transmettrait la double table de passage en septembre 2019. Il n'existe pas de raison pour laquelle un chercheur ne pourrait pas accéder à la méthodologie qui a permis de produire les résultats publiés dans un bulletin d'information scientifique comme <i>Pop&amp;Soc</i>.</p> <p>Ensuite, la question de la répliquabilité en sciences n'est pas une opinion personnelle. Elle constitue la norme dans la plupart des sciences, et il paraît important que cette répliquabilité puisse s'exercer librement, particulièrement sur des résultats spectaculaires ou contre-intuitifs qui obtiennent un écho public important. En l'occurrence, les auteurs mettent en évidence qu'en France en 2008, le prénom le plus fréquent chez les petits-fils d'immigrés du Maghreb est « Nicolas » (derrière « Yanis »). En outre, ils observent qu'au fil des générations les prénoms spécifiquement arabo-musulmans sont de moins en moins fréquents si bien qu'ils forment moins du quart des prénoms des petits-enfants d'au moins un immigré du Maghreb. Or, ces faits sont étonnants au regard des savoirs disponibles et il paraît important de procéder à cet exercice de répliquabilité.</p> <p>Enfin, sur le plan déontologique, en tant que directrice de la publication de <i>Pop&amp;Soc</i>, vous n'ignorez pas que les auteurs d'une publication scientifique <i>doivent</i> déontologiquement communiquer tout ce qu'ils <i>peuvent</i> légalement communiquer pour attester de la réalité comme de la robustesse de leurs résultats, afin de faire progresser les savoirs. Je ne pense pas que le CNRS et l'INED puissent différer au regard de ces exigences déontologiques.</p> <p>C'est pourquoi je me permets de renouveler ma demande de communication de ces deux éléments — la double table de passage permettant d'attribuer à chaque prénom une origine linguistique probable (détaillée et recodée), et le programme informatique de l'étude — qui permettront de mener à bien la répliquabilité des résultats de cette publication de l'Ined.</p> <p>Respectueusement,</p> <p>Jean-François Mignot</p>		

<b>Intitulé</b> : Intégrité scientifique à l'Ined		
<b>Date</b> : 22 octobre 2019		
<b>De</b> : Jean-François Mignot	<b>À</b> : Henri Leridon	<b>Copie</b> : -
<p>Monsieur le référent intégrité scientifique de l'Ined,</p> <p>Démographe au CNRS, je me permets de vous contacter pour obtenir les éléments d'information nécessaires à la répliquabilité de l'article intitulé « Quels prénoms les immigrés donnent-ils à leurs enfants en France ? », publié dans <i>Population et sociétés</i> en avril 2019 (<a href="https://www.ined.fr/fichier/s_rubrique/29081/565.population.societes.avril2019.immigres.prenoms.france.fr.pdf">https://www.ined.fr/fichier/s_rubrique/29081/565.population.societes.avril2019.immigres.prenoms.france.fr.pdf</a>).</p> <p>Dans le cadre d'un projet de recherche sur les déterminants du choix du prénom en France, je disposerai bientôt de la base de données de l'enquête TeO enrichie de la variable prénom, par le biais du Comité du secret statistique. Mais la répliquabilité de l'article susmentionné nécessite au moins deux autres éléments méthodologiques :</p>		

· d'une part, la double table de passage qui permet (i) de passer de chaque prénom de l'enquête TeO à son origine linguistique détaillée (selon la méthodologie de Humpert & Schneiderheinze 2000), et (ii) de cette origine linguistique détaillée à l'origine linguistique regroupée telle que présentée dans l'article ;

· d'autre part, le programme informatique qui a permis aux auteurs de produire leurs résultats (choisir et empiler les bases, choisir et recoder les variables, définir les champs et pondérations, etc., et réaliser les analyses).

Accepteriez-vous de me communiquer ces éléments, dans le cadre du déroulement normal de l'activité scientifique ?

Si je me permets de vous contacter, c'est que j'ai déjà contacté les auteurs de l'article, mais depuis avril 2019 ils refusent de me communiquer la méthodologie de leur publication. J'ai donc dû contacter le rédacteur en chef de la revue et la directrice de la publication, mais là encore, malgré mes demandes répétées, je ne parviens toujours pas à accéder à la méthodologie de l'article. Dans un message électronique du 5 juin 2019 l'un des auteurs, Monsieur Simon, m'avait pourtant écrit qu'il me transmettrait la double table de passage en septembre 2019, mais il n'a pas tenu parole.

Or, les auteurs mettent en évidence qu'en France en 2008, le prénom le plus fréquent chez les petits-fils d'immigrés du Maghreb est « Nicolas » (derrière « Yanis »). En outre, ils observent qu'au fil des générations les prénoms spécifiquement arabo-musulmans sont de moins en moins fréquents si bien qu'ils forment moins du quart des prénoms des petits-enfants d'au moins un immigré du Maghreb. Or, ces faits sont étonnants au regard des savoirs disponibles, si bien qu'il paraît important de procéder à cet exercice de réplification.

Merci d'avance pour votre aide. En tant que référent intégrité scientifique de l'Ined, vous êtes la seule personne qui puisse encore me permettre d'accéder à la méthodologie de cet article, indispensable pour répliquer les résultats publiés. Malgré mes échanges avec les auteurs et la hiérarchie de l'Ined, je ne connais toujours pas les raisons pour lesquelles un chercheur ne pourrait pas accéder à la méthodologie qui a permis de produire les résultats publiés dans un bulletin d'information scientifique comme *Pop&Soc*.

Respectueusement,

Jean-François Mignot

**Intitulé :** Re: Quels prénoms les immigrés donnent-ils à leurs enfants en France ?

**Date :** 14 novembre 2019

**De :** Magda Tomasini

**À :** Jean-François Mignot

**Copie :** -

Bonsoir monsieur

vous trouverez en PJ un avis du référent intégrité scientifique de l'Ined que j'ai saisi sur le sujet qui vous préoccupe.

Vous en souhaitant bonne lecture, bien cordialement,

Magda Tomasini

**Pièces attachées :**

**Réponse à Jean-François Mignot (PROJET)**

(H. Leridon, référent intégrité scientifique, 16 octobre 2019)

La directrice de l'Ined, Magda Tomasini, m'a transmis votre demande relative à l'article de MM Coulmont et Simon paru dans *Population et Sociétés* d'avril 2019, et m'a invité à vous répondre en tant que Référent intégrité scientifique.



Je note d'abord que vous entretenez une confusion entre les principes de « reproduction » et de « réplification ». Seul le premier implique de disposer à la fois des données utilisées et de la liste détaillée de tous les processus de traitement opérés, afin de reproduire *exactement* le travail des auteurs. Ce n'est ni plus ni moins qu'un contrôle du travail effectué par ceux-ci. La répliquabilité consiste, elle, à tenter de reproduire l'ensemble de la démarche des auteurs, pas toujours avec les données d'origine (on peut utiliser des données « semblables ») ni en utilisant exactement les mêmes processus et programmes. Il s'agit, en pratique, de savoir si le résultat publié est généralisable, ou s'il faut le considérer comme de portée limitée, parce que non parfaitement répliquable. Il n'est donc pas cohérent de demander la communication des « éléments nécessaires à la réplification de cet article, dans le cadre du déroulement normal de l'activité scientifique », alors que les éléments que vous demandez visent à une exacte reproduction de l'étude.

Les recommandations d'ouverture de la science par un accès facilité aux publications, d'une part, et aux données d'autre part, sont de mieux en mieux cadrées. La communication des procédures et programmes utilisés ne l'est pas encore. Pour le moment, ce sont principalement les revues qui doivent se porter garantes des bonnes pratiques des auteurs d'articles publiés. Ce que vous appelez le « déroulement normal de l'activité scientifique », et votre empressement à obtenir la réponse à vos demandes, ne renvoient en fait à aucune pratique établie, et les réponses ne peuvent être apportées qu'au cas par cas.

A titre d'exemple, la revue (en ligne) *Demographic Research* vient de lancer une nouvelle catégorie d'articles, les « replications » (cf. J. Bajak, *Demog. Res.*, 41, 32, 19 oct. 2019) :

*Replications are carefully prepared and executed studies aimed at replicating other results on the same topic, published either in Demographic Research or elsewhere in the literature. The replication can be carried out in the same context as the original work or in a different context, and the results should illuminate and reflect the similarities to and differences from the original study.*

Cette approche montre que l'on peut chercher à « répliquer » une étude sans disposer des données originales, mais en s'efforçant de suivre la démarche de l'auteur de l'étude originale.

Dans le cas qui nous occupe, il me semble qu'un accès au fichier source (incluant les prénoms) et des informations sur les principes du codage mis en œuvre (travail réalisé par une entreprise spécialisée, utilisant un algorithme connu) devraient vous permettre de répliquer dans de bonnes conditions le travail publié. Si vous parvenez à un résultat différent, cela prouvera que ces résultats sont sensibles aux modes de codage et d'analyse, ce qui enrichira la compréhension du problème, et ceci dans le « cadre normal de l'activité scientifique ».

Quant à la table de correspondance élaborée par les auteurs entre prénom et catégorie culturelle, qui constitue une donnée nouvelle, elle sera incorporée à la base TéO au plus tard un an après la première publication utilisant ces nouvelles données, soit en avril 2020.

## ❖ Échanges avec l'INED impliquant la Mission à l'intégrité scientifique du CNRS

Pour respecter l'anonymat des membres de la Mission à l'intégrité scientifique (MIS) du CNRS, les individus impliqués sont ici dénommés « MIS1 », « MIS2 » et « MIS3 ». De même pour le médiateur intégrité scientifique de l'ENS Paris-Saclay, dénommé « MIS-ENS ». De même pour un agent de l'INSEE, dénommé « INSEE1 ».

<b>Intitulé</b> : Quels prénoms les immigrés donnent-ils à leurs enfants en France ?		
<b>Date</b> : 17 février 2020		
<b>De</b> : Jean-François Mignot	<b>À</b> : Patrick Simon, Baptiste Coulmont	<b>Copie</b> : MIS1, MIS2, Henri Leridon
<p>Chers collègues,</p> <p>Les référents intégrité scientifique du CNRS (en copie de ce message) ayant eu l'occasion d'échanger avec le référent intégrité scientifique de l'INED (en copie aussi), je me permets de vous recontacter pour vous demander les éléments qui me permettraient de comprendre la méthodologie qui vous a permis d'aboutir aux résultats publiés dans le <i>Pop&amp;Soc</i> d'avril 2019. Cela me permettrait aussi de reproduire à l'identique vos analyses sur la même base de données (TeO enrichie de la variable prénom).</p> <p>Précisément, je souhaiterais obtenir :</p> <ul style="list-style-type: none"> <li>· Le <b>programme informatique</b> qui vous a permis d'obtenir, à partir des bases TeO brutes, chacun des résultats publiés, y compris les graphiques et tableaux ; cela me permettra de savoir quelles tables TeO vous avez choisi et comment vous les avez empilées, comment vous avez sélectionné les individus, comment vous avez construit vos variables, comment vous avez pondéré vos individus, et quelles analyses précisément vous avez réalisées pour aboutir aux résultats publiés ;</li> <li>· La <b>double table de passage</b> qui, à chaque prénom, associe (i) son origine linguistique détaillée selon la méthodologie de Humpert et Schneiderheinze (2000), et (ii) son origine linguistique regroupée en 7 ou 8 catégories selon la méthodologie de votre article (« français », « arabo-musulman », « latin », « africain », « asiatique », « internationaux ou anglophones », « autre groupe » et/ou « inclassable ») ; cela me permettra de savoir, par exemple, comment les prénoms de votre Tableau page 3 (Yanis, Nicolas, Mehdi, Sarah, Inès, Lina, etc.) sont classifiés par Humpert et Schneiderheinze, et par vous.</li> </ul> <p>En vous remerciant par avance,</p> <p>Jean-François Mignot</p>		

<b>Intitulé</b> : Re: Quels prénoms les immigrés donnent-ils à leurs enfants en France ?		
<b>Date</b> : 3 mars 2020		
<b>De</b> : Jean-François Mignot	<b>À</b> : Patrick Simon, Baptiste Coulmont	<b>Copie</b> : MIS1, MIS2, Henri Leridon
<p>Chers collègues,</p> <p>Je me permets de vous écrire de nouveau pour vous faire part de mon incompréhension, car je n'ai pas reçu de votre part de réponse à mon message du 17 février, ni même un simple accusé de réception. Pourtant, les référents à l'intégrité scientifique du CNRS m'ont indiqué que M. Leridon, dans sa réponse du 24 janvier à la Mission à l'Intégrité Scientifique du CNRS, écrivait : « Des éclaircissements peuvent, sans attendre, être demandés aux auteurs comme je le mentionnais aussi : « Si nécessaire, vous pouvez demander aux auteurs de préciser davantage certaines modalités de leur démarche ». L'auteur a été informé de cette éventualité. »</p> <p>En vous remerciant par avance de votre réponse,</p> <p>Jean-François Mignot</p>		

<b>Intitulé</b> : Re: Quels prénoms les immigrés donnent-ils à leurs enfants en France ?		
<b>Date</b> : 4 mars 2020		
<b>De</b> : Baptiste Coulmont	<b>À</b> : Jean-François Mignot	<b>Copie</b> : Patrick Simon, MIS1, MIS2, Henri Leridon
<p>Bonjour,</p> <p>En effet je n'ai pas accusé réception de votre mail.</p> <p>Je n'étais pas destinataire des échanges entre les référents intégrité, et je suivrai la décision de l'INED.</p> <p>Cordialement,</p> <p>B. Coulmont</p>		

<b>Intitulé</b> : Re: Quels prénoms les immigrés donnent-ils à leurs enfants en France ?		
<b>Date</b> : 7 mars 2020		
<b>De</b> : Patrick Simon	<b>À</b> : Jean-François Mignot	<b>Copie</b> : Baptiste Coulmont, MIS1, MIS2, Henri Leridon, Magda Tomasini
<p>Cher collègue,</p> <p>Vous réitérez une demande que vous avez formulée il y a quelques mois et à laquelle la direction de l'Ined a répondu, en substance, que les classifications des prénoms seront ajoutées à la base TeO diffusée par Quételet un an après la publication du Population &amp; Sociétés d'avril 2019, c'est-à-dire en avril 2020. La date est maintenant très proche. Vous pourrez reproduire à l'identique toutes les analyses que vous souhaiterez en utilisant la base enrichie des prénoms dont vous disposez au CASD en la combinant avec celle de Quételet.</p> <p>Vous posez dans votre message des questions plus précises sur les tables TeO que nous avons utilisées et les pondérations. Voici quelques éléments à ce sujet : nous avons utilisé la base individus associée à la base enfants, ce qui nous a permis d'observer les trois générations impliquées dans nos analyses. Il n'y a qu'une pondération à utiliser pour TeO qui s'appelle poids, comme vous le découvrirez en lisant la documentation relative à l'enquête. Souhaitez-vous que nous vous expliquions comment apparier ces bases ? Je ne doute pas que vous ayez dans votre laboratoire des ingénieurs de recherche du CNRS capables d'effectuer ces opérations, mais si besoin Baptiste et moi-même pourrons vous donner des conseils.</p> <p>Vous demandez comment nous avons construits nos variables : la plupart des variables construites sont renseignées dans le dictionnaire des codes diffusé avec la base Quételet. L'Insee et l'équipe de recherche a passé beaucoup de temps pour établir les spécifications et les renseigner pour que des utilisateurs extérieurs, dont vous faites partie, puissent exploiter la base TeO sans coût d'entrée excessif. De nombreux utilisateurs nous ont remercié pour ce travail de mise à disposition et c'est la fierté de nos institutions d'avoir fourni à la communauté scientifique un outil de recherche collective en un temps relativement bref. S'il y a des variables spécifiques sur lesquelles vous souhaitez des informations supplémentaires, n'hésitez pas à nous les indiquer précisément par leur nom ou leur caractéristique, nous vous fournirons leur spécification en SAS ou en Stata.</p> <p>Quelques précisions s'agissant des tables de passage entre les prénoms et les classifications effectuées par Humpert &amp; Schneiderheinze (H&amp;S). L'Ined a demandé à H&amp;S, qui est une entreprise privée, de produire des classifications onomastiques suivant des typologies d'origine. Ces classifications ont été achetées à H&amp;S et sont la propriété de l'Ined. Vous avez écrit directement à H&amp;S en leur demandant de vous fournir le fichier qu'ils nous ont vendu, les mettant dans une situation embarrassante. Il vous est parfaitement loisible de recourir aux services de cette entreprise à partir de la liste de prénoms dont vous disposez pour leur demander également une classification, ce qui suppose de les rémunérer et vous, ou le CNRS par votre intermédiaire, serez propriétaire de la dite table de passage. Dans l'état actuel du droit de la propriété, je ne vois pas selon quelle norme H&amp;S ou l'Ined devraient vous fournir un équipement scientifique pour lequel un contrat a été signé. Comme la direction de l'Ined vous l'a indiqué, et que je vous ai rappelé au début de ce message, les classifications vous seront accessibles en avril et vous pourrez indirectement reconstituer la table de passage.</p>		

J'espère que ces éléments répondent à vos demandes et nous attendons avec intérêt de lire vos analyses sur les prénoms, ainsi que la reproduction de nos propres analyses publiées dans le bulletin Population & Sociétés d'avril 2019. Nous sommes bien entendu disposés à répondre à vos questions précises sur les variables lorsque vous aurez commencé les traitements et, éventuellement, rencontré des difficultés à développer vos analyses.

Dans l'attente, je vous prie de croire, cher collègue, en mes sentiments les meilleurs.

Patrick Simon

<b>Intitulé</b> : Re: Quels prénoms les immigrés donnent-ils à leurs enfants en France ?		
<b>Date</b> : 9 mars 2020		
<b>De</b> : Jean-François Mignot	<b>À</b> : Patrick Simon, Baptiste Coulmont	<b>Copie</b> : MIS1, MIS2, Henri Leridon, Magda Tomasini
Chers collègues, Je vous recontacterai, si c'est utile, en avril. Jean-François Mignot		

<b>Intitulé</b> : Re: Quels prénoms les immigrés donnent-ils à leurs enfants en France ?		
<b>Date</b> : 16 avril 2020		
<b>De</b> : Jean-François Mignot	<b>À</b> : Patrick Simon, Baptiste Coulmont	<b>Copie</b> : MIS1, MIS2, Henri Leridon, Magda Tomasini
Chers collègues, J'espère que vous et vos proches vous portez bien. Je me permets de vous recontacter suite à votre dernier message, qui indiquait : « la direction de l'Ined a répondu, en substance, que les classifications des prénoms seront ajoutées à la base TeO diffusée par Quêtelet un an après la publication du Population & Sociétés d'avril 2019, c'est-à-dire en avril 2020 ». Cela fait désormais plus d'un an que vous avez publié votre <i>Pop&amp;Soc</i> (c'était le 10 avril 2019), mais à ce jour les classifications des prénoms ne sont toujours pas disponibles avec la base TeO diffusée sur le réseau Quetelet. Ces classifications ne sont d'ailleurs pas près d'être disponibles sur le réseau Quetelet, puisqu'à ce jour l'Adisp ne les a pas reçues. C'est pourquoi je reviens vers vous pour vous demander de bien vouloir, conformément à votre engagement, me communiquer lesdites classifications, ou tout simplement votre double table de passage, que ce soit par message électronique ou WeTransfer ou tout autre moyen adapté. Cordialement, et en vous souhaitant une bonne fin de semaine, Jean-François Mignot		

<b>Intitulé</b> : Re: Quels prénoms les immigrés donnent-ils à leurs enfants en France ?		
<b>Date</b> : 19 avril 2020		
<b>De</b> : Patrick Simon	<b>À</b> : Jean-François Mignot	<b>Copie</b> : Baptiste Coulmont, MIS1, MIS2, Henri Leridon, Magda Tomasini, INSEE1
Cher collègue, J'espère que ce message vous trouve en bonne santé ainsi que vos proches. Vous vous inquiétez de la disponibilité à Quêtelet de la base TeO augmentée des variables de classification des prénoms. Je comprends votre impatience, mais il ne vous aura pas échappé que les conditions de confinement ont considérablement ralenti les processus de recherche et administratif. Nous avons à l'Ined préparé les fichiers permettant l'enrichissement de la base et il nous reste quelques détails à régler. Mais le processus est un peu plus complexe que celui que vous semblez imaginer. C'est en effet l'Insee qui est chargé de la diffusion des données TeO. Nous allons donc transmettre ces fichiers à l'Insee la semaine prochaine et de là le protocole interne de l'Insee, tel qu'il est actif en période de		

confinement, s'appliquera pour transmettre les fichiers à l'ADISP. Je mets mon collègue de l'Insee [INSEE1] en copie de ce message car il a la responsabilité de la mise à disposition à l'Insee. Il vous dira mieux que moi les délais de traitement actuel. J'insiste sur les conditions particulières provoquées par le confinement car l'accès aux serveurs sécurisés de l'Insee depuis l'extérieur n'est sans doute pas activable de n'importe où et cela peut impacter l'activité des services concernés. Eu égard aux impératifs de ces services, il n'est pas certain que la mise à disposition des données constitue une priorité.

Concernant votre demande de transmission de données confidentielles, à savoir la table de passage des prénoms à leur position dans la classification, nous vous avons indiqué qu'elle ne vous sera pas directement transmise, et le cas échéant certainement pas via une plateforme non sécurisée comme We Transfer. Je laisse les correspondants éthique du CNRS et de l'Ined apprécier cette demande pour le moins étonnante.

Je ne doute pas que vous partagiez nos inquiétudes et incertitudes dans cette période de mise à l'épreuve et que vous saurez vous armer de patience.

Bien à vous,

Patrick Simon

**Intitulé :** Re: Quels prénoms les immigrés donnent-ils à leurs enfants en France ?

**Date :** 20 avril 2020

**De :** Jean-François Mignot

**À :** Patrick Simon, Baptiste Coulmont

**Copie :** Baptiste Coulmont, MIS1, MIS2, Henri Leridon, Magda Tomasini, INSEE1, MIS3

Chers collègues,

Le fait que le confinement puisse ralentir certaines procédures de l'INSEE est parfaitement compréhensible. Ce qui l'est peut-être un peu moins, ce sont les raisons pour lesquelles, depuis plus d'une année maintenant, vous refusez de communiquer les éléments méthodologiques qui permettraient de reproduire les résultats que vous avez publiés dans votre article de 2019, alors même qu'**aucun de ces éléments méthodologiques ne comprend des données individuelles et n'est donc soumis au secret statistique.**

En effet je ne vous ai jamais demandé l'accès aux tables de données de l'enquête TeO, que ce soit directement ou par le biais du réseau Quetelet ; et pour cause : j'y ai déjà accès par le biais du CASD (<https://www.casd.eu/project/les-determinants-du-choix-du-prenom-en-france/>). Ce que je vous demande, c'est votre double table de passage de structure suivante :

**Double table de passage de chaque prénom à son origine**

	Catégorie linguistique <b>détaillée</b> (en plusieurs dizaines de modalités) selon la méthodologie de Humpert et Schneiderheinze (2000)	Catégorie linguistique <b>regroupée</b> (en 7 ou 8 modalités) selon la méthodologie de Coulmont et Simon (2019)
Prénom 1		
Prénom 2		
...		
Prénom 18 163		

Comme chacun peut le constater, cette double table de passage ne porte pas sur des individus (ni a fortiori sur des individus enquêtés par la statistique publique et susceptibles d'être ré-identifiables et donc protégés par le secret statistique), mais sur des prénoms. Par conséquent, pourquoi refusez-vous de me communiquer cette double table de passage, qui n'est aucunement soumise au secret statistique et ne requiert pas de passer par l'INSEE ou le réseau Quetelet pour être mise à disposition d'un collègue ? Qu'est-ce qui vous empêche de me communiquer cette double table de passage alors

qu'elle est légalement communicable, y compris par des moyens standard comme le message électronique ou WeTransfer ?

D'autre part, pourquoi refusez-vous de me communiquer le programme informatique qui, une fois appliqué aux bases TeO (auxquelles j'ai accès par le biais du CASD), permettrait de reproduire les résultats de votre publication ?

Depuis un an que vous refusez de me communiquer ces deux éléments méthodologiques que la plupart des chercheurs auraient communiqué en moins d'une semaine, vous n'avez jamais cru bon de justifier votre refus.

En vous remerciant par avance pour vos éclaircissements,  
Jean-François Mignot

[Le message ci-dessous m'est probablement parvenu par erreur, suite à une mauvaise manipulation informatique de Patrick Simon.]

<b>Intitulé :</b> Re: Quels prénoms les immigrés donnent-ils à leurs enfants en France ?		
<b>Date :</b> 25 mai 2020		
<b>De :</b> Patrick Simon	<b>À :</b> Jean-François Mignot	<b>Copie :</b> -
<p>mais en septembre je n'avais pas fini l'article</p> <p>-----</p> <p>Le 25/09/2019 à 22:11, SIMON Patrick a écrit :</p> <p>Cher collègue,</p> <p>Nous avons pris du retard dans la rédaction de l'article. Je ne manquerai de vous tenir informé de la mise à disposition des variables quand nous aurons terminé.</p> <p>Bien à vous,</p> <p>Patrick Simon</p> <p>-----</p> <p>Le 23/09/2019 à 09:56, Jean-François Mignot a écrit :</p> <p>Bonjour,</p> <p>Comme convenu en juin, je me permets de vous contacter pour que vous me communiquiez la double table de passage permettant de passer de chaque prénom TeO à son origine linguistique détaillée (Humpert &amp; Schneiderheinze 2000), et de cette origine linguistique détaillée à son origine linguistique regroupée (Coulmont et Simon 2019).</p> <p>Merci d'avance et bonne rentrée à vous,</p> <p>Jean-François Mignot</p>		

[16 juin 2020 : Discussion orale avec Baptiste Coulmont]

<b>Intitulé :</b> Quels prénoms les immigrés donnent-ils à leurs enfants en France ?		
<b>Date :</b> 01/12/2020		
<b>De :</b> Jean-François Mignot	<b>À :</b> Magda Tomasini	<b>Copie :</b> Henri Leridon, Baptiste Coulmont, Patrick Simon, Gilles Pison, MIS1, MIS2, MIS3, MIS-ENS
<p>Madame la directrice de l'INED,</p> <p>Plus d'un an et demi après la publication du Pop&amp;Soc n°565, les informations relatives aux prénoms ajoutées à l'enquête TeO par l'INED ont enfin été rendues disponibles aux chercheurs.</p>		



Toutefois il apparaît que la nomenclature de prénoms ajoutée à TeO n'est pas celle utilisée dans l'article, comme en témoigne la disparité entre d'une part les modalités de la nomenclature de prénoms ajoutée par l'INED (en attaché), et d'autre part les modalités de la nomenclature de l'article (encadré page 2 de l'article : [https://www.ined.fr/fichier/s\\_rubrique/29081/565.population.societes.avril2019.immigres.prenoms.france.fr.pdf](https://www.ined.fr/fichier/s_rubrique/29081/565.population.societes.avril2019.immigres.prenoms.france.fr.pdf)). Manquent, par exemple, les modalités relatives aux prénoms asiatiques ou africains. Dois-je comprendre que l'INED ne compte finalement pas respecter les engagements pris dans votre message du 3 octobre 2019 et dans les messages suivants de Messieurs Leridon, Coulmont et Simon ? Comment l'attitude de l'INED est-elle compatible avec une recherche ouverte, transparente et reproductible ?

Jean-François Mignot

<b>Intitulé :</b> Quels prénoms les immigrés donnent-ils à leurs enfants en France ?		
<b>Date :</b> 01/12/2020		
<b>De :</b> Magda Tomasini	<b>À :</b> Jean-François Mignot	<b>Copie :</b> Henri Leridon, Baptiste Coulmont, Patrick Simon, Gilles Pison, MIS1, MIS2, MIS3, MIS-ENS
<p>Cher monsieur,</p> <p>je souhaiterais que vous vous adressiez à moi sur un autre ton : je ne suis pas en charge directement de la collecte et de la diffusion de variables d'enquête.</p> <p>Je vous invite à vous adresser aux producteurs d'enquête pour comprendre le sens des variables disponibles.</p> <p>Bonne journée</p>		

<b>Intitulé :</b> Re : Quels prénoms les immigrés donnent-ils à leurs enfants en France ?		
<b>Date :</b> 02/12/2020		
<b>De :</b> Jean-François Mignot	<b>À :</b> Baptiste Coulmont, Patrick Simon	<b>Copie :</b> Magda Tomasini, Henri Leridon, Gilles Pison, MIS1, MIS2, MIS3, MIS-ENS
<p>Bonjour Messieurs Simon et Coulmont,</p> <p>Je viens d'avoir accès à la nomenclature de prénoms que vous avez communiquée à l'Insee, pour communication au réseau Quetelet et au CASD et pour mise à disposition des chercheurs. Malheureusement, elle ne correspond pas à la nomenclature que vous avez utilisée dans votre article, et je ne comprends pas comment on peut déduire de cette classification celle que vous utilisez dans votre article. Pourriez-vous m'éclairer sur ce point ? Et plus simplement, pourquoi n'avez-vous pas directement mis à disposition des chercheurs la classification précisément utilisée dans votre article ?</p> <p>Finalement, au bout de 18 mois d'attente, je trouve regrettable que la méthodologie de votre article, que la plupart des chercheurs auraient communiquée en moins d'une semaine, ne soit toujours pas disponible pour la communauté.</p> <p>Dans l'attente de vos réponses. Cordialement,</p> <p>Jean-François Mignot</p>		

<b>Intitulé :</b> Re : Quels prénoms les immigrés donnent-ils à leurs enfants en France ?		
<b>Date :</b> 04/12/2020		
<b>De :</b> Patrick Simon	<b>À :</b> Jean-François Mignot	<b>Copie :</b> Baptiste Coulmont, Magda Tomasini, Henri Leridon, Gilles Pison, MIS1, MIS2, MIS3, MIS-ENS
<p>Cher collègue,</p> <p>Nous avons mis à la disposition de la communauté les classifications de base à partir desquelles nous avons travaillé, ce qui est usuel dans les mises à disposition des enquêtes. C'est sans doute parce que vous n'en avez pas fait que vous n'êtes pas familier des normes de diffusion des variables. Vous imaginez bien que les équipes de recherche ayant exploité l'enquête Trajectoires et Origines ont produit un nombre incalculable de variables construites qui ne sont pas consignées dans les répertoires disponibles au CASD ou à Quêtelet. Du reste, les variables que vous produirez vous mêmes ne seront pas plus communiquées que celles de 320 chercheur.e.s qui ont demandé et obtenu la base de données. Vous nous faites un mauvais procès en complotisme en nous reprochant de ne pas diffuser les outils de travail que nous avons produit pour écrire notre article. Mais comme vous êtes très insistant, voici le format que nous avons utilisé pour la nomenclature des prénoms de l'article, qui s'applique à Cat1 et pas à Cat0 que nous n'avons pas utilisé.</p> <p>Value onomasb  100,126,128,132,139,142,148,149,157,611,121,122,125,138,144,154,152,160,162,165,166,169,430,441,651='divers'  129='France'  134,137,153,161,323,327,661='Europe Sud, latino'  135,168,368,696,697='Anglo et inter'  163='Turquie'  221,252,224,622,422,425,439,450,641,643,662='Arabo-musulman'  231-251,262-283,621,623='Af sub'  427,431-437,442,446,449,457-479,644,645='Asie'  0,698='unclassified';</p> <p>Bon travail et n'oubliez pas de nous envoyer le résultat de vos contre-analyses, puisque tel est votre projet. Nous serons heureux de vous lire.</p> <p>Bien à vous,  Patrick Simon</p>		

<b>Intitulé :</b> Re : Quels prénoms les immigrés donnent-ils à leurs enfants en France ?		
<b>Date :</b> 05/12/2020		
<b>De :</b> Jean-François Mignot	<b>À :</b> Baptiste Coulmont, Patrick Simon	<b>Copie :</b> Magda Tomasini, Henri Leridon, Gilles Pison, MIS1, MIS2, MIS3, MIS-ENS
<p>Messieurs Coulmont et Simon,</p> <p>Vous avez communiqué à la communauté scientifique une nomenclature bidon (variable CAT0), mais comme j'ai été « très insistant » vous venez de me faire l'honneur et le privilège de me communiquer une partie de la nomenclature que vous utilisez dans votre article (à partir de CAT1). Comment dois-je vous remercier ?</p> <p>La nomenclature que vous venez de communiquer comprend, d'après vos indications, des prénoms « <b>divers</b> » qui sont absents de la nomenclature de votre article : merci de m'indiquer comment vous avez procédé.</p> <p>Quant au « mauvais procès » que je vous intente, figurez-vous que, de mon côté, la méthodologie de mes publications est disponible sur demande de n'importe quel collègue et sans délai. C'est ce dont témoigne, notamment, le fait que le 18 mai 2017, quand un collègue (d'ailleurs associé à l'INED) m'a demandé de lui communiquer le programme informatique nécessaire à la reproduction d'un de mes articles, je lui ai communiqué ledit programme de façon inconditionnelle et dans la journée même : rien que de très normal dans le cadre de l'activité scientifique.</p>		

Un mauvais procès consisterait plutôt, par exemple, à vous reprocher d'avoir catégorisé les prénoms arméniens (modalité 422) comme « arabo-musulmans ». Cette catégorisation, certes inédite et qui ne manque pas de piquant, porte sur un effectif de prénoms trop faible pour risquer d'impacter vos résultats.

Enfin, avant de mettre un terme aux échanges si gratifiants qui nous lient depuis plus de 18 mois, reste une dernière question : quand allez-vous me communiquer le **programme informatique** que je vous demande depuis mon message du 30 septembre 2019 ? Que vais-je devoir faire, cette fois-ci, pour que votre magnanimité vous pousse jusqu'à respecter la déontologie de notre métier ?

Jean-François Mignot

**Intitulé :** Re : Quels prénoms les immigrés donnent-ils à leurs enfants en France ?

**Date :** 18/12/2020

**De :** Jean-François Mignot

**À :** Magda Tomasini

**Copie :** Baptiste Coulmont,  
Patrick Simon, Henri Leridon,  
Gilles Pison, MIS-ENS

Madame la directrice de l'INED,

Cela fait désormais plus de 20 mois que l'INED a publié le Pop&Soc n°565, et plus de 20 mois aussi que j'ai contacté ses auteurs pour connaître et comprendre la méthodologie qui a pu leur permettre de produire les résultats qu'ils ont publiés.

Comme vous pouvez le constater, les auteurs continuent de refuser de répondre aux questions méthodologiques élémentaires que je me permets de leur poser. Les bribes d'information qu'ils ont communiquées sont inutilisables, tant qu'ils refusent de préciser comment ils sont passés de la nomenclature qu'ils ont finalement communiquée (à partir de CAT1) à la nomenclature de leur article, qui ne comprend pas de catégorie « divers ». Quant à leur programme informatique, ils n'ont jamais justifié leur refus de le communiquer.

Madame la directrice, est-ce que l'INED assume de refuser de communiquer la méthodologie de cette publication ? Si non, quand comptez-vous mettre un terme à cette mascarade ?

Jean-François Mignot

## Annexe 4B : Les programmes informatiques communiqués par les auteurs

Cette Annexe 4B, associée au Chapitre 4, p. 15, comprend les 3 programmes informatiques R qui m'ont finalement été communiqués par Coulmont et Simon par le biais du CASD le 6 janvier 2021, respectivement intitulés « 01-importation », « 02-recodage » et « 03-tables ».

### ❖ 1<sup>er</sup> programme informatique communiqué par les auteurs (01-importation)

```
#####
#
# importation et nettoyage des bases TeO
#
#####

# installation des packages
#####
# install.packages("reshape2")
# install.packages("haven")
# install.packages("devtools") # pour installer la version github de ggplot2
# library(devtools)
# dev_mode(on=T)
# install_github("hadley/ggplot2")

# chargement des packages nécessaires
#####

library(reshape2)
# utilisation du package haven (qui fonctionne pour l'importation de fichiers SAS)
library(haven) # windows shortcut CTRL+R
library(plyr)
library(ggplot2)
library(stringi) # pour enlever les accents

# répertoire de travail pour l'importation des bases
#####

#setwd("REPERTOIRE")

#####
# chargement des tables individus, THL et enfants
#####

# individus
#####

# base TeO non anonymisée, avec les prénoms
indiv<-read_sas("indiv_moa.sas7bdat") # 1544 variables, 21761 individus
# attention, les variables vides sont codées "" et non pas "NA"

# base TeO anonymisée avec variables construites
indiv_final <- read_sas("base_indiv_28052012.sas7bdat")

#### sauver en RDATA et pas en CSV !

# thl
#####
thl <- read_sas("thl_moa.sas7bdat") # 106 variables, 75121 individus

# enfants
#####

enfants <- read_sas("enfants_moa.sas7bdat") # 30 variables, 31476 individus

#####
#
# Changement du répertoire de travail
#
# pour les nettoyages et calculs d'indices
```

```

#
#####

#setwd("REPertoire")

#####
#
# Le prÃ©nom
#
#####

# Examen de la variable indiv$Prenom
# limitÃ©e Ã 15 caractÃ©res
# donc les 2e et 3e prÃ©noms sont souvent coupÃ©s
# comme "Nicole Catherin"
# prÃ©noms composÃ©s comprennent rarement un "-" (une cinquantaine)
# prÃ©noms composÃ©s "thi X Y" assez frÃ©quents (71 "thi" en premier prÃ©nom)
# stratÃ©gie : ne garder que le premier prÃ©nom,
# y compris quand prÃ©nom composÃ© avec un tiret

# Prenom1 : premier prÃ©nom
#####

indiv$Prenom1<-gsub('-',',',indiv$Prenom)
indiv$Prenom1<-gsub('[(.*)^][A-Z]','',indiv$Prenom1)
indiv$Prenom1<-gsub(' ',',',indiv$Prenom1) # reste parfois un espace final
indiv$Prenom1 <- stri_trans_general(indiv$Prenom1, "Latin-ASCII") # on enleve les quelques accents

enfants$Prenom1<-gsub('-',',',enfants$E_PRENOM)
enfants$Prenom1<-gsub('[(.*)^][A-Z]','',enfants$Prenom1)
enfants$Prenom1<-gsub(' ',',',enfants$Prenom1)
enfants$Prenom1 <- stri_trans_general(enfants$Prenom1, "Latin-ASCII") # on enleve les quelques accents

thl$Prenom1<-gsub('-',',',thl$PRENOM)
thl$Prenom1<-gsub('[(.*)^][A-Z]','',thl$Prenom1)
thl$Prenom1<-gsub(' ',',',thl$Prenom1)
thl$Prenom1 <- stri_trans_general(thl$Prenom1, "Latin-ASCII") # on enleve les quelques accents

# crÃ©ation d'un identifiant unique : IDENTIFIANT
# pour servir de clÃ© d'appariement entre les bases
#####

indiv$IDENTIFIANT<-paste(indiv$IDENT,indiv$Ne,sep="-")
indiv_final$IDENTIFIANT <- paste(indiv_final$IDENT,indiv_final$Ne,sep="-")
thl$IDENTIFIANT<-paste(thl$IDENT,thl$NOI,sep="-")
# attention, 7924 enfants n'ont pas de E_NOI_THL car ils
# ne rÃ©sident pas avec leur parent enquÃªtÃ© (EGO)
enfants$IDENTIFIANT<-paste(enfants$IDENT,enfants$E_NOI_THL,sep="-")

# trouver, pour chaque enfant, le lieu de naissance de ses parents
#####
# la stratÃ©gie suivie consiste Ã enrichir le THL, qui contient EGO et une partie des enfants
# on commence par indiquer, pour tous les individus des mÃ©nages du THL
# quel est le numÃ©ro de EGO (personne enquÃªtÃ©e, dans le mÃ©nage)
# c'est le plus souvent l'individu NOI nÃ©1 mais pas toujours

m<-match(thl$IDENT,indiv$IDENT)
thl$EGO<-indiv$Ne[m] # le numÃ©ro de EGO dans le mÃ©nage Ã©tudiÃ© (c'est le plus souvent 1)

# puis on indique le sexe d'EGO (pour rÃ©partir pÃ©re et mÃ©re des enfants)
thl$EGO_SEXE<-indiv$Sexe[m]

# on indique pour chaque individu du THL l'identifiant de leurs parents
# si ces individus ont des parents dans le THL,
# sinon Ã§a donne un identifiant inutilisable
# attention MER2E (NOI de la mere) est indiquÃ© "4" et non pas "04"
# pour avoir une clÃ© d'appariement, il faut ajouter les "0"

thl$MER2E <- paste("0",thl$MER2E,sep="")
thl$PER2E <- paste("0",thl$PER2E,sep="")

```

```

thl$MEREID<-paste(thl$IDENT,thl$MER2E,sep="-")
thl$PEREID<-paste(thl$IDENT,thl$PER2E,sep="-")

# identification du lieu de naissance du pÃ¨re et de la mÃ¨re
# de maniÃ¨re entiÃ¨rement interne Ã la base THL
m<-match(thl$MEREID,thl$IDENTIFIANT)
thl$MERE_LN <- thl$LNAIS[m]
m<-match(thl$PEREID,thl$IDENTIFIANT)
thl$PERE_LN <- thl$LNAIS[m]

# maintenant indiquer dans THL le lieu de naissance
# des parents de EGO (quand ils manquent dans le THL)

m <- match(thl$IDENTIFIANT,indiv$IDENTIFIANT)
thl$MERE_LN <- ifelse(is.na(thl$MERE_LN),indiv$Lnaism[m],thl$MERE_LN)
thl$PERE_LN <- ifelse(is.na(thl$PERE_LN),indiv$Lnaisp[m],thl$PERE_LN)

# maintenant indiquer dans THL le lieu de naissance
# des parents de son conjoint
# indiv_final$Cjte : NOI du conjoint de l'enquÃªtÃ© (quand le conjoint est dans le THL)
# on recode le lieu de naissance des parents du conjoint de EGO
indiv_final$Lieu naiscamr2 <- ifelse(indiv_final$Lieu naiscamr%in%c(NA,""),NA,ifelse(indiv_final$Lieu naiscamr%in%c(0,1),1,2))
indiv_final$Lieu naiscapr2 <- ifelse(indiv_final$Lieu naiscapr%in%c(NA,""),NA,ifelse(indiv_final$Lieu naiscapr%in%c(0,1),1,2))
indiv_final$IDENTIFIANT_CONJOINT <- paste(indiv_final$IDENT,indiv_final$Cjte,sep="-")
m <- match(thl$IDENTIFIANT,indiv_final$IDENTIFIANT_CONJOINT)
thl$MERE_LN <- ifelse(is.na(thl$MERE_LN),indiv_final$Lieu naiscamr2[m],thl$MERE_LN)
thl$PERE_LN <- ifelse(is.na(thl$PERE_LN),indiv_final$Lieu naiscapr2[m],thl$PERE_LN)

# Lieunaiscamr et Lieunaiscapr ne sont pas renseignÃ©s quand
# le conjoint est immigrÃ© (car nÃ© Ã©tranger Ã l'Ã©tranger)
# donc on va coder que si le conjoint est immigrÃ©
# ses deux parents sont nÃ©s Ã l'Ã©tranger

indiv_final$INDICLIEUNAISPARENTCONJOINT <- ifelse(indiv_final$Group1ca==1,2,1)
m <- match(thl$IDENTIFIANT,indiv_final$IDENTIFIANT_CONJOINT)
thl$MERE_LN <- ifelse(is.na(thl$MERE_LN),indiv_final$INDICLIEUNAISPARENTCONJOINT[m],thl$MERE_LN)
thl$PERE_LN <- ifelse(is.na(thl$PERE_LN),indiv_final$INDICLIEUNAISPARENTCONJOINT[m],thl$PERE_LN)

#####
# Maintenant les enfants de EGO
#
# on veut connaitre le lieu de naissance du conjoint de EGO, 2e parent
#
# avec comme problÃ¨me le fait que certains 2e parent ne sont plus dans le logement
#
# i.e. thl$LIENEGO==02
# quand l'un des parents n'habite pas le logement.
# i.e. thl$MERE2E ou thl$PERE2E == 0

# opÃ©ration Ã faire :
# comparer la date de naissance des enfants
# et les unions des parents

# exemple : head(subset(thl,thl$LIENEGO=="02"&is.na(thl$MERE_LN)))
# "Lydia" (IDENT:152) vit avec son pÃ¨re et pas avec sa mere (dÃ©cÃ©dÃ©e ?)
# pÃ¨re de Lydia : Omar, nÃ© en 1965
# mais problÃ¨me : Omar n'a qu'une seule union dÃ©clarÃ©e
#indiv[152,c("Nbunions")]
#indiv[152,c("C_Relac")]
#indiv[152,c("C_Finap")]

# autre exemple "kelly" IDENT:287
#indiv[287,1:50]
#indiv[287,c("Nbunions")]
#indiv[287,c("C_Relac")]
#indiv[287,c("C_Finap")]
#Mais lÃ le pÃ¨re dÃ©clare deux unions
#subset(thl,thl$IDENT=="287")
# et la premiÃ¨re union date de 1997, un an avant naissance de leur premier enfant
#indiv[287,c("C_Relap")]

# si le parent qui est dans le thl

```



```

# ne déclare qu'une seule union
# alors donner à l'autre parent
# les caractéristiques liées à cette union
# ou les caractéristiques du parents s'il y a eu ZERO unions

#
thl$EGO_ID <- paste(thl$IDENT,thl$EGO,sep="-")

# on recode le lieu de naissance du *conjoint* de EGO (pas des parents du conjoint, ce qu'on a déjà fait)
indiv_final$Lieunaiscar2 <- ifelse(indiv_final$Lieunaiscar%in%c(NA,""),NA,ifelse(indiv_final$Lieunaiscar%in%c(0,1),1,2))

# ci pour indiquer dans le THL le lieu de naissance du conjoint indiqué dans indiv_final
m <- match(thl$EGO_ID,indiv_final$IDENTIFIANT)

# on indique le nombre d'unions de EGO, pour savoir qui est l'autre parent
thl$EGO_NBUNION <- indiv_final$Nbunions[m]

# si EGO (le parent) n'y a eu qu'une seule union
thl$MERE_LN <- ifelse( is.na( thl$MERE_LN ) & thl$EGO_NBUNION == 1 & thl$LIENEGO == "02",
  indiv_final$Lieunaiscar2[m], thl$MERE_LN)

thl$PERE_LN <- ifelse( is.na( thl$PERE_LN ) & thl$EGO_NBUNION == 1 & thl$LIENEGO == "02",
  indiv_final$Lieunaiscar2[m], thl$PERE_LN)

# si EGO, le parent, a eu zéro union mais qu'il/elle a quand même des enfants
# alors on indique lieu naissance du parent car le lieu de naissance du partenaire n'est pas indiqué

thl$MERE_LN <- ifelse( is.na( thl$MERE_LN ) & thl$EGO_NBUNION == 0 & thl$LIENEGO == "02",
  indiv_final$Lnaise[m], thl$MERE_LN)

thl$PERE_LN <- ifelse( is.na( thl$PERE_LN ) & thl$EGO_NBUNION == 0 & thl$LIENEGO == "02",
  indiv_final$Lnaise[m], thl$PERE_LN)

# Si il y a eu plus de 2 unions, le processus est plus complexe :
# si 2 unions ou plus, il faut regarder la date de naissance des enfants
# et comparer avec les dates de début et fin des unions

# head(subset(thl,is.na(thl$PERE_LN)&thl$LIENEGO=="02"))

thl$EGO_CRELAP<-indiv_final$C_Relap[m] # date début première relation de EGO
thl$EGO_CFINAP<-indiv_final$C_Finap[m] # date fin première relation conjoint
thl$EGO_CRELAC<-indiv_final$C_Relac[m] # date début conjoint actuel ego
thl$EGO_CRELAA<-indiv_final$C_Relaa[m] # date début relation actuelle personnes pas en couple

thl$AUTPAR_PC<-as.numeric(thl$ANAIS<=thl$EGO_CFINAP)
# 1 si enfant né avant fin de la première union
thl$AUTPAR_CA<-as.numeric(thl$ANAIS>=thl$EGO_CRELAC)
# 1 si enfant est né après début relation avec conjoint actuel
thl$AUTPAR_AA<-as.numeric(thl$ANAIS>=thl$EGO_CRELAA)
# 1 si enfant né après début relation amoureuse actuelle

# si l'enfant est né avant la fin de la première union
# alors on lui donne comme lieu de naissance de l'autre parent le lieu de naissance du conjoint 1ère union

indiv_final$Lieunaispcr2 <- ifelse( indiv_final$Lieunaispcr %in% c(0,1), 1, 2)

thl$MERE_LN <- ifelse( is.na( thl$MERE_LN ) & thl$LIENEGO == "02" & thl$AUTPAR_PC == 1,
  indiv_final$Lieunaispcr2[m], thl$MERE_LN)

thl$PERE_LN <- ifelse( is.na( thl$PERE_LN ) & thl$LIENEGO == "02" & thl$AUTPAR_PC == 1,
  indiv_final$Lieunaispcr2[m], thl$PERE_LN)

# si l'enfant est né après début de relation avec conjoint actuel ou relation amoureuse actuelle
# alors on lui donne comme lieu de naissance de l'autre parent le lieu de naissance du conjoint / relation actuelle

thl$MERE_LN <- ifelse( is.na( thl$MERE_LN ) & thl$LIENEGO == "02" & (thl$AUTPAR_CA == 1 | thl$AUTPAR_AA ==1),
  indiv_final$Lieunaispcr2[m], thl$MERE_LN)

thl$PERE_LN <- ifelse( is.na( thl$PERE_LN ) & thl$LIENEGO == "02" & (thl$AUTPAR_CA == 1 | thl$AUTPAR_AA ==1),
  indiv_final$Lieunaispcr2[m], thl$PERE_LN)

# après tout cela il reste environ 660 enfants du THL sans précision sur lieu de naissance de l'autre parent

```

```

# on code le lieu de naissance de l'autre parent Ã partir d'indices
# le lieu de naissance de EGO
# la nationalitÃ de l'enfant
# le lieu de naissance de l'enfant
# cela se trouve dans la base "enfants"

# variable enfants$LNP : indicateur du lieu de naissance des parents
# (si enfant n'Ã pas Ãtranger Ã l'Ãtranger alors parents probablement Ãtrangers)
enfants$LNP <- ifelse(enfants$E_LN == 1 & enfants$E_NAT == 1, 1, 2)
# attention, ce n'est pas une variable fine
# elle a pour but de renseigner sur les enfants du THL pour lesquels on n'a pas d'information

# clÃ d'appariement entre bases enfants et base THL

m <- match(thl$IDENTIFIANT, enfants$IDENTIFIANT)

#thl$ENFANT_LNP <- enfants$LNP[m]

thl$MERE_LN <- ifelse( is.na( thl$MERE_LN ) & thl$LIENEGO == "02" ,
  enfants$LNP[m], thl$MERE_LN)

thl$PERE_LN <- ifelse( is.na( thl$PERE_LN ) & thl$LIENEGO == "02" ,
  enfants$LNP[m], thl$PERE_LN)

# arrivÃ Ã cette Ãtape, il n'y a plus d'enfants du THL pour lesquels retrouver le lieu de naissance des parents

# maintenant, il faut extraire des informations sur les enfants
# qui ne sont plus dans le mÃnage

enfants_autres<-subset(enfants,enfants$E_NOI_THL=="")
# il y en a 7924

# il faut procÃder aux memes comparaisons.

# indiquer le sexe de EGO le parent interrogÃ
# aller chercher lieu de naissance de EGO parent interrogÃ
# comparer les dates de naissances et dates des relations de EGO

m <- match(enfants_autres$IDENT, indiv_final$IDENT)

enfants_autres$EGO_SEXE <- indiv_final$Sexee[m]
enfants_autres$PERE_LN <- ifelse(enfants_autres$EGO_SEXE==1, indiv_final$Lnaise[m], NA)
enfants_autres$MERE_LN <- ifelse(enfants_autres$EGO_SEXE==2, indiv_final$Lnaise[m], NA)

# Nombre d'unions
enfants_autres$EGO_NBUNION <- indiv_final$Nbunions[m]

# si parent a eu zÃro union (on indique lieu naissance du parent)
# car le conjoint n'est pas renseignÃ
enfants_autres$MERE_LN <- ifelse( is.na( enfants_autres$MERE_LN ) & enfants_autres$EGO_NBUNION == 0 ,
  indiv_final$Lnaise[m], enfants_autres$MERE_LN)
enfants_autres$PERE_LN <- ifelse( is.na( enfants_autres$PERE_LN ) & enfants_autres$EGO_NBUNION == 0 ,
  indiv_final$Lnaise[m], enfants_autres$PERE_LN)

# si parent a eu une union
# alors on a le lieu de naissance du conjoint
enfants_autres$MERE_LN <- ifelse( is.na( enfants_autres$MERE_LN ) & enfants_autres$EGO_NBUNION == 1 ,
  indiv_final$Lieuunaiscar2[m], enfants_autres$MERE_LN)
enfants_autres$PERE_LN <- ifelse( is.na( enfants_autres$PERE_LN ) & enfants_autres$EGO_NBUNION == 1 ,
  indiv_final$Lieuunaiscar2[m], enfants_autres$PERE_LN)

# si le parent a eu plusieurs unions

enfants_autres$EGO_CRELAP<-indiv_final$C_Relap[m] # date dÃbut premiÃre relation
enfants_autres$EGO_CFINAP<-indiv_final$C_Finap[m] # date fin premiere relation conjoint
enfants_autres$EGO_CRELAC<-indiv_final$C_Relac[m] # date debut conjoint actuel ego
enfants_autres$EGO_CRELAA<-indiv_final$C_Relaa[m] # date dÃbut relation actuelle personnes pas en couple

enfants_autres$AUTPAR_PC<-as.numeric(enfants_autres$E_ANAIS<=enfants_autres$EGO_CFINAP)
# 1 si enfant n'Ã pas avant fin de la premiÃre union
enfants_autres$AUTPAR_CA<-as.numeric(enfants_autres$E_ANAIS>=enfants_autres$EGO_CRELAC)
# 1 si enfant est nÃ aprÃs dÃbut relation avec conjoint actuel

```

```

enfants_autres$AUTPAR_AA<-as.numeric(enfants_autres$E_ANAIS>=enfants_autres$EGO_CRELAA)
# 1 si enfant n'aurait pas de relation amoureuse actuelle

enfants_autres$MERE_LN <- ifelse( is.na( enfants_autres$MERE_LN ) & enfants_autres$AUTPAR_PC == 1,
  indiv_final$Lieunaispcr2[m], enfants_autres$MERE_LN)
enfants_autres$PERE_LN <- ifelse( is.na( enfants_autres$PERE_LN ) & enfants_autres$AUTPAR_PC == 1,
  indiv_final$Lieunaispcr2[m], enfants_autres$PERE_LN)

# si l'enfant est au début de relation avec conjoint actuel ou relation amoureuse actuelle
# alors on lui donne comme lieu de naissance de l'autre parent le lieu de naissance du conjoint / relation actuelle

enfants_autres$MERE_LN <- ifelse( is.na( enfants_autres$MERE_LN ) & (enfants_autres$AUTPAR_CA == 1 |
enfants_autres$AUTPAR_AA ==1),
  indiv_final$Lieunaispcr2[m], enfants_autres$MERE_LN)
enfants_autres$PERE_LN <- ifelse( is.na( enfants_autres$PERE_LN ) & (enfants_autres$AUTPAR_CA == 1 |
enfants_autres$AUTPAR_AA ==1),
  indiv_final$Lieunaispcr2[m], enfants_autres$PERE_LN)

# il reste environ 600 enfants pour lesquels il faut imputer un lieu de naissance des parents

# on considère qu'un enfant n'appartient à l'un ou l'autre d'un parent n'appartient à l'un ou l'autre

enfants_autres$LNP <- ifelse(enfants_autres$E_LN == 1 & enfants_autres$E_NAT == 1, 1, 2)

enfants_autres$MERE_LN <- ifelse(is.na(enfants_autres$MERE_LN), enfants_autres$LNP, enfants_autres$MERE_LN)
enfants_autres$PERE_LN <- ifelse(is.na(enfants_autres$PERE_LN), enfants_autres$LNP, enfants_autres$PERE_LN)

# maintenant on reporte sur la table "enfants" les informations du THL et de enfants_autres

# d'abord les enfants qui résident avec leurs parents dans le logement
m <- match(enfants$IDENTIFIANT, thl$IDENTIFIANT)
enfants$MERE_LN <- thl$MERE_LN[m]
enfants$PERE_LN <- thl$PERE_LN[m]

# ensuite les enfants non cohabitants
# qui ne disposent pas de NOI, il faut donc une autre clé
enfants_autres$PRENOM_ID <- paste(enfants_autres$Prenom1, enfants_autres$IDENTIFIANT)
enfants$PRENOM_ID <- paste(enfants$Prenom1, enfants$IDENTIFIANT)
m <- match(enfants$PRENOM_ID, enfants_autres$PRENOM_ID)
enfants$MERE_LN <- ifelse(is.na(enfants$MERE_LN), enfants_autres$MERE_LN[m], enfants$MERE_LN)
enfants$PERE_LN <- ifelse(is.na(enfants$PERE_LN), enfants_autres$PERE_LN[m], enfants$PERE_LN)

# on recalcule enfants$LNP
enfants$LNP <- ifelse(enfants$MERE_LN==1 & enfants$PERE_LN==1, 1, 2)

#####
#
# INDICE
#
#
#####

# on va combiner les informations du THL et celles de enfants_autres
# pour constituer le FNI

# dans le THL, extraire EGO, enfants, conjoints seulement
# et ne garder que prénom, lieu de naissance des parents, année de naissance, et pondération
# dans enfants_autres, extraire prénom, lieu de naissance des parents et année de naissance, et pondération

#####
#
#
# FNI non pondéré
#
#
#####

# à retravailler en ajoutant les individus du THL pour lesquels on a LIENEGO==10 (frères et sœurs)

tmp <- subset(thl, thl$LIENEGO %in% c("00", "01", "02"))[, c("Prenom1",
  "IDENTIFIANT",
  "ANAI",
  "ANARRIV",
  "LANAI",

```

```

        "MERE_LN",
        "PERE_LN",
        "Poids",
        "LIENEGO")]}

tmp2<-
subset(enfants_autres)[,c("Prenom1", "IDENTIFIANT", "E_ANAIS", "E_ARRIV", "E_LN", "MERE_LN", "PERE_LN", "Poids")]
tmp2$LIENEGO<-"99" # pour indiquer "enfant non r  sidant"

colnames(tmp2)<-c("Prenom1",
  "IDENTIFIANT",
  "ANAIS",
  "ANARRIV",
  "LNAIS",
  "MERE_LN",
  "PERE_LN",
  "Poids",
  "LIENEGO")

fni_source<-rbind(tmp,tmp2)
# on enleve des non r  ponses sur le lieu de naissance des parents
#   sa posait probl  me, car une partie des individus n'avaient plus de FNI
# fni_source<-subset(fni_source, (!fni_source$PERE_LN%in%c("",NA) & !fni_source$MERE_LN %in% c("",NA)))

fni_source$Poids<-as.numeric(as.character(fni_source$Poids))

# indicateur synth  tique du lieu de naissance des parents
fni_source$PAR_LN<-ifelse(fni_source$PERE_LN==1 & fni_source$MERE_LN==1, 0 , 1)
fni_source$NB<-1

#####
# probl  me des pr  noms uniques.
# Probl  me des pr  noms uniques
# ces pr  noms sont rapproch  s des pr  noms qui leur ressemblent
# et qui apparaissent au moins deux fois
#####

prenoms_frequents <- as.data.frame(table(fni_source$Prenom1))

# rep  rage des pr  noms voisins, pour les pr  noms qui sont tr  s peu donn  s
#
#####

# Probl  me des pr  noms uniques
# ces pr  noms sont rapproch  s des pr  noms qui leur ressemblent
# et qui apparaissent au moins deux fois

##### Probl  me d'encodage pour certains pr  noms
##### a nettoyer d  s le d  part
##### par exemple MICKA\xc3L
#####

prenoms_frequents <- as.data.frame(table(fni_source$Prenom1))

# packages n  cessaires
#install.packages("stringdist")
library(stringdist) # distance entre les pr  noms
#install.packages("FastKNN")
library(FastKNN) # plus proches voisins
library(igraph)

liste <- prenoms_frequents$Var1
dm <- stringdistmatrix(liste,liste,weight=c(.5,1,1,.5),method="osa")
k=1
nn = matrix(0,length(liste),k) # n x k
for (i in 1:length(liste))
  nn[i,] = k.nearest.neighbors(i, dm, k = k)

rownames(nn) <- liste
nn <- as.data.frame(nn)
nn$voisin1 <- liste[nn[,1]]

m <- match(rownames(nn),prenoms_frequents$Var1)
nn$N <- prenoms_frequents$Freq[m]

```

```

m <- match(nn$voisin1,prenoms_frequents$Var1)
nn$N1 <- prenoms_frequents$Freq[m]

test <- nn[,c("voisin1","N","N1")]
test$PR <- as.character(rownames(test))
colnames(test) <- c("Voisin","NB_PR","NB_Voisin","PR")
test$Voisin <- as.character(test$Voisin)
g <- subset(test,test$NB_PR==1&test$NB_Voisin==1)

g <- graph.data.frame(g[,c("PR","Voisin")])
#plot(g)
V(g)$groupe <- clusters(g)$membership
V(g)$groupe_taille <- clusters(g)$size[clusters(g)$membership]
df <- as.data.frame(cbind(V(g)$name,V(g)$groupe,V(g)$groupe_taille))
colnames(df) <- c("Prenom","Groupe","Taille")
df$Prenom<-as.character(df$Prenom)
#df
a_garder <- data.frame(Indice=1:max(as.numeric(as.character(df$Groupe))),Remplacement=NA)
for (i in (1:max(as.numeric(as.character(df$Groupe)))) {
  a_garder[i,] <- c(i,as.character(sample(subset(df,df$Groupe==i)$Prenom,1)))
}
m <- match(df$Groupe,a_garder$Indice)
df$Remplacement <- a_garder$Remplacement[m]
m <- match(test$PR,df$Prenom)
test$Remplacement <- df$Remplacement[m]
test$NB_Remplacement <- df$Taille[m]

test$Remplacement <- ifelse(is.na(test$Remplacement)&test$NB_Voisin>1&test$NB_PR==1,test$Voisin,
  ifelse(is.na(test$Remplacement)&test$NB_PR>1,test$PR,test$Remplacement))

head(test,n=150)
remplacement<-test

save(remplacement,file="remplacement-prenoms-voisins.Rdata")

# fichier "remplacement" issu d'un travail de classification
#####
load("remplacement-prenoms-voisins.Rdata") #objet remplacement

m <- match(fni_source$Prenom1,remplacement$PR)
fni_source$Prenom2 <- remplacement$Remplacement[m]

prenoms_frequents2 <- as.data.frame(table(fni_source$Prenom2))

# calcul du FNI "basique", en gardant les prÃ©noms uniques
#
# on calcule le FNI avec les prÃ©noms de dÃ©part
# (y compris les prÃ©noms qu'on ne trouve qu'une seule fois)
#####
fni <- dcast(fni_source, Prenom1~PAR_LN, value.var="NB", sum)
colnames(fni)<-c("Prenom1","NONMIG","MIG")
fni$FNI<-(fni$NONMIG/sum(fni$NONMIG))/(fni$NONMIG/sum(fni$NONMIG)+fni$MIG/sum(fni$MIG))
fni$N <- fni$MIG+fni$NONMIG

# on associe chaque prÃ©nom Ã son FNI
m<-match(fni_source$Prenom1,fni$Prenom1)
fni_source$FNI <- fni$FNI[m]
fni_source$QUINQ <- 5*floor(fni_source$ANAI/5) # quinquennats

#####
#
#
# il faut prendre en compte le caractÃ©re historique des prÃ©noms
# et donc calculer le FNI par dÃ©cennie ou par groupe de 20 ans
# 2010-1990, 1990-1970, 1970-1940
# car un prÃ©nom "franÃ§ais" en 1940 peut etre devenu un prÃ©nom "de descendant de migrant" en 2000
# ce qui contribue Ã rapprocher les indices de 1940-1950 alors qu'il n'y a pas simultanÃ©itÃ© des pratiques

```

```

#
# #####
#options(scipen=99999) # pour Ã©viter scientific notation

# fni_source$PERIODE<-cut(fni_source$ANAIIS,breaks=c(1900,1970,1990,2010),dig.lab=10)

#fni_source<-subset(fni_source, !is.na(fni_source$ANAIIS))
fni_source$FNI_MOBILE<-NA # la variable FNI_MOBILE est calculÃ©e sur une pÃ©riode mobile de 20 ans
fni_source$PRENOM_ANNEE <- paste(fni_source$ANAIIS,fni_source$Prenom1)

for (i in 1920:2010){
  tmp<-subset(fni_source,fni_source$ANAIIS>(i-10)&fni_source$ANAIIS<=(i+10))
  fni <- dcast(tmp, Prenom1~PAR_LN, value.var="NB", sum)
  colnames(fni)<-c("Prenom1","NONMIG","MIG")
  fni$FNI<-(fni$NONMIG/sum(fni$NONMIG))/(fni$NONMIG/sum(fni$NONMIG)+fni$MIG/sum(fni$MIG))
  fni$PRENOM_ANNEE <- paste(i,fni$Prenom1)
  m<-match(fni_source$PRENOM_ANNEE,fni$PRENOM_ANNEE)
  fni_source$FNI_MOBILE <- ifelse(is.na(fni_source$FNI_MOBILE),fni$FNI[m],fni_source$FNI_MOBILE)
  print(i)
}

#####
#
# TroisiÃ¨me FNI :
# Calcul du FNI aprÃ¨s recodage des prÃªnoms (prÃªnoms uniques associÃ©s Ã leur plus proche voisin)
# la variable crÃ©Ã©e est FNI2
#
#####

fni <- dcast(fni_source, Prenom2~PAR_LN, value.var="NB", sum)
colnames(fni)<-c("Prenom2","NONMIG","MIG")
fni$FNI<-(fni$NONMIG/sum(fni$NONMIG))/(fni$NONMIG/sum(fni$NONMIG)+fni$MIG/sum(fni$MIG))
fni$N <- fni$MIG+fni$NONMIG
# on associe chaque prÃªnom Ã son FNI
m<-match(fni_source$Prenom2,fni$Prenom2)
fni_source$FNI2 <- fni$FNI[m] # FNI2 : FNI calculÃ© Ã partir des prÃªnoms recodÃ©s

#####
#
#
# Associer les FNI aux bases "enfants" et "indiv_final"
#
#
#####

# associer un FNI Ã chaque individu de la base indiv_final
# et associer un FNI Ã chaque individu de la base "enfants"

prenoms_frequents <- as.data.frame(table(fni_source$Prenom1))
m <- match(fni_source$Prenom1,prenoms_frequents$Var1)
fni_source$NB_Prenom1 <- prenoms_frequents$Freq[m] # indique le nombre total de prÃªnom dans la base, pour chaque prÃªnom

m <- match(indiv_final$IDENTIFIANT,fni_source$IDENTIFIANT)
indiv_final$FNI <- fni_source$FNI[m]
indiv_final$FNI2 <- fni_source$FNI2[m]
indiv_final$FNI_MOBILE <- fni_source$FNI_MOBILE[m]
indiv_final$NB_Prenom1 <- fni_source$NB_Prenom1[m]
indiv_final$PAR_LN <- ifelse(indiv_final$lnaispts_b==1,1,2)
indiv_final$Prenom1 <- fni_source$Prenom1[m]

# les enfants
#####
# l'appariement est un peu plus complexe, car il y a des enfants non rÃ©sidant
# qui n'ont pas de NOI, et donc la variable "IDENTIFIANT" n'est pas un identifiant unique

enfants$IDENTIFIANT2 <- paste(enfants$IDENTIFIANT,enfants$Prenom1)
fni_source$IDENTIFIANT2 <- paste(fni_source$IDENTIFIANT,fni_source$Prenom1)

```



```

m <- match(enfants$IDENTIFIANT2,fni_source$IDENTIFIANT2)
enfants$FNI <- fni_source$FNI[m]
enfants$FNI2 <- fni_source$FNI2[m]
enfants$FNI_MOBILE <- fni_source$FNI_MOBILE[m]
enfants$NB_Prenom1 <- fni_source$NB_Prenom1[m]

#####
#
#
# Enrichissement de la table "indiv_final"
# avec le FNI moyen des enfants de EGO
#
#
#####
# c'est la proc dure d'appariement la plus simple
# mais alors qu'il y a plus de 31 000 enfants
# seuls 13 000 EGO ont des enfants

indiv_final$NB <- 1
fni_moyen_enfants_ego <- aggregate(FNI2~IDENT,data=enfants,mean)
m <- match(indiv_final$IDENT,fni_moyen_enfants_ego$IDENT)
indiv_final$FNI_moyen_enfants <- fni_moyen_enfants_ego$FNI2[m]
# si EGO a un enfant au moins, on cr e une variable FNI moyen de ses enfants
# attention, l'enfant peut  tre n  en France ou non
# on cr e donc une autre variable, pour les enfants n s en France

fni_moyen_enfants_ego_fr <- aggregate(FNI2~IDENT,data=subset(enfants,enfants$E_LN==1),mean)
m <- match(indiv_final$IDENT,fni_moyen_enfants_ego_fr$IDENT)
indiv_final$FNI_moyen_enfants_fr <- fni_moyen_enfants_ego_fr$FNI2[m]

# nettoyage
rm(tmp,tmp2,remplacement,prenoms_frequents,prenoms_frequents2,fni_moyen_enfants_ego,fni_moyen_enfants_ego_fr,enfants_au
tres,fni_source,fni,i,m)

#####
#
#
# Association avec les donn es pr noms classifi es par groupe national et linguistique
#
#
#
# note : les pr noms complets ont  t  classifi s (pr noms compos s)
# d'o  n cessit  de proc der   l'appariement ici
#
#
# on ajoute variables CAT0, CAT1...

# l' tape suivante est facultative
# il s'agit de limiter le nombre de pr noms "inclassables"
# en prenant pour r f rence le premier pr nom

# on g n re CAT0_corr et CAT1_corr
# dans les faits  a change peu de choses

library(tidyverse)

# classification c'est, dans THL: PRENOM CAT1 et CAT0

```

```
# la classification a été faite sur l'ensemble des prénoms "Maria Augusta" et pas seulement "Maria"
# cela pose quelques problèmes car nous travaillons ensuite sur les premiers prénoms uniquement
```

```
classification <- classification %>%
  mutate(Prenom1 = gsub('-', '', Prenom1), # on enleve les tirets et les _
         Prenom1 = gsub('[(.*)^ ]+[A-Z]+', '', Prenom1), # on ne garde que la première partie du prénom
         Prenom1 = gsub(' ', '', Prenom1), # on enleve certains trailing spaces qui restent
         Prenom1 = stri_trans_general(Prenom1, "Latin-ASCII") )
```

```
# associer la catégorisation "majoritaire"
# car Marie Akoye a été classée "inclassable" en raison du 2e prénom
# or on ne travaille que sur le premier prénom
# donc il faut que Marie Akoye soit classée comme Marie : 129
```

```
# on identifie les prénoms qui se retrouvent associés à des CAT0 différents
prenoms_problemes <- classification %>%
  group_by(Prenom1, CAT0) %>%
  summarize(N=n()) %>%
  group_by(Prenom1) %>%
  summarize(total = n()) %>% arrange(-total) %>%
  filter(total>1)
```

```
correction_CAT0 <- classification %>% filter(Prenom1 %in% prenoms_problemes$Prenom1) %>%
  group_by(Prenom1, CAT0) %>% summarize(N=n()) %>%
  group_by(Prenom1) %>% mutate(keep = (rank(-N, ties.method="random") == 1) ) %>%
  filter(keep) %>%
  select(Prenom1, CAT0_corr = CAT0)
```

```
classification <- classification %>%
  left_join(correction_CAT0, by="Prenom1") %>%
  mutate(CAT0_corr = ifelse(is.na(CAT0_corr), CAT0, CAT0_corr) )
```

```
# on identifie les prénoms qui se retrouvent associés à des CAT1 différents
prenoms_problemes <- classification %>%
  group_by(Prenom1, CAT1) %>%
  summarize(N=n()) %>%
  group_by(Prenom1) %>%
  summarize(total = n()) %>% arrange(-total) %>%
  filter(total>1)
```

```
correction_CAT1 <- classification %>% filter(Prenom1 %in% prenoms_problemes$Prenom1) %>%
  group_by(Prenom1, CAT1) %>% summarize(N=n()) %>%
  group_by(Prenom1) %>% mutate(keep = (rank(-N, ties.method="random") == 1) ) %>%
  filter(keep) %>%
  select(Prenom1, CAT1_corr = CAT1)
```

```
classification <- classification %>%
  left_join(correction_CAT1, by="Prenom1") %>%
  mutate(CAT1_corr = ifelse(is.na(CAT1_corr), CAT1, CAT1_corr) )
```

```
classification <- classification %>%
  group_by(Prenom1, CAT0_corr, CAT1_corr) %>%
  summarize(N=n()) %>% select(-N)
```

```
# nouveau code : on associe juste les prénoms
```

```
indiv_final <- indiv_final %>%
  left_join(classification, by = "Prenom1")
```

```

thl <- thl %>%
  left_join(classification, by = "Prenom1")

enfants <- enfants %>%
  left_join(classification, by = "Prenom1")

# sauvegarde
save(indiv_final,enfants,thl,file="TeO.Rdata")

rm(list=ls())

```

```
#####
```

```
# ouf ! on est arriv   au bout de la premi  re   tape
```

## ❖ 2   programme informatique communiqu   par les auteurs (02-recodage)

```
#####
```

```
#####
#####
#####
#####
##### P R E P A R A T I O N #####
#####
#####
#####
```

```
#
# Code qui pr  pare les bases pour les traitements suivants
#
```

```
#####
#####
#####
#####
# Recodage et cr  ation d'indicateurs
# Appariement Enfants / Parents
```

```
# D  terminants du choix des pr  noms
# Cr  ation d'un fichier avec toutes les variables utilis  es ensuite
```

```
# encodage de ce fichier : UTF-8
```

```
# Packages
# install.packages(tidyverse)
library(tidyverse)
library(stringi)
# espace de travail
# setwd("REPertoire")
```

```
#####
# donn  es
#####
# ce sont les donn  es enfants, THL et indiv
# avec association du FNI
# issus des traitements initiaux
load(file="TeO.Rdata")
```

```
#####
#
#
# Cr  ation d'indicateurs et de variables sur indiv_final
# pour ensuite   tudier pr  nom de ego et pr  nom des enfants de ego
#
#####
```

```
#####
# Lieu de naissance : en clair
#####
```

```

# dans indiv_final, lieu de naissance de Ego et de ses parents

# Recoder les régions en clair

# Attention, ici la "France", c'est aussi la Belgique et la Suisse...

indiv_final$REGION_EGO <- ifelse(indiv_final$Regionnaise2 %in% c(1101,1201,1202,1203,1204,1205,4410,4404),"France",
  ifelse(indiv_final$Regionnaise2 %in% c(2101,2201,2301),"ALG-MAR-TUN",
    ifelse(indiv_final$Regionnaise2 %in% c(2401,2402,2403,2404,2405,2406,2407,2408,2409), "SAH",
      ifelse(indiv_final$Regionnaise2 %in% c(4101,4201,4301), "POR-ESP-ITA",
        ifelse(indiv_final$Regionnaise2 %in%
c(2501,2502,2503,2504,2505,2506,2507,2508,2509,2510,2511,2512), "Afrique centrale",
          ifelse(indiv_final$Regionnaise2 %in% c(2601:2901), "Autre Afrique",
            ifelse(indiv_final$Regionnaise2 %in% c(3101:3600), "Asie (avec Turquie)",
              ifelse(is.na(indiv_final$Regionnaise2),"NR","AUTRE"))))))))

indiv_final$REGION_PERE2 <- ifelse(indiv_final$Regionnaisp2 %in% c(1101,1201,1202,1203,1204,1205,4410,4404),"France",
  ifelse(indiv_final$Regionnaisp2 %in% c(2101,2201,2301),"ALG-MAR-TUN",
    ifelse(indiv_final$Regionnaisp2 %in% c(2401,2402,2403,2404,2405,2406,2407,2408,2409), "SAH",
      ifelse(indiv_final$Regionnaisp2 %in% c(4101,4201,4301), "POR-ESP-ITA",
        ifelse(indiv_final$Regionnaisp2 %in%
c(2501,2502,2503,2504,2505,2506,2507,2508,2509,2510,2511,2512), "Afrique centrale",
          ifelse(indiv_final$Regionnaisp2 %in% c(2601:2901), "Autre Afrique",
            ifelse(indiv_final$Regionnaisp2 %in% c(3101:3600),"Asie (avec Turquie)",
              ifelse(is.na(indiv_final$Regionnaisp2),"NR","AUTRE"))))))))

indiv_final$REGION_MERE2 <- ifelse(indiv_final$Regionnaism2 %in% c(1101,1201,1202,1203,1204,1205,4410,4404),"France",
  ifelse(indiv_final$Regionnaism2 %in% c(2101,2201,2301),"ALG-MAR-TUN",
    ifelse(indiv_final$Regionnaism2 %in% c(2401,2402,2403,2404,2405,2406,2407,2408,2409), "SAH",
      ifelse(indiv_final$Regionnaism2 %in% c(4101,4201,4301), "POR-ESP-ITA",
        ifelse(indiv_final$Regionnaism2 %in%
c(2501,2502,2503,2504,2505,2506,2507,2508,2509,2510,2511,2512), "Afrique centrale",
          ifelse(indiv_final$Regionnaism2 %in% c(2601:2901), "Autre Afrique",
            ifelse(indiv_final$Regionnaism2 %in% c(3101:3600), "Asie (avec Turquie)",
              ifelse(is.na(indiv_final$Regionnaism2),"NR","AUTRE"))))))))

indiv_final$REGION_CONJOINT2 <- ifelse(indiv_final$Regionnaisca2 %in%
c(1101,1201,1202,1203,1204,1205,4410,4404),"France",
  ifelse(indiv_final$Regionnaisca2 %in% c(2101,2201,2301),"ALG-MAR-TUN",
    ifelse(indiv_final$Regionnaisca2 %in% c(2401,2402,2403,2404,2405,2406,2407,2408,2409), "SAH",
      ifelse(indiv_final$Regionnaisca2 %in% c(4101,4201,4301), "POR-ESP-ITA",
        ifelse(indiv_final$Regionnaisca2 %in%
c(2501,2502,2503,2504,2505,2506,2507,2508,2509,2510,2511,2512), "Afrique centrale",
          ifelse(indiv_final$Regionnaisca2 %in% c(2601:2901), "Autre Afrique",
            ifelse(indiv_final$Regionnaisca2 %in% c(3101:3600), "Asie (avec Turquie)",
              ifelse(is.na(indiv_final$Regionnaisca2),"NR","AUTRE"))))))))

#####
#
# Mixité du couple
#
#####

# variable indiv_final$MIXlimmiPC (pour les EGO immigrés)

indiv_final$COUPLE_MIXTE_IMMI <- factor(indiv_final$MIXlimmiPC, exclude=NULL) # on garde les NA
levels(indiv_final$COUPLE_MIXTE_IMMI)[is.na(levels(indiv_final$COUPLE_MIXTE_IMMI))]="Sans info" # nom vide pour
quelques individus
indiv_final$COUPLE_MIXTE_IMMI <- plyr::revalue(indiv_final$COUPLE_MIXTE_IMMI, c("1"="immigré ou descendant,
mère pays",
  "2"="immigré ou descendant, autre pays",
  "3"="immigré ou descendant, mère pays",
  "4"="immigré ou descendant, autre pays",
  "5"="groupe majoritaire, Français, ou DOM",
  "6"="groupe majoritaire, Français, ou DOM",
  "7"="groupe majoritaire, Français, ou DOM",
  "8"="immigré ou descendant, mère pays",
  "9"="groupe majoritaire, Français, ou DOM"))

```

```
# variable indiv_final$MIXIdesimmiPC (pour les EGO descendants d'immigrÃ©s)
# recoder
# recoder les NA quand GROUP1==3 car il peut avoir eu des enfants sans avoir Ã©tÃ© en couple

indiv_final$COUPLE_MIXTE_DESC <- factor(indiv_final$MIXIdesimmiPC, exclude=NULL) # on garde les NA
levels(indiv_final$COUPLE_MIXTE_DESC)[is.na(levels(indiv_final$COUPLE_MIXTE_DESC))]="Sans info" # nom vide pour
quelques individus
indiv_final$COUPLE_MIXTE_DESC <- plyr::revalue(indiv_final$COUPLE_MIXTE_DESC, c("1"="immigrÃ© mÃªme pays",
                                     "2"="immigrÃ© autre pays",
                                     "3"="descendant mÃªme pays",
                                     "6"="descendant mÃªme pays",
                                     "4"="descendant autre pays",
                                     "5"="groupe majoritaire ou DOM",
                                     "7"="groupe majoritaire ou DOM"))

# Ã cette Ã©tape : on a de nouvelles variables dans indiv_final


#####
#####
#####
#####
#####
#####
#####
#####
##### A P P A R I E M E N T #####
##### ENFANTS ET PARENTS #####
#
#
#####

#####
#
# Associer Ã chaque "enfant" de la base "enfants" les caractÃ©ristiques
# de son pÃ¨re et de sa mÃ¨re
#
# diplÃ©me, religion, Ã¢ge, durÃ©e de sÃ©jour, etc...
# but : avoir une grande base en combinant avec un extrait de indiv_final
#
#
#####

# la difficultÃ© de l'exercice : TeO ne recueille pas les filiations complÃ©tes

# crÃ©ation d'un identifiant unique : IDENTIFIANT
# pour servir de clÃ© d'appariement entre les bases
#####

#indiv$IDENTIFIANT<-paste(indiv$IDENT,indiv$Ne,sep="-")
indiv_final$IDENTIFIANT <- paste(indiv_final$IDENT,indiv_final$Ne,sep="-")
thl$IDENTIFIANT<-paste(thl$IDENT,thl$NOI,sep="-")
# attention, 7924 enfants n'ont pas de E_NOI_THL car ils
# ne rÃ©sident pas avec leur parent enquÃªtÃ© (EGO)
enfants$IDENTIFIANT<-paste(enfants$IDENT,enfants$E_NOI_THL,sep="-")

enfants$PRENOM_ID <- paste(enfants$E_PRENOM,enfants$IDENT)

# dÃ©composer indiv en 3 bases en fonction de conjoint, relation actuelle, premiÃ¨re relation
# avec les variables qui nous intÃ©ressent (pays naissance des parents du conjoint/relation)
# et on s'Ã©lectionne ensuite en fonction de l'annÃ©e de naissance de l'enfant
# table(indiv$C_lnp ,useNA = "always")

premier <- indiv_final[,c("IDENT","Nbunions","C_Relap","C_Anaisp","Regionnaisc2","C_Finap")]
premier$TYPE <- "PremierConjoint"
conjoint <- indiv_final[,c("IDENT","Nbunions","C_Relac","Anaïscă","Regionnaiscă2")] # attention C_Anaisc mal renseignÃ©
(utiliser Anaïsca)
```

```

conjoint$FIN <- 2009 # on fait comme si les unions actuelles couraient jusqu'en 2009, pour éviter les NA
conjoint$TYPE <- "ConjointActuel"
actuel <- indiv_final[c("IDENT", "Nbunions", "C_Relaa", "Anaïsa", "Regionnaïsa2")]
actuel$FIN <- 2009
actuel$TYPE <- "RelationActuelle"
colnames(premier) <- c("IDENT", "Nbunions", "DEBUT", "ANAIS_AUTPAR", "REGION_AUTPAR", "FIN", "TYPE")
colnames(conjoint) <- c("IDENT", "Nbunions", "DEBUT", "ANAIS_AUTPAR", "REGION_AUTPAR", "FIN", "TYPE")
colnames(actuel) <- c("IDENT", "Nbunions", "DEBUT", "ANAIS_AUTPAR", "REGION_AUTPAR", "FIN", "TYPE")

test <- rbind(premier, conjoint, actuel)
test <- test[order(test$IDENT),]

#
# cas le plus simple :
# l'enfant vit avec ses deux parents : c'est renseigné dans le THL
#

##### on part des informations du THL

m <- match(enfants$IDENTIFIANT, thl$IDENTIFIANT)
thl$COUPLEPARENTS <- (thl$MERIE==1 & thl$PERIE==1)
enfants$VIT_AVEC_COUPLEPARENTS <- thl$COUPLEPARENTS[m]
enfants$AUTRE_PARENT <- ifelse(enfants$VIT_AVEC_COUPLEPARENTS, "ConjointActuel", NA)
# 19943 enfants vivent avec leurs deux parents
# pour eux, tout est simple : on sait que l'autre parent est le conjoint actuel

# pour les autres... (11533) il faut chercher
# ça peut être le conjoint actuel s'il réside ailleurs ?
# on va faire l'hypothèse que ça ne peut pas être le conjoint actuel *si l'enfant est dans le THL*
# c'est soit le premier conjoint, soit la relation actuelle (non cohabitante)
test2
subset(enfants, is.na(enfants$AUTRE_PARENT))[, c("IDENT", "E_NOI_THL", "E_ANAIS", "E_PRENOM", "PRENOM_ID")]
test3 <- merge(test, test2, by="IDENT")
test3 <- subset(test3, !is.na(test3$DEBUT)) # on garde les relations renseignées

test3$KEEP <- ifelse(test3$E_ANAIS >= test3$DEBUT & test3$E_ANAIS <= test3$FIN, # si la naissance est dans les bornes
test3$TYPE, # alors on donne à KEEP le TYPE
NA ) # sinon c'est NA

test3 <- subset(test3, !is.na(test3$KEEP))
l <- test3[duplicated(test3$PRENOM_ID), "PRENOM_ID"]

test3 <- subset(test3, !test3$PRENOM_ID %in% l) # on enlève les individus dupliqués (75) pour lesquels il est impossible de
trouver l'autre parent

# On a renseigné 8903 enfants
# il reste 2630 enfants pour lesquels aucune information sur l'autre parent n'est disponible

m <- match(enfants$PRENOM_ID, test3$PRENOM_ID)
enfants$AUTRE_PARENT <- ifelse(is.na(enfants$AUTRE_PARENT), test3$KEEP[m], enfants$AUTRE_PARENT)

# on examine plus précisément les cas difficiles
# et là on attend la plage temporelle
test2
subset(enfants, is.na(enfants$AUTRE_PARENT))[, c("IDENT", "E_NOI_THL", "E_ANAIS", "E_PRENOM", "PRENOM_ID")]
test3 <- merge(test, test2, by="IDENT")
test3$KEEP <- ifelse(test3$E_ANAIS >= (test3$DEBUT-2) & test3$E_ANAIS <= (test3$FIN+2), # si la naissance est dans des bornes
largies
test3$TYPE, # alors on donne à KEEP le TYPE
NA ) # sinon c'est NA
test3 <- subset(test3, !is.na(test3$KEEP))
l <- test3[duplicated(test3$PRENOM_ID), "PRENOM_ID"]
test3 <- subset(test3, !test3$PRENOM_ID %in% l)
m <- match(enfants$PRENOM_ID, test3$PRENOM_ID)
enfants$AUTRE_PARENT <- ifelse(is.na(enfants$AUTRE_PARENT), test3$KEEP[m], enfants$AUTRE_PARENT)
# on peut même ajouter une variable "degré de certitude de l'appariement"

# il reste 2249 enfants sans information sur l'autre parent
# on indique "non renseigné"
enfants$AUTRE_PARENT[is.na(enfants$AUTRE_PARENT)] <- "NonRenseigne"

#
#
# et maintenant on doit appairer enfants avec les variables de indiv_final
#

```

```

#

# extraction des informations sur le premier conjoint
premier <- indiv_final[,c("IDENT",
  "Sexee", # sexe de EGO pour avoir sexe de autre parent et d'Ã©terminer qui est pere, qui est mere
  "Nbunions",
  "C_Relap",
  "C_Anaisp",
  "Regionnaispc2",
  "Lieunaispcr",
  "Group1pc",
  "Regionnaispcp2",
  "Regionnaispcm2",
  "Statumigpcp",
  "Statumigpcm",
  "TYPEPMARP", # type du premier mariage passÃ©
  "C_Isced_P", # niveau de diplome
  "C_Finap")]
premier$TYPE <- "PremierConjoint"
premier$CS_CJ <- NA
premier$CS_PERECJ <- NA

conjoint <- indiv_final[,c("IDENT",
  "Sexee",
  "Nbunions",
  "C_Relac",
  "Anaisca", # attention C_Anaisc mal renseignÃ©e (utiliser Anaisca)
  "Regionnaisca2",
  "Group1ca",
  "Lieunaiscar",
  "Regionnaiscap2",
  "Regionnaiscam2",
  "Statumigcap",
  "Statumigcam",
  "TYPEMARA",
  "C_Isced_C", # niveau de diplome
  "CS_CJ", # CS conjoint : que pour le conjoint actuel,
  "CS_PERECJ")] # CS pÃ©re du conjoint : idem
conjoint$FIN <- 2009 # on fait comme si les unions actuelles couraient jusqu'en 2009, pour Ã©viter les NA
conjoint$TYPE <- "ConjointActuel"

actuel <- indiv_final[,c("IDENT",
  "Sexee",
  "Nbunions",
  "C_Relaa",
  "Anaisca",
  "Regionnaisca2",
  "Group1ca",
  "Lieunaiscar",
  "Regionnaiscap2",
  "Regionnaiscam2",
  "Statumigcap",
  "Statumigcam",
  "TYPEMARA",
  "C_Isced_A")]
actuel$FIN <- 2009
actuel$TYPE <- "RelationActuelle"
actuel$CS_CJ <- NA
actuel$CS_PERECJ <- NA

colnames(premier) <- c("IDENT",
  "Sexee",
  "Nbunions",
  "DEBUT",
  "ANAIIS_AUTPAR",
  "REGION_AUTPAR",
  "LIEUNAIIS_AUTPAR",
  "GROUP1_AUTPAR",
  "REGION_PERE_AUTPAR",
  "REGION_MERE_AUTPAR",
  "STATUMIG_PERE_AUTPAR",
  "STATUMIG_MERE_AUTPAR",
  "TYPEMARIAGE", "DIPLOME_AUTPAR",
  "FIN", "TYPE", "CS_AUTPAR", "CS_PERE_AUTPAR")

```



```

colnames(conjoint) <- c("IDENT",
  "Sexee",
  "Nbunions",
  "DEBUT",
  "ANAIIS_AUTPAR",
  "REGION_AUTPAR",
  "GROUPE1_AUTPAR",
  "LIEUNAIS_AUTPAR",
  "REGION_PERE_AUTPAR",
  "REGION_MERE_AUTPAR",
  "STATUMIG_PERE_AUTPAR",
  "STATUMIG_MERE_AUTPAR",
  "TYPEMARIAGE", "DIPLOME_AUTPAR",
  "CS_AUTPAR", "CS_PERE_AUTPAR",
  "FIN", "TYPE")
colnames(actuel) <- c("IDENT",
  "Sexee",
  "Nbunions",
  "DEBUT",
  "ANAIIS_AUTPAR",
  "REGION_AUTPAR",
  "GROUPE1_AUTPAR",
  "LIEUNAIS_AUTPAR",
  "REGION_PERE_AUTPAR",
  "REGION_MERE_AUTPAR",
  "STATUMIG_PERE_AUTPAR",
  "STATUMIG_MERE_AUTPAR",
  "TYPEMARIAGE", "DIPLOME_AUTPAR",
  "FIN", "TYPE", "CS_AUTPAR", "CS_PERE_AUTPAR")

test <- rbind(premier, conjoint, actuel)
test <- test[order(test$IDENT),]

# df est la base enfants enrichie avec infos sur 2e parent
# note : le "conjoint actuel" est bien renseigné (profession, profession du père, diplomes)
# ce n'est pas le cas pour les Relations amoureuses actuelles (on a le diplôme, pas les professions du partenaire et de son pere)
# ce n'est pas le cas pour le Premier conjoint (on a juste le diplôme)
# ca limite l'intérêt de la prise en compte de cette variable
df <- merge(enfants, test, by.x=c("IDENT", "AUTRE_PARENT"), by.y=c("IDENT", "TYPE"), all.x=T)

#####
#####
##
##
## Association des variables de "indiv_final" à "df"
##
##
##
#####
#####

# associer les variables de "indiv_final" à df :
# on fait attention à nommer de la même manière _EGO et _AUTPAR
# but : rendre compatible cette base "df" avec "indiv_final" pour les combiner
m <- match(df$IDENT, indiv_final$IDENT)
df$Sexee <- indiv_final$Sexee[m]
df$ANAIIS_EGO <- indiv_final$Anaise[m]
df$REGION_EGO <- indiv_final$Regionnaise2[m] # région de naissance de EGO (le parent enquêté)
df$GROUPE1_EGO <- indiv_final$GROUPE1[m]
df$LIEUNAIS_EGO <- indiv_final$Lieuнаiser[m]
df$REGION_PERE_EGO <- indiv_final$Regionnaise2[m]
df$REGION_MERE_EGO <- indiv_final$Regionnaism2[m]
df$STATUMIG_PERE_EGO <- indiv_final$STATUMIGP[m]
df$STATUMIG_MERE_EGO <- indiv_final$STATUMIGM[m]
df$DIPLOME_EGO <- indiv_final$SF_Isced[m]
df$CS_EGO <- indiv_final$CS42_EGO[m] # nomenclature CS42 pour être raccord avec CS_CJ
df$CS_PERE_EGO <- indiv_final$CS_PERE[m]

df_p <- subset(df, df$Sexee==1) # table des enfants dont EGO est père
df_m <- subset(df, df$Sexee==2) # table des enfants dont EGO est mère

```

```

noms_colonnes <- colnames(df_p)
noms_colonnes <- gsub("AUTPAR","MERE",noms_colonnes)
noms_colonnes <- gsub("EGO","PERE",noms_colonnes)
colnames(df_p) <- noms_colonnes
df_p$E_PERE <- NA # pour l'ancienne colonne E_AUTPAR (vit avec l'autre parent)

noms_colonnes <- colnames(df_m)
noms_colonnes <- gsub("AUTPAR","PERE",noms_colonnes)
noms_colonnes <- gsub("EGO","MERE",noms_colonnes)
colnames(df_m) <- noms_colonnes
df_m$E_MERE <- NA

# df est maintenant compatible avec "indiv_final" (modulo quelques arrangements)
df <- rbind(df_p,df_m)
# attention E_LN : France 1, etranger 2
# attention indiv_final$Lieunaiser : 0 France 1 DOM 2 ETRANGER

# on cr e une variable GROUP1
df$GROUP1 <- ifelse(df$E_LN%in%c(2,8,9) & df$E_NAT%in%c(2,8,9),1, # enfant n  l tranger et aujourd'hui de
nationalit  tr re
  ifelse(df$E_LNDEP%in%c("99","9A","9B","9C","9D","9E","9F","9G","9H","9I"),2,
    ifelse((df$GROUP1_MERE==1) | (df$GROUP1_PERE==1),3,
      ifelse((df$GROUP1_MERE==2) | (df$GROUP1_PERE==2),4, 5))))
# df$GROUP1[is.na(df$GROUP1)] <- 6 ### cr er peut  tre une variable "GROUP1 == 6" mal renseign 

enfants_parents <- df

enfants_parents$NB <- 1

#
# Recoder les r gions en clair
#
#

enfants_parents$REGION_E <- ifelse(enfants_parents$E_LNPIALF=="", "France",
  ifelse(enfants_parents$E_LNPIALF %in% c("PRT","ESP","ITA"), "POR-ESP-ITA",
    ifelse(enfants_parents$E_LNPIALF %in% c("DZA","TUN","MAR"), "ALG-MAR-TUN",
      ifelse(enfants_parents$E_LNPIALF %in%
c("SEN","MRT","GMB","GNB","GIN","MLI","BFA","NER","TCD","SDN"), "SAH","AUTRE"))))

enfants_parents$REGION_PERE2 <- ifelse(enfants_parents$REGION_PERE %in%
c(1101,1201,1202,1203,1204,1205,4410,4404), "France",
  ifelse(enfants_parents$REGION_PERE %in% c(2101,2201,2301), "ALG-MAR-TUN",
    ifelse(enfants_parents$REGION_PERE %in%
c(2401,2402,2403,2404,2405,2406,2407,2408,2409), "SAH",
      ifelse(enfants_parents$REGION_PERE %in% c(4101,4201,4301), "POR-ESP-ITA",
        ifelse(enfants_parents$REGION_PERE %in%
c(2501,2502,2503,2504,2505,2506,2507,2508,2509,2510,2511,2512), "Afrique centrale",
          ifelse(enfants_parents$REGION_PERE %in% c(2601:2901), "Autre Afrique",
            ifelse(enfants_parents$REGION_PERE %in% c(3101:3600), "Asie (avec Turquie)",
              ifelse(is.na(enfants_parents$REGION_PERE), "NR", "AUTRE"))))))))

enfants_parents$REGION_MERE2 <- ifelse(enfants_parents$REGION_MERE %in%
c(1101,1201,1202,1203,1204,1205,4410,4404), "France",
  ifelse(enfants_parents$REGION_MERE %in% c(2101,2201,2301), "ALG-MAR-TUN",
    ifelse(enfants_parents$REGION_MERE %in%
c(2401,2402,2403,2404,2405,2406,2407,2408,2409), "SAH",
      ifelse(enfants_parents$REGION_MERE %in% c(4101,4201,4301), "POR-ESP-ITA",
        ifelse(enfants_parents$REGION_MERE %in%
c(2501,2502,2503,2504,2505,2506,2507,2508,2509,2510,2511,2512), "Afrique centrale",
          ifelse(enfants_parents$REGION_MERE %in% c(2601:2901), "Autre Afrique",
            ifelse(enfants_parents$REGION_MERE %in% c(3101:3600), "Asie (avec Turquie)",
              ifelse(is.na(enfants_parents$REGION_MERE), "NR", "AUTRE"))))))))

# les grands parents
# nombre de grands parents Maghreb
# quand l'information est NA on consid re que le grand parent N'EST PAS NE au Maghreb

```

```

enfants_parents$NB_GP_MAG <- rowSums( cbind((enfants_parents$REGION_PERE_MERE %in% c(2101,2201,2301)),
      (enfants_parents$REGION_PERE_PERE %in% c(2101,2201,2301)),
      (enfants_parents$REGION_MERE_MERE %in% c(2101,2201,2301)),
      (enfants_parents$REGION_MERE_PERE %in% c(2101,2201,2301))), na.rm=T)

enfants_parents$NB_GP_EURSUD <- rowSums( cbind((enfants_parents$REGION_PERE_MERE %in% c(4101,4201,4301)),
      (enfants_parents$REGION_PERE_PERE %in% c(4101,4201,4301)),
      (enfants_parents$REGION_MERE_MERE %in% c(4101,4201,4301)),
      (enfants_parents$REGION_MERE_PERE %in% c(4101,4201,4301))), na.rm=T)

#####
# CatÃ©gorie des prÃ©noms, en clair

indiv_final <- indiv_final %>%
  mutate(type_prenom = case_when(CAT1_corr %in% c(129) ~ "franÃ§ais",
    CAT1_corr %in% c(168,697,368,157,611) ~ "international_us",
    CAT1_corr %in% c(622,641,662,221,252,643) ~ "arabe_plus",
    CAT1_corr %in% c(661,696,137,153,161) ~ "international_sud",
    CAT1_corr %in% c(621,623,231,232,238,243,244,249,251,262,263,269,283) ~ "afrique",
    CAT1_corr %in% c(698) ~ "inclassables",
    CAT1_corr %in% c(163) ~ "turquie",
    CAT1_corr %in% c(422:479,645,644) ~ "asie",
    TRUE ~ "divers"))

enfants <- enfants %>%
  mutate(type_prenom = case_when(CAT1_corr %in% c(129) ~ "franÃ§ais",
    CAT1_corr %in% c(168,697,368,157,611) ~ "international_us",
    CAT1_corr %in% c(622,641,662,221,252,643) ~ "arabe_plus",
    CAT1_corr %in% c(661,696,137,153,161) ~ "international_sud",
    CAT1_corr %in% c(621,623,231,232,238,243,244,249,251,262,263,269,283) ~ "afrique",
    CAT1_corr %in% c(698) ~ "inclassables",
    CAT1_corr %in% c(163) ~ "turquie",
    CAT1_corr %in% c(422:479,645,644) ~ "asie",
    TRUE ~ "divers"))

#####
#####
##### Sauvegarde gÃ©nÃ©rale
#####

save(enfants,indiv_final,thl,enfants_parents,file="TeO_utile.Rdata")

rm(list=ls())

❖ 3e programme informatique communiquÃ© par les auteurs (03-tables)
#####
##### transmission d'un "prÃ©nom spÃ©cifique"
##### en fonction de l'origine
##### Afrique du Nord (sans sahel)
##### et europe du sud
#####

library(tidyverse)

# * * * * *
#
# charger TeO utile
#
# * * * * *

## charger d'abord "enfants" (rÃ©sultat de l'appariement avec pÃ©res et mÃ©res de indiv_final)
# dans objet "enfants_parents"
load(file="TeO_utile.Rdata")

```

```

# on sélectionne une sous-base qui ne contient que les variables qui nous intéressent
# dans une version précédente, on gardait GROUP1==3, mais c'est pas bon
# on garde GROUP1==3 car on s'intéresse aux ego nés en France de parents immigrés
ego <- indiv_final %>%
  # filter(GROUP1==3) %>%
  select(REGION_PERE2, REGION_MERE2, REGION_EGO,
         ORIGINE_TOUS,
         Lieunaiss, # lieu de naissance d'EGO : 0 = France
         Inaisprts_b, # lieu de naissance des parents de EGO
         T_Enfmer, # nombre d'enfants que la mere de EGO a eu
         Anaise, FNI2, GROUP1, Poids, IDENT,
         FNI_MOBILE,
         Sexee, GROUP1, GROUP2,
         discredm, discriemploi_m, discriosphere_m, compteurlog_m, discrierv_m, discri, D_Nom,
         duretu,
         P_SALMENS_EGO, P_Nsremu, DESC, # desc : descendant d'immigré, modalités région d'origine
         Prenom1,
         ZUS2007, Inaisprts_b, # lieu de naissance des parents
         Active,
         CAT0_corr, CAT1_corr)

# idem pour la base enfants
# mais ici, il y a tous les enfants, nés en France ou non

## base enfants

enfego <- enfants %>%
  left_join(indiv_final, by="IDENT") %>%
  select(REGION_EGO, AGE08,
         ORIGINE_TOUS,
         CSNQ_ego,
         GROUP1, Group1ca, # pour étudier le couple conjugal/parental
         REGION_CONJOINT2, # région du conjoint actuel
         E_AUTPAR, # pour savoir si l'autre parent est dans le logement
         E_LN, E_SEXE, E_ANAIS, ANARRIVE,
         MERE_LN, PERE_LN,
         NBENF, ## nombre d'enfants de EGO
         duretu, # durée des études de EGO parent
         MIXlimmiPC,
         COUPLE_MIXTE_IMMI, ## couple mixte pour les EGO immigrés
         COUPLE_MIXTE_DESC, ## couple mixte, pour EGO descendant d'immigré
         REGION_PERE2, REGION_MERE2, ## région de naissance des parents de EGO
         DAdActIRIS_PCT_IMMI, ## domicile immigrés iris actuel
         DAdAntIRIS_PCT_IMMI, # domicile immigré iris antérieur
         A_Rorig, ## amis de la même origine, pour ceux qui ont déclaré des amis
         A_Rrelig, ## amis de la même religion, pour ceux qui ont déclaré des amis
         FNI2.x, Sexee,
         FNI_MOBILE.x,
         Anaise, Anaisca, # année de naissance du conjoint actuel
         Poids.x, IDENT, Prenom1.x, ##### remettre Prenom1 ? Prenom1.x = prénom de l'enfant
         CAT0_corr.x, CAT1_corr.x, # classification des prénoms
         E_PRENOM, E_LVIE, # prénom et lieu de vie
         L_Demen) # Année du déménagement

# attention : comme Prenom1 se trouve dans indiv_final et dans enfants,
# Prenom1.x est le prénom de l'enfant
# vérifier que ça ne pose pas de problème pour la suite !

enfego <- enfego %>% mutate(Duree_inst_naissance = E_ANAIS-ANARRIVE)
enfego <- enfego %>% mutate(iris_immi =
  ifelse(DAdActIRIS_PCT_IMMI %in% c(0:6), "iris_peu_immi", ifelse(DAdActIRIS_PCT_IMMI %in% c(7,8), "iris_moy_immi", "iris_sup_immi")
  )))

ego$cat_prenom <- ifelse(ego$CAT1_corr %in% c(129), "français",
  ifelse(ego$CAT1_corr %in% c(168,697,368,160,611), "international_us",
    ifelse(ego$CAT1_corr %in% c(622,641,662,221,252,621,623), "arabe_plus",
      ifelse(ego$CAT1_corr %in% c(696,698,137,153,161), "international_sud", "autre"))))
ego <- ego %>% mutate(type_prenom = case_when(CAT1_corr %in% c(129) ~ "1-français",
  CAT1_corr %in% c(168,697,368,157,611) ~ "international_us",
  CAT1_corr %in% c(622,641,662,221,252,643) ~ "arabe_plus",
  CAT1_corr %in% c(661,696,137,153,161) ~ "international_sud",

```

```

CAT1_corr %in% c(621,623,231,232,238,243,244,249,251,262,263,269,283) ~ "afrique",
CAT1_corr %in% c(698) ~ "inclassables",
CAT1_corr %in% c(163) ~ "turquie",
CAT1_corr %in% c(422:479,645,644) ~ "asie",
TRUE ~ "divers"))
enfego$cat_prenom <- ifelse(enfego$CAT1_corr.x %in% c(129),"franÃ§ais",
  ifelse(enfego$CAT1_corr.x %in% c(168,697,368,160,611),"international_us",
    ifelse(enfego$CAT1_corr.x %in% c(622,641,662,221,252,621,623),"arabe_plus",
      ifelse(enfego$CAT1_corr.x %in% c(696,698,137,153,161),"international_sud", "autre")))))
enfego <- enfego %>% mutate(type_prenom = case_when(CAT1_corr.x %in% c(129) ~ "1-franÃ§ais",
  CAT1_corr.x %in% c(168,697,368,157,611) ~ "international_us",
  CAT1_corr.x %in% c(622,641,662,221,252,643) ~ "arabe_plus",
  CAT1_corr.x %in% c(661,696,137,153,161) ~ "international_sud",
  CAT1_corr.x %in% c(621,623,231,232,238,243,244,249,251,262,263,269,283) ~ "afrique",
  CAT1_corr.x %in% c(698) ~ "inclassables",
  CAT1_corr.x %in% c(163) ~ "turquie",
  CAT1_corr.x %in% c(422:479,645,644) ~ "asie",
  TRUE ~ "divers"))

enfants_parents$cat_prenom <- ifelse(enfants_parents$CAT1_corr %in% c(129),"franÃ§ais",
  ifelse(enfants_parents$CAT1_corr %in% c(168,697,368,160,611),"international_us",
    ifelse(enfants_parents$CAT1_corr %in% c(622,641,662,221,252,621,623),"arabe_plus",
      ifelse(enfants_parents$CAT1_corr %in% c(696,698,137,153,161),"international_sud", "autre")))))
enfants_parents <- enfants_parents %>% mutate(type_prenom = case_when(CAT1_corr %in% c(129) ~ "1-franÃ§ais",
  CAT1_corr %in% c(168,697,368,157,611) ~ "international_us",
  CAT1_corr %in% c(622,641,662,221,252,643) ~ "arabe_plus",
  CAT1_corr %in% c(661,696,137,153,161) ~ "international_sud",
  CAT1_corr %in% c(621,623,231,232,238,243,244,249,251,262,263,269,283) ~
"afrique",
  CAT1_corr %in% c(698) ~ "inclassables",
  CAT1_corr %in% c(163) ~ "turquie",
  CAT1_corr %in% c(422:479,645,644) ~ "asie",
  TRUE ~ "divers"))

indiv_final$cat_prenom <- ifelse(indiv_final$CAT1_corr %in% c(129),"franÃ§ais",
  ifelse(indiv_final$CAT1_corr %in% c(168,697,368,160,611),"international_us",
    ifelse(indiv_final$CAT1_corr %in% c(622,641,662,221,252,621,623),"arabe_plus",
      ifelse(indiv_final$CAT1_corr %in% c(696,698,137,153,161),"international_sud", "autre")))))

#
tmp_1 <- ego %>% mutate(var=case_when(GROUP1 %in% c(5,2,4) & Lieunaizer==0 ~ "natif",
  GROUP1 == 1 ~ "immigre",
  GROUP1 == 3 ~ "descendant",
  TRUE ~ "autre")) %>%
filter((REGION_PERE2=="POR-ESP-ITA")|(REGION_MERE2=="POR-ESP-ITA")|REGION_EGO=="POR-ESP-ITA") %>%
group_by(Anaise,var) %>%
summarize(p=weighted.mean(FNI2,Poids),n=n(),
  p_mobile=weighted.mean(FNI_MOBILE,Poids),
  p_spe = sum(Poids[cat_prenom=="international_sud"])/sum(Poids)) %>%
mutate(annee=Anaise,type="ego") %>%
ungroup() %>%
select(-Anaise)

tmp_2 <- enfants_parents %>% mutate(var=case_when(GROUP1_MERE %in% c(5,2,4) & GROUP1_PERE %in% c(5,2,4) &
E_LN==1 ~ "natif",
  GROUP1_MERE == 1 & GROUP1_PERE == 1 & E_LN == 2 ~ "immigre",
  (GROUP1_MERE == 1 | GROUP1_PERE == 1) & E_LN == 1 ~ "descendant",
  (GROUP1_MERE == 3 | GROUP1_PERE == 3) & E_LN == 1 ~ "enfant de descendant",
  TRUE ~ "autre")) %>%
filter((REGION_PERE2=="POR-ESP-ITA")|(REGION_MERE2=="POR-ESP-ITA") | NB_GP_EURSUD > 0) %>%
group_by(E_ANAIS,var) %>%
summarize(n=n(),
  p_mobile = weighted.mean(FNI_MOBILE,Poids,na.rm=T),
  p_spe = sum(Poids[type_prenom=="international_sud"])/sum(Poids)) %>%
filter(E_ANAIS>1977,E_ANAIS<2009) %>%
ungroup() %>%
mutate(annee=E_ANAIS,type="enfant") %>%
select(-E_ANAIS)

prenoms_sudeurope <- bind_rows(tmp_1,tmp_2)

```

```

tmp_1 <- ego %>% mutate(var=case_when(GROUP1 %in% c(5,2,4) & Lieunaiser==0 ~ "natif",
                                     GROUP1 == 1 ~ "immigre",
                                     GROUP1 == 3 ~ "descendant",
                                     TRUE ~ "autre")) %>%
  filter((REGION_PERE2=="ALG-MAR-TUN")|(REGION_MERE2=="ALG-MAR-TUN")| (REGION_EGO=="ALG-MAR-
TUN")) %>%
  group_by(Anaise,var) %>%
  summarize(p=weighted.mean(FNI2,Poidsi),n=n(),
            p_mobile=weighted.mean(FNI_MOBILE,Poidsi),
            p_spe = sum(Poidsi[type_prenom %in% c("arabe_plus")])/sum(Poidsi)) %>%
  mutate(annee=Anaise,type="ego") %>%
  ungroup() %>%
  select(-Anaise)

tmp_2 <- enfants_parents %>% mutate(var=case_when(GROUP1_MERE %in% c(5,2,4) & GROUP1_PERE %in% c(5,2,4) &
E_LN==1 ~ "natif",
                                     GROUP1_MERE == 1 & GROUP1_PERE == 1 & E_LN == 2 ~ "immigre",
                                     (GROUP1_MERE == 1 | GROUP1_PERE == 1) & E_LN == 1 ~ "descendant",
                                     (GROUP1_MERE == 3 | GROUP1_PERE == 3) & E_LN == 1 ~ "enfant de descendant",
                                     TRUE ~ "autre")) %>%
  filter((REGION_PERE2=="ALG-MAR-TUN") | (REGION_MERE2=="ALG-MAR-TUN") | NB_GP_MAG > 0) %>%
  group_by(E_ANAIS,var) %>%
  summarize(n=n(),
            p_mobile=weighted.mean(FNI_MOBILE,Poidsi,na.rm=T),
            p_spe = sum(Poidsi[type_prenom %in% c("arabe_plus")])/sum(Poidsi)) %>%
  filter(E_ANAIS>1977,E_ANAIS<2009) %>%
  ungroup() %>%
  mutate(annee=E_ANAIS,type="enfant") %>%
  select(-E_ANAIS)

prenoms_afnord <- bind_rows(tmp_1,tmp_2)

evolution <-
  bind_rows(
    prenom_afnord %>% group_by(var) %>%
      summarize(p_spe = weighted.mean(p_spe,n),
                annee=weighted.mean(annee,n),
                N=sum(n),
                origine="Afrique du Nord")
    ,
    prenom_sudeurope %>% group_by(var) %>%
      summarize(p_spe = weighted.mean(p_spe,n),
                annee=weighted.mean(annee,n),
                N=sum(n),
                origine="Europe du Sud") )

# prÃ©noms les plus frÃ©quents
# pour les immigrÃ©s,
# les descendants
# les enfants de descendants

tmp_1 <- ego %>% mutate(var=case_when(GROUP1 %in% c(5,2,4) & Lieunaiser==0 ~ "natif",
                                     GROUP1 == 1 ~ "immigre",
                                     GROUP1 == 3 ~ "descendant",
                                     TRUE ~ "autre")) %>%
  filter((REGION_PERE2=="POR-ESP-ITA")|(REGION_MERE2=="POR-ESP-ITA")|REGION_EGO=="POR-ESP-ITA") %>%
  group_by(Prenom1,var) %>%
  summarize(p=weighted.mean(FNI2,Poidsi),n=n(),
            p_mobile=weighted.mean(FNI_MOBILE,Poidsi),
            p_spe = sum(Poidsi[cat_prenom=="international_sud"])/sum(Poidsi)) %>%
  mutate(prenom=Prenom1,type="ego") %>%
  ungroup()

tmp_2 <- enfants_parents %>% mutate(var=case_when(GROUP1_MERE %in% c(5,2,4) & GROUP1_PERE %in% c(5,2,4) &
E_LN==1 ~ "natif",
                                     GROUP1_MERE == 1 & GROUP1_PERE == 1 & E_LN == 2 ~ "immigre",
                                     (GROUP1_MERE == 1 | GROUP1_PERE == 1) & E_LN == 1 ~ "descendant",

```

```

      (GROUP1_MERE == 3 | GROUP1_PERE == 3) & E_LN == 1 ~ "enfant de descendant",
      TRUE ~ "autre")) %>%
filter((REGION_PERE2=="POR-ESP-ITA")|(REGION_MERE2=="POR-ESP-ITA") | NB_GP_EURSUD > 0) %>%
group_by(Prenom1,var) %>%
summarize(n=n(),
  p_mobile = weighted.mean(FNI_MOBILE,Poids1,na.rm=T),
  p_spe = sum(Poids1[type_prenom=="international_sud"])/sum(Poids1)) %>%
# filter(E_ANAIS>1977,E_ANAIS<2009) %>%
ungroup() %>%
mutate(prenom=Prenom1,type="enfant")

prenoms_sudeurope <- bind_rows(tmp_1,tmp_2)

prenoms_sudeurope %>% group_by(var,prenom) %>% summarize(N=sum(n)) %>%
group_by(var) %>% mutate(rang=rank(-N,ties.method = "random")) %>%
filter(rang<20,var %in% c("immigre","descendant","enfant de descendant")) %>%
select(-N) %>% spread(key=var,value=prenom)

# et pour l'Afrique du Nord

tmp_1 <- ego %>% mutate(var=case_when(GROUP1 %in% c(5,2,4) & Lieunaiser==0 ~ "natif",
  GROUP1 == 1 ~ "immigre",
  GROUP1 == 3 ~ "descendant",
  TRUE ~ "autre")) %>%
filter((REGION_PERE2=="ALG-MAR-TUN")|(REGION_MERE2=="ALG-MAR-TUN")|REGION_EGO=="ALG-MAR-
TUN") %>%
group_by(Prenom1,var) %>%
summarize(p=weighted.mean(FNI2,Poids1),n=n(),
  p_mobile=weighted.mean(FNI_MOBILE,Poids1),
  p_spe = sum(Poids1[cat_prenom=="arabe_plus"])/sum(Poids1)) %>%
mutate(prenom=Prenom1,type="ego") %>%
ungroup()

tmp_2 <- enfants_parents %>% mutate(var=case_when(GROUP1_MERE %in% c(5,2,4) & GROUP1_PERE %in% c(5,2,4) &
E_LN==1 ~ "natif",
  GROUP1_MERE == 1 & GROUP1_PERE == 1 & E_LN == 2 ~ "immigre",
  (GROUP1_MERE == 1 | GROUP1_PERE == 1) & E_LN == 1 ~ "descendant",
  (GROUP1_MERE == 3 | GROUP1_PERE == 3) & E_LN == 1 ~ "enfant de descendant",
  TRUE ~ "autre")) %>%
filter((REGION_PERE2=="ALG-MAR-TUN")|(REGION_MERE2=="ALG-MAR-TUN") | NB_GP_MAG > 0) %>%
group_by(Prenom1,var) %>%
summarize(n=n(),
  p_mobile = weighted.mean(FNI_MOBILE,Poids1,na.rm=T),
  p_spe = sum(Poids1[type_prenom=="arabe_plus"])/sum(Poids1)) %>%
# filter(E_ANAIS>1977,E_ANAIS<2009) %>%
ungroup() %>%
mutate(prenom=Prenom1,type="enfant")

prenoms_afnord <- bind_rows(tmp_1,tmp_2)

prenoms_afnord %>% group_by(var,prenom) %>% summarize(N=sum(n)) %>%
group_by(var) %>% mutate(rang=rank(-N,ties.method = "random")) %>%
filter(rang<10,var %in% c("immigre","descendant","enfant de descendant")) %>% View()
select(-N) %>% spread(key=var,value=prenom)

```



## **Annexe 5A : La base de données**

Cette Annexe 5A, associée au Chapitre 5, p. 24, comprend la méthodologie de la reproduction « au plus près » concernant la base de données. Je précise l'échantillon, les variables et leurs (re)codages, les pondérations et la table que j'utilise, avant d'évaluer sa validité. Je présente alors quelques résultats préliminaires, qui permettent de replacer les analyses de Coulmont et Simon (2019) dans leur contexte. Le programme informatique de la reproduction « au plus près » figure dans l'Annexe 5B, p. 183.

### **❖ Des programmes informatiques insuffisants**

Les 3 programmes informatiques R que les auteurs et l'INED ont finalement communiqués le 6 janvier 2021 (Annexe 4B, p. 124) sont inutilisables et incomplets (Chapitre 4, p. 15), si bien qu'ils ne dispensent pas de reprendre l'analyse des prénoms au début.

### **❖ L'échantillon maximal**

L'enquête TeO a été menée en face à face auprès de 21 761 personnes d'âge actif (18-50 ou 18-60 ans selon le sous-échantillon, comme indiqué en Annexe 1B, p. 91) vivant en ménage ordinaire en France métropolitaine en 2008. Son taux de réponse a été de 61 %. Des 22 tables issues de la version nominative de l'enquête TeO<sup>61</sup> que le CASD communique aux chercheurs, seules 3 renseignent le prénom de certaines personnes<sup>62</sup> et pourraient donc avoir été utilisées par Coulmont et Simon (2019) pour analyser des prénoms :

- une première table, appelée `THL2_PRENOM`, renseigne le prénom et certaines caractéristiques des 75 121 habitants du logement des enquêtés ; comme le montre un tri à plat de la variable (`LIENEGO`) indiquant le lien entre chaque habitant du logement et l'enquêté (Insee 2011, p. 291), cet effectif comprend les 21 761 enquêtés et leurs 23 552 enfants qui habitent le même logement qu'eux, ainsi que les 29 808 autres habitants de leur logement (qui comprennent leurs 13 243 conjoints à la date de l'enquête, 7 647 parents, 6 713 frères et sœurs, 127 petits-enfants, 81 grands-parents, 639 beaux-enfants, 391 beaux-parents, 697 autres relations familiales, 89 amis et 181 individus sans lien familial ni amical avec l'enquêté) ;

---

<sup>61</sup> Baptiste Coulmont a eu accès à cette base de données nominative de l'enquête TeO en 2013 (comme indiqué [ici](#) et [là](#)) et en 2016-2017, pendant sa période de [délégation](#) dans l'unité de l'INED où travaille Patrick Simon (Coulmont 2017a, 2017b et 2017c). Patrick Simon a quant à lui directement accès à la base. De mon côté j'ai déposé auprès du Comité du secret statistique une « Demande d'accès à des données confidentielles pour un nouveau projet » le 5 juin 2019, elle a été examinée et a fait l'objet d'un avis favorable le 11 octobre 2019, puis elle a été acceptée le 13 décembre 2019. J'ai alors pu m'inscrire au Centre d'accès sécurisé aux données (CASD) le 16 décembre 2019, établir les contrats relatifs à [mon projet de recherche](#) le 28 janvier 2020 (facture de 2 856 euros) et assister à une séance d'enrôlement de trois heures le 3 février 2020. J'ai ensuite préparé l'exploitation statistique de cette base, avant de la réaliser sur la SD-Box de juin 2020 à juillet 2021 (Annexe 5B, p. 183). J'en profite pour dire l'importance majeure, pour la répliquabilité — et donc la fiabilité — de la recherche en sciences sociales, d'institutions comme les Archives de données issues de la statistique publique (ADISP) et le Centre d'accès sécurisé aux données (CASD).

<sup>62</sup> Ces 3 tables sont `THL2_PRENOM`, `INDIV2_PRENOM` et `ENFANTS2_PRENOM`. 17 autres tables ne contiennent aucun prénom et ne sont pas utilisées dans la présente reproduction : il s'agit des tables `THL_B`, `THL2`, `THL2_PAYS3`, `INDIV_B`, `INDIV2`, `INDIV2_LANG`, `INDIV2_PAYS3`, `ENFANTS_B`, `ENFANTS2`, `ENFANTS2_PAYS3`, `TRAJMIG2`, `TRAJMIG2_PAYS3`, `TRAJPRO2`, `AUTLOG2`, `AUTLOG2_PAYS3`, `CONTEXT` et `JEUNES`. La 21<sup>e</sup> table, `POLREL2`, ne contient pas de prénom mais je l'utilise quand même pour renseigner les affiliations religieuses déclarées par l'enquêté. Enfin la 22<sup>e</sup> table, `INDIV_B_PRENOM`, contient bien des prénoms mais ce sont les mêmes que ceux contenus dans la table `INDIV2_PRENOM`, qu'il est préférable d'utiliser. Pour plus de précisions, voir Annexe 5B, p. 183.

- une deuxième table, appelée INDIV2\_PRENOM, renseigne le prénom et de nombreuses autres caractéristiques des 21 761 individus enquêtés ; en d’autres termes, la table INDIV2\_PRENOM ne renseigne pas de prénom qui ne soit déjà renseigné dans la table THL2\_PRENOM ;
- une troisième table, appelée ENFANTS2\_PRENOM, renseigne le prénom et certaines autres caractéristiques des 31 476 enfants des enquêtés, c’est-à-dire leurs 23 552 enfants qui habitent le même logement que les enquêtés mais aussi leurs 7 924 enfants qui habitent un autre logement ou sont décédés à la date de l’enquête ; en d’autres termes, les seuls prénoms que renseigne la table ENFANTS2\_PRENOM mais que ne renseigne pas déjà la table THL2\_PRENOM sont ceux des 7 924 enfants de l’enquêté qui habitent un autre logement ou sont décédés à la date de l’enquête.

L’enquête TeO enrichie de la variable « prénom » renseigne donc le prénom des 75 121 habitants du logement de l’enquêté ainsi que le prénom des 7 924 enfants de l’enquêté qui habitent un autre logement que le sien ou qui sont décédés — mais aucun autre prénom. Par conséquent, pour pouvoir reproduire les analyses de Coulmont et Simon (2019) sur *tous* les échantillons possibles, je dois construire une table qui englobe l’échantillon maximal, c’est-à-dire l’échantillon couvrant tous les habitants du logement de l’enquêté ainsi que les enfants de l’enquêté qui habitent un autre logement ou qui sont décédés. Cette table de reproduction, que j’appellerai REPRO, sera donc composée de  $75\,121 + 7\,924 = 83\,045$  individus (en lignes).

### ❖ Les variables

Cette table de reproduction REPRO est enrichie de certaines informations (variables, colonnes) issues de la table THL2\_PRENOM (à partir de laquelle REPRO est construite), mais aussi de variables issues des tables INDIV2\_PRENOM, ENFANTS2\_PRENOM et POLREL2. Dans son message du 7 mars 2020, Patrick Simon m’écrivait : « Vous demandez comment nous avons construits [sic] nos variables : la plupart des variables construites sont renseignées dans le dictionnaire des codes diffusé avec la base Quételet. » À l’aide du [questionnaire](#) et du [dictionnaire des codes](#) de l’enquête (Insee 2011) et des conseils d’utilisation et de codage qu’ils donnent,<sup>63</sup> j’ai identifié les variables nécessaires et suffisantes à la reproduction de l’article, soit principalement les variables suivantes : la pondération ; le sexe ; l’année de naissance ; la combinaison entre le lien à la migration (G1, G2, G3) et l’origine géographique (Europe du Sud, Maghreb, etc.) ; l’importance de la religion dans la vie ; et le prénom.<sup>64</sup> J’examine ci-dessous ces variables une à une pour vérifier si elles ne contiennent pas d’erreur — et si nécessaire je les recode, pour aboutir aux variables de la table REPRO qui sont utilisées dans la reproduction « au plus près ».

### Les 8 variables (dont 2 identifiants) issues de la table THL2\_PRENOM

À partir de la table THL2\_PRENOM, qui porte sur les 75 121 habitants du logement des enquêtés, la table REPRO reprend telles quelles 8 variables (Annexe 5B, p. 183) :

- les 2 variables permettant d’identifier chaque membre du logement : IDENT (identifiant de l’enquêté, pour chaque habitant du logement de l’enquêté), qui va de 1 à 21761 et compte 21 761 modalités, soit autant que d’enquêtés, ce qui est attendu ; et NOI (numéro d’ordre de l’individu dans le logement), qui va de 1 à 19

<sup>63</sup> Par exemple, « Aux variables de contrôle, il faut préférer les variables de nature "calculée", qui sont des variables qui ont été construites a posteriori à des fins de synthèse et d’analyse. Par exemple pour étudier le lien migratoire, les variables GROUP1 et GROUP2 ont été spécifiquement construites » (Insee 2011, p. 12). Par ailleurs, si l’enquêté est descendant de deux parents immigrés nés dans des lieux différents, on lui associe le lieu de naissance de son père, plutôt que celui de sa mère ou aucun des deux ([site TeO](#)).

<sup>64</sup> J’exclus ici les identifiants et d’autres variables méthodologiques.

et compte 19 modalités, c'est-à-dire que le logement enquêté le plus peuplé compte 19 habitants ; ces variables permettent d'apparier à la table THL2\_PRENOM les informations issues des autres tables ;

- POIDSI (pondération de l'habitant du logement) ; cette variable va de 13 à 59 700 ;
- SEXE (sexe de l'habitant du logement) ; cette variable a 2 modalités (Homme ; Femme), d'effectifs presque identiques ;
- ANAIS (année de naissance de l'habitant du logement) ; cette variable va de 1900 à 2009 ;
- REGIONNAIS2 (lieu de naissance de l'habitant du logement) ; cette variable compte 70 modalités (Insee 2011, p. 511-512) ;
- PRENOM (prénom de l'habitant du logement) ; les 75 121 habitants du logement de l'enquête portent 16 938 prénoms différents ; le prénom de chacun des habitants du logement (mais aussi des enfants non cohabitants de l'enquête) est demandé à l'enquête pour faciliter la passation de la partie du questionnaire relative au « tableau des habitants du logement » ; c'est donc le prénom d'usage (pas nécessairement identique au premier prénom de l'état civil), tel que déclaré par l'enquête ; ce prénom d'usage peut être simple ou composé (jusqu'à 4 « composants », pour un prénom asiatique), et il est renseigné en lettres capitales (sans accent ni apostrophe ni trait d'union) ;
- LIENEGO (lien avec l'enquête) ; cette variable va de 0 à 90 et compte 13 modalités (Enquête ; Conjoint ; Enfant ; Parent ; Frère, sœur ; Petit-enfant ; Grand-parent ; Beau-fils, belle-fille ; Beau-parent ; Autre lien familial ; Lien familial indéterminé ; Ami ; Autre lien non familial) ; cette variable permet de réaliser la reproduction « au plus près » sur plusieurs champs distincts.

#### **Les 41 variables (dont 2 identifiants) issues de la table INDIV2\_PRENOM**

À partir de la table INDIV2\_PRENOM, qui porte sur les 21 761 enquêtés, la table REPRO reprend 41 variables (Annexe 5B, p. 183) :

- les 2 variables permettant d'identifier l'enquête parmi les membres du logement : IDENT (identifiant de l'enquête, pour chaque enquête), qui va de 1 à 21 761 et compte 21 761 modalités différentes, soit autant que d'enquêtes, ce qui est attendu ; et NOI<sup>65</sup> (numéro d'ordre de l'enquête dans le logement), qui va de 1 à 9 et comprend 9 modalités ;
- POIDSI (pondération de l'enquête) ; cette variable va de 13 à 59 700 ;
- SEXEE (sexe de l'enquête) ; cette variable a 2 modalités (Homme ; Femme), d'effectifs proches ;
- ANAISE (année de naissance de l'enquête) ; cette variable va de 1948 à 1990, ce qui correspond au champ de l'enquête ;
- PRENOME (prénom de l'enquête) ; les 21 761 enquêtés portent 6 699 prénoms différents ;
- les 3 variables GROUP2 (statut de l'enquête par rapport à la migration), GROUP2CA (statut du conjoint actuel de l'enquête par rapport à la migration) et GROUP2PC (statut du premier conjoint de l'enquête par rapport à la migration) ; ces variables comptent respectivement 16, 15 et 15 modalités différentes ;<sup>66</sup>

<sup>65</sup> Cette variable n'est rien d'autre que la variable NE de la table INDIV2\_PRENOM, renommée NOI.

<sup>66</sup> L'enquête TeO collecte des informations relevant de « statistiques ethniques », comme le lieu de naissance de leurs parents et leur religion (autodéclarée) ; sur ce sujet, voir Le Minez 2020. Avant cette enquête et en l'absence de « statistiques ethniques », certains chercheurs devaient utiliser les prénoms

- les 9 variables de lieu de naissance : REGIONNAISE2 (lieu de naissance de l'enquêté), REGIONNAISP2 (lieu de naissance du père de l'enquêté ou de l'homme qui l'a élevé), REGIONNAISM2 (lieu de naissance de la mère de l'enquêté ou de la femme qui l'a élevé), REGIONNAISCA2 (lieu de naissance du conjoint actuel de l'enquêté), REGIONNAISCAP2 (lieu de naissance du père du conjoint actuel de l'enquêté), REGIONNAISCAM2 (lieu de naissance de la mère du conjoint actuel de l'enquêté), REGIONNAISPC2 (lieu de naissance du premier conjoint de l'enquêté), REGIONNAISPCP2 (lieu de naissance du père du premier conjoint de l'enquêté), REGIONNAISPCM2 (lieu de naissance de la mère du premier conjoint de l'enquêté) ; ces variables comptent chacune 70 modalités (Insee 2011, p. 511-512) ;
- les 2 variables de nationalité à la naissance des parents de l'enquêté : NATPNIV2 (nationalité à la naissance du père de l'enquêté ou de l'homme qui l'a élevé) et NATMNIV2 (nationalité à la naissance de la mère de l'enquêté ou de la femme qui l'a élevé) ; ces variables comptent chacune 65 modalités (Insee 2011, p. 504-505) ;
- les 2 variables de choix de nationalité des parents de l'enquêté nés dans une ancienne colonie française : N\_INDPER (choix de nationalité du père si né dans une ancienne colonie française) et N\_INDMER (choix de nationalité de la mère si née dans une ancienne colonie française) ; ces variables vont de 1 à 9 et comptent 5 modalités (A pris la nationalité du pays devenu indépendant ou la double nationalité ; Est resté français ; Était déjà décédé ; N'était pas encore né ; Ne sait pas) ;
- les 13 variables d'origine « dominante » de l'enquêté et de ses proches<sup>67</sup> : ORIG\_ENQ (origine de l'enquêté), ORIG\_PEREENQ (origine du père de l'enquêté), ORIG\_MEREENQ (origine de la mère de l'enquêté), ORIG\_FRERESOEURENQ (origine du frère ou de la sœur de l'enquêté), ORIG\_CA (origine du conjoint actuel de l'enquêté), ORIG\_PERECA (origine du père du conjoint actuel de l'enquêté), ORIG\_MERECA (origine de la mère du conjoint actuel de l'enquêté), ORIG\_ENFCA (origine de l'enfant du conjoint actuel de l'enquêté), ORIG\_PC (origine du premier conjoint de l'enquêté), ORIG\_PEREPC (origine du père du premier conjoint de

des individus comme indicateur de lien à la migration et d'origine géographique (Simon 2008, p. 160-163). Depuis cette enquête et la disponibilité de ces « statistiques ethniques », nul besoin d'utiliser les prénoms pour indiquer des variables directement renseignées, et on peut désormais utiliser les prénoms des individus comme indicateur d'assimilation intergénérationnelle (Coulmont et Simon 2019).

<sup>67</sup> Ces variables ont été construites dans la table INDIV2\_PRENOM à partir des variables suivantes de la même table : GROU1, GROU1CA, GROU1PC, GROU2, GROU2CA, GROU2PC, REGIONNAISE2, REGIONNAISP2, REGIONNAISM2, REGIONNAISCA2, REGIONNAISCAP2, REGIONNAISCAM2, REGIONNAISPC2, REGIONNAISPCP2 et REGIONNAISPCM2. Pour comprendre la logique de la construction des variables commençant par « ORIG\_ », précisons ici le codage de la variable ORIG\_ENQ (origine « dominante » de l'enquêté), en 8 étapes : (i) si l'enquêté est immigré, on lui associe son lieu de naissance ; (ii) si l'enquêté est natif d'un DOM, on lui associe le DOM ; (iii) si l'enquêté est descendant d'une mère immigrée uniquement, on lui associe le lieu de naissance de sa mère ; (iv) si l'enquêté est descendant d'un père immigré uniquement, on lui associe le lieu de naissance de son père ; (v) si l'enquêté est descendant de deux parents immigrés nés dans le même lieu, on lui associe le lieu de naissance de ses parents ; (vi) si l'enquêté est descendant de deux parents immigrés nés dans des lieux différents, on lui associe — conformément à la convention adoptée dans l'enquête TeO — le lieu de naissance de son père ([site TeO](#)) ; (vii) si l'enquêté est descendant d'un ou deux natifs d'un DOM, on lui associe le DOM ; (viii) si l'enquêté est membre du groupe « majoritaire », on lui associe ce groupe « majoritaire ». Au final cette variable ORIG\_ENQ (origine de l'enquêté) est bien renseignée pour les 21 761 enquêtés. Pour plus de précisions sur la construction des variables commençant par « ORIG\_ », voir Annexe 5B, p. 183.

l'enquêté), ORIG\_MEREPC (origine de la mère du premier conjoint de l'enquêté), ORIG\_GRDPARENTPATERNELENQ (origine des grands-parents paternels de l'enquêté) et ORIG\_GRDPARENTMATERNELENQ (origine des grands-parents maternels de l'enquêté) ; ces variables ont toutes le même codage à 3 chiffres, dans lequel le premier chiffre indique le lien de l'individu à la migration (Groupe dit « majoritaire » sans ascendant immigré sur deux générations ; Natif d'un DOM ; Enfant d'1 ou 2 natifs d'un DOM ; Petit-enfant d'1 à 4 natifs d'un DOM ; Immigré (G1) ; Enfant d'1 ou 2 immigrés (G2) ; Petit-enfant d'1 à 4 immigrés (G3) ; Arrière-petit-enfant d'1 à 8 immigrés (G4) ou natifs d'un DOM ; Ne sait pas), et les deux derniers chiffres indiquent le lieu de naissance de l'ascendant immigré de l'individu — s'il a un ascendant immigré — ou le lieu de naissance de l'individu s'il est lui-même immigré (France métropolitaine ; DOM ; Europe du Sud ; Autre Europe ; Maghreb ; Autre Afrique ; Turquie ; Asie (sauf Turquie) ; Ailleurs dans le monde ; Combinaison d'origines sans dominante ; Ne sait pas ; Non réponse) ;

- 4 variables de langues parlées par les parents de l'enquêté dans son enfance : LNP1\_GR (1<sup>e</sup> langue parlée par le père de l'enquêté dans son enfance), LNP2\_GR (2<sup>e</sup> langue parlée par le père de l'enquêté dans son enfance), LNM1\_GR (1<sup>e</sup> langue parlée par la mère de l'enquêté dans son enfance) et LNM2\_GR (2<sup>e</sup> langue parlée par la mère de l'enquêté dans son enfance) ; ces variables comprennent 28 modalités (Insee 2011, p. 501) ;
- R\_IMPVI (importance qu'accorde aujourd'hui l'enquêté à la religion dans sa vie) ; cette variable comprend 6 modalités (Pas du tout d'importance ; Un peu d'importance ; Assez d'importance ; Beaucoup d'importance ; Refuse de répondre ; Ne sait pas) ;
- X\_VUFRI (opinion de l'enquêté sur : « On me voit comme un Français ») ; cette variable comprend 6 modalités (Tout à fait d'accord ; Plutôt d'accord ; Plutôt pas d'accord ; Pas du tout d'accord ; Ne sait pas ; Refuse de répondre).

### **Les 11 variables (dont 2 identifiants) issues de la table ENFANTS2\_PRENOM**

À partir de la table ENFANTS2\_PRENOM, qui porte sur les 31 476 enfants des enquêtés (y compris leurs 7 924 enfants non cohabitants ou décédés), la table REPRO reprend 11 variables (Annexe 5B, p. 183) :

- les 2 variables permettant d'identifier les enfants de l'enquêté : IDENT (identifiant de l'enquêté, pour chaque enfant de l'enquêté), qui va de 3 à 21 760 et compte 13 388 modalités différentes, c'est-à-dire que 13 388 enquêtés (sur un total de 21 761) ont déclaré avoir eu au moins un enfant ; et NOI<sup>68</sup> (numéro d'ordre de l'enfant de l'enquêté dans le logement pour chaque enfant cohabitant de l'enquêté, et numéro identifiant arbitraire pour chaque enfant de l'enquêté non cohabitant ou décédé à la date de l'enquête) ;
- POIDSE (pondération de l'enfant de l'enquêté) ; cette variable va de 7 à 29 850 ;
- E\_SEXE (sexe de l'enfant de l'enquêté) ; cette variable a 3 modalités (Homme ; Femme ; Inconnu) ; les effectifs de garçons et de filles sont proches ;
- E\_ANAIS (année de naissance de l'enfant de l'enquêté) ; cette variable va de 1962 à 2009 et compte quelques non réponses ;

<sup>68</sup> Cette variable est la variable E\_NOI\_THL de la table ENFANTS2\_PRENOM, à laquelle on ajoute un NOI arbitraire pour les 7 924 enfants de l'enquêté non cohabitants ou décédés à la date de l'enquête, afin de pouvoir identifier ces enfants non cohabitants ou décédés à la date de l'enquête de la table ENFANTS2\_PRENOM, sans quoi on ne peut apparier les informations les concernant aux habitants du logement de l'enquêté de la table THL2\_PRENOM. Pour plus de précisions, voir Annexe 5B, p. 183.

- REGIONNAISENF2 (lieu de naissance de l'enfant de l'enquêté) ; cette variable compte 70 modalités (Insee 2011, p. 511-512) ;
- E\_PRENOM (prénom de l'enfant de l'enquêté) ; les 31 476 enfants des enquêtés portent 8 844 prénoms différents ;
- E\_VIE (est-ce que l'enfant de l'enquêté est toujours en vie) ; cette variable va de 1 à 9 et compte quelques non réponses, mais 98,9 % des 31 476 enfants des enquêtés sont toujours en vie à la date de l'enquête ;
- E\_LVIE (lieu où vit l'enfant) ; cette variable va de 1 à 9 et compte 334 non réponses, mais 96,4 % des 31 476 enfants des enquêtés vivent en métropole à la date de l'enquête ;
- ORIG\_ENFENQ (origine « dominante » de l'enfant de l'enquêté) ; cette variable a le même codage à 3 chiffres que les autres variables débutant par « ORIG\_ » (voir ci-dessus) ;
- AUTREPARENT (est-ce que l'enfant de l'enquêté a pour autre parent le conjoint actuel ou le premier conjoint de l'enquêté) ; cette variable a 2 modalités (Conjoint actuel ; Premier conjoint).

### **Les 6 variables (dont 1 identifiant) issues de la table POLREL2**

À partir de la table POLREL2, qui porte sur les 21 761 enquêtés, la table REPRO reprend 6 variables (Annexe 5B, p. 183) :

- IDENT (identifiant de l'enquêté, pour chaque enquêté), qui va de 1 à 21761 et compte 21 761 modalités différentes, soit autant que d'enquêtés, ce qui est attendu ;
- 5 variables d'affiliation religieuse : RELEGO1 (religion de l'enquêté), qui est renseignée par les réponses aux questions : « Aujourd'hui, avez-vous une religion ? [Si oui] Laquelle ? », RELCA1 (religion du conjoint actuel de l'enquêté), RELPC1 (religion du premier conjoint de l'enquêté), RELPERE1 (religion du père de l'enquêté) et RELMERE1 (religion de la mère de l'enquêté) ; ces variables sont recodées en 6 modalités (Sans religion ; Chrétien ; Musulman ; Juif ; Bouddhiste, hindouiste, autre, multi-religions ; Refus de répondre ou ne sait pas).

### **❖ La construction de la table REPRO**

On connaît maintenant la structure de la table REPRO : ses  $75\,121 + 7\,924 = 83\,045$  lignes, et ses  $(8-2) + (41-2) + (11-2) + (6-1) = 59$  colonnes, les identifiants étant exclus du décompte des colonnes. Ne reste plus qu'à rassembler ces informations dans une seule et même table. Pour cela, je procède par étapes (voir Annexe 5B, p. 183) :

- je pars des 75 121 habitants du logement de l'enquêté (lignes) et des  $(8-2 = 6)$  variables (colonnes) de la table THL2\_PRENOM (POIDSI, SEXE, ANAIS, REGIONNAIS2, PRENOM, LIENEG0) ;
- j'y ajoute les 7 924 enfants des enquêtés (lignes) de la table ENFANTS2\_PRENOM qui habitent un autre logement ou sont décédés à la date de l'enquête, aboutissant ainsi au total de 83 045 individus (lignes) ;
- j'y ajoute<sup>69</sup> les  $(11-2 = 9)$  variables (colonnes) issues de la table ENFANTS2\_PRENOM, appariées aux seuls enfants des enquêtés (POIDSE, E\_SEXE, E\_ANAIS, REGIONNAISENF2, E\_VIE, E\_LVIE, E\_PRENOM, ORIG\_ENFENQ et AUTREPARENT) ;

<sup>69</sup> Dans le programme informatique, cette commande et la précédente s'exécutent simultanément. Pour plus de précisions, voir Annexe 5B, p. 183.

- j'y ajoute ensuite les (41-2 = 39) variables (colonnes) issues de la table INDIV2\_PRENOM :
  - 6 variables appariées aux seuls enquêtés (POIDSI, SEXEE, ANAISE, PRENOME, R\_IMPVIIE et X\_VUFRI) ;
  - 33 variables appariées à tous les habitants du logement (REGIONNAISE2, REGIONNAISP2, REGIONNAISM2, REGIONNAISCA2, REGIONNAISCAP2, REGIONNAISCAM2, REGIONNAISPC2, REGIONNAISPCP2, REGIONNAISPCM2, GROUP2, GROUP2CA, GROUP2PC, ORIG\_ENQ, ORIG\_PEREENQ, ORIG\_MEREENQ, ORIG\_FRERESOEURENQ, ORIG\_CA, ORIG\_PERECAL, ORIG\_MERECAL, ORIG\_ENFCA, ORIG\_PC, ORIG\_PEREPC, ORIG\_MEREPC, ORIG\_GRDPARENTPATERNELENQ, ORIG\_GRDPARENTMATERNELENQ, NATPNIV2, NATMNIV2, N\_INDPER, N\_INDMER, LNP1\_GR, LNP2\_GR, LNM1\_GR et LNM2\_GR) ;
- j'y ajoute enfin les (6-1 = 5) variables issues de la table POLREL2, appariées à tous les habitants du logement (RELEGO1, RELCA1, RELPC1, RELPERE1 et RELMERE1) ;
- j'obtiens ainsi la table REPRO, qui comprend 83 045 individus (lignes) et 59 variables (colonnes) qui suffisent pour la reproduction « au plus près ».

#### ❖ Le recodage des variables de la table REPRO

Je combine et recode ces 59 variables de la table REPRO (et je leur applique des labels) pour aboutir aux 24 variables utilisées dans la reproduction « au plus près », présentées dans le tableau 5.1. Ces recodages appellent assez peu de commentaires dans la mesure où ils sont minimalistes et visent à reproduire les analyses de Coulmont et Simon (2019) « au plus près », sinon à l'identique. Toutes les précisions figurent dans l'Annexe 5B, p. 183.



**Tableau 5.1. Les 24 variables utilisées dans la reproduction « au plus près » (enquête TeO) <sup>70</sup>**

	Variable (table REPRO)		Variable(s) et table(s) d'origine ( <a href="#">CASD</a> )	Effectif non pondéré renseigné
	Nom (intitulé)	Modalités		
<b>1</b>	<b>POIDS_IND (pondération de l'individu)</b>	34 412 modalités (de 12 à 59 701)	POIDSI (THL2_PRENOM) et POIDSE (ENFANTS2_PRENOM)	83 045
<b>2</b>	<b>SEXE_IND (sexe de l'individu)</b>	3 modalités (Homme ; Femme ; Non réponse)	SEXE (THL2_PRENOM) et E_SEXE (ENFANTS2_PRENOM)	83 035
<b>3</b>	<b>ANAIIS_IND (année de naissance de l'individu)</b>	101 modalités (de 1900 à 2009)	ANAIIS (THL2_PRENOM) et E_ANAIIS (ENFANTS2_PRENOM)	82 961
<b>4</b>	<b>PRENOM_IND (prénom de l'individu)</b>	18 161 modalités	PRENOM (THL2_PRENOM) et E_PRENOM (ENFANTS2_PRENOM)	83 045
<b>5</b>	<b>LIENEGO_IND (lien de l'individu avec l'enquêté)</b>	15 modalités (Enquêté ; Conjoint ; Enfant cohabitant ; Enfant non cohabitant en vie ; Enfant non cohabitant décédé ; Parent ; Frère ou sœur ; Petit-enfant ; Grand-parent ; Beau-fils ou belle-fille ; Beau-parent ; Autre lien familial ; Lien familial indéterminé ; Ami ; Autre lien non familial)	LIENEGO (THL2_PRENOM) et E_VIE (ENFANTS2_PRENOM)	83 045
<b>6</b>	<b>REGIONNAIS_IND (lieu de naissance de l'individu)</b>	10 modalités (France métropolitaine ; DOM ; Maghreb ; Autre Afrique ; Asie ; Turquie ; Europe du Sud ; Autre Europe ; Ailleurs dans le monde ; Non réponse)	REGIONNAIS2 (THL2_PRENOM) et REGIONNAISENF2 (ENFANTS2_PRENOM)	83 027
<b>7</b>	<b>RESIDENCE_IND (lieu de résidence de l'individu)</b>	2 modalités (France métropolitaine ; Ailleurs)	E_LVIE (ENFANTS2_PRENOM)	83 045

<sup>70</sup> Ce tableau n'inclut pas les identifiants IDENT (issu des tables sources THL2\_PRENOM, INDIV2\_PRENOM, ENFANTS2\_PRENOM et POLREL2) et NOI (issu des tables sources THL2\_PRENOM, INDIV2\_PRENOM et ENFANTS2\_PRENOM), ni les variables de nomenclature des prénoms par origine ORIG\_PRENOM\_IND et CAT1\_INED\_REC (présentées et examinées en détail dans l'Annexe 6B, p. 285).

	Variable (table REPRO)		Variable(s) et table(s) d'origine ( <a href="#">CASD</a> )	Effectif non pondéré renseigné
	Nom (intitulé)	Modalités		
8	GROUP2_IND (lien à la migration de l'enquête et de son conjoint actuel)	18 modalités (Insee 2011, p. 241-243)	GROUP2 et GROUP2CA (INDIV2_PRENOM)	35 004
9-10	REGIONNAISPERE_IND et REGIONNAISMERE_IND (lieu de naissance du père / de la mère de l'individu)	70 modalités (Insee 2011, p. 511-512)	REGIONNAISE2, REGIONNAISP2, REGIONNAISM2, REGIONNAISCA2, REGIONNAISCAP2, REGIONNAISCAM2 et REGIONNAISPC2 (INDIV2_PRENOM)	67 403
11	ORIG_IND (origine « dominante » de l'individu)	33 modalités (voir tableau 5.3)	GROUP1, GROUP1CA, GROUP1PC, GROUP2, GROUP2CA, GROUP2PC, REGIONNAISE2, REGIONNAISP2, REGIONNAISM2, REGIONNAISCA2, REGIONNAISCAP2, REGIONNAISCAM2, REGIONNAISPC2, REGIONNAISPCP2 et REGIONNAISPCM2 (INDIV2_PRENOM) ; C_COHAC, C_COHAP et C_FINAP (ENFANTS2_PRENOM)	81 780
12	AUTREPARENT (l'autre parent de l'enfant de l'enquête est-il le conjoint cohabitant ou le conjoint précédent de l'enquête ?)	2 modalités	E_ANAIS, C_COHAC, C_COHAP et C_FINAP (ENFANTS2_PRENOM)	27 692
13-14	NATPNIV2 et NATMNIV2 (nationalité à la naissance du père / de la mère de l'enquête ou de l'homme / la femme qui l'a élevé)	65 modalités (Insee 2011, p. 504-505)	NATPNIV2 et NATMNIV2 (INDIV2_PRENOM)	21 761
15-16	N_INDPER et N_INDMER (choix de nationalité du père / de la mère si né(e) dans une ancienne colonie française)	5 modalités (A pris la nationalité du pays devenu indépendant ou la double nationalité ; Est resté français ; Était déjà décédé ; N'était pas encore né ; Ne sait pas)	N_INDPER et N_INDMER (INDIV2_PRENOM)	21 761
17-20	LNP1_GR, LNP2_GR, LNM1_GR et LNM2_GR (langues parlées par les parents de l'enquête dans son	28 modalités (Insee 2011, p. 501)	LNP1_GR, LNP2_GR, LNM1_GR et LNM2_GR (INDIV2_PRENOM)	21 761

	Variable (table REPRO)		Variable(s) et table(s) d'origine ( <a href="#">CASD</a> )	Effectif non pondéré renseigné
	Nom (intitulé)	Modalités		
	enfance)			
21	RELIG_IND (religion de l'individu)	6 modalités (Sans religion ; Chrétiens ; Musulmans ; Juifs ; Bouddhistes, hindouistes, autres, multi-religions ; Refus de répondre ou ne sait pas)	RELEGO1, RELCA1, RELPC1, RELPERE1 et RELMERE1 (POLREL2)	42 651
22	R_IMPVIE (religiosité de l'enquêté, i.e. importance qu'accorde aujourd'hui l'enquêté à la religion dans sa vie)	6 modalités (Pas du tout d'importance ; Un peu d'importance ; Assez d'importance ; Beaucoup d'importance ; Refus de répondre ; Ne sait pas)	R_IMPVIE (INDIV2_PRENOM)	15 239
23	ORIG_RELIG_IND (origine religieuse de l'individu)	36 modalités (combinaisons père-mère des 6 modalités de RELIG_IND)	RELEGO1, RELCA1, RELPC1, RELPERE1 et RELMERE1 (POLREL2)	59 082
24	X_VUFRI (opinion de l'enquêté sur : « On me voit comme un Français »)	6 modalités (Tout à fait d'accord ; Plutôt d'accord ; Plutôt pas d'accord ; Pas du tout d'accord ; Vous ne savez pas ; Vous ne souhaitez pas répondre)	X_VUFRI (INDIV2_PRENOM)	21 761

### ❖ L'échantillon retenu

Les auteurs m'ont indiqué avoir exploité « la base individus associée à la base enfants » (message de Patrick Simon du 7 mars 2020). Je suppose que par « individus » ils entendent les enquêtés seulement,<sup>71</sup> sans leurs conjoints. Mais j'ignore si par « enfants » ils entendent tous les enfants de l'enquêté ou un sous-échantillon d'entre eux. En outre, les programmes informatiques que les auteurs m'ont finalement communiqués (6 janvier 2021) incluent une exploitation de la base THL, et non seulement des bases INDIV et ENFANTS. Enfin, malgré de nombreuses tentatives, je ne suis jamais parvenu à identifier un échantillon qui permettrait de reproduire à l'identique les résultats des auteurs.

Par conséquent, je reproduis les analyses des auteurs sur **l'échantillon qui me semble s'approcher le plus d'un échantillon représentatif des habitants de 0-50 ans / 0-60 ans en France métropolitaine en 2008**. Cet échantillon, présenté dans le tableau 5.2, comprend :

- *tous* les enquêtés (N = 21 761),<sup>72</sup> qui vivent en ménage ordinaire en France métropolitaine en 2008 et qui sont représentatifs des personnes de 18-50 ans nées en France métropolitaine d'au moins un parent immigré ou né dans un DOM, et des personnes de 18-60 ans immigrées ou nées dans un DOM ou appartenant à la population « majoritaire » ;
- *certaines* des enfants de l'enquêté (N = 20 447), afin d'éviter les doubles comptes :
  - d'une part les enfants vivants *majeurs* de l'enquêté qui habitent en métropole et sont petits-enfants (G3) d'au moins un immigré ou petits-enfants (G3) d'au moins une personne née dans un DOM, car ils ne sont pas compris parmi les enquêtés (N = 1 081) ;
  - d'autre part les enfants vivants *mineurs* de l'enquêté qui habitent en métropole, car eux non plus ne sont pas compris parmi les enquêtés, à l'exception des enfants vivants mineurs de l'enquêté « majoritaire » qui appartiennent eux-mêmes à l'un des groupes « minoritaires » car ils sont déjà compris parmi les enfants mineurs des enquêtés des groupes « minoritaires » (N = 19 366) ;
- *aucun* autre membre du ménage de l'enquêté (N = 0), afin là encore d'éviter les doubles comptes.

Au total, **cet échantillon est représentatif des personnes de 18-50 / 18-60 ans en ménage ordinaire et de leurs enfants de 0-17 ans, habitant en France métropolitaine en 2008 (N = 42 208)**. Le champ ainsi délimité ne souffre pas de doubles comptes, alors que les autres champs — en ajoutant d'autres enfants des enquêtés et/ou d'autres membres du ménage des enquêtés qui sont *déjà* statistiquement représentés par les enquêtés et/ou certains de leurs

<sup>71</sup> La variable CHAMP1850 de la table INDIV2\_PRENOM permet de sélectionner, parmi les 21 761 enquêtés, les 18 864 enquêtés de 18-50 ans pour travailler sur le champ d'âge pour lequel l'enquête est représentative quel que soit le lien de l'enquêté à la migration (Insee 2011, p. 172). Mais comme rien ne semble indiquer que l'article de Coulmont et Simon (2019) porte sur ce champ restreint, j'analyse la totalité des enquêtés (N = 21 761).

<sup>72</sup> Pour rappel, les enquêtés TeO sont issus de 5 sous-échantillons de personnes d'âge actif vivant en ménage ordinaire en France métropolitaine en 2008 (Annexe 4A, p. 94) : (i) les immigrés, nés étrangers à l'étranger entre 1948 et 1990 (âgés de 18 à 60 ans en 2008) ; (ii) les personnes nées dans un département d'Outre-Mer (DOM) entre 1948 et 1990 (âgées de 18 à 60 ans) ; (iii) les enfants d'au moins un parent immigré, c'est-à-dire nés en France métropolitaine entre 1958 et 1990 d'au moins une personne née étrangère à l'étranger (âgés de 18 à 50 ans) ; (iv) les enfants d'au moins un parent né dans un DOM, c'est-à-dire nés en France métropolitaine entre 1958 et 1990 (âgés de 18 à 50 ans) ; (v) les autres personnes, nées entre 1948 et 1990 (âgées de 18 à 60 ans), n'appartenant à aucun des groupes précédents, et aussi appelées population « majoritaire » ; ce groupe inclut des personnes qui ont aussi un lien à la migration vers la France métropolitaine, comme les Français nés français à l'étranger et leurs enfants, les rapatriés de l'ex-empire colonial français et leurs enfants, les personnes originaires des COM et leurs enfants, et les personnes qui ont une ascendance immigrée lointaine (dont au moins un aïeul était immigré).

enfants — incluent des doublons et déforment donc la distribution des prénoms, par rapport à ce qu'elle était en métropole en 2008.

Dans les analyses du présent document, les résultats sont présentés sur le champ couvrant les personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008. D'autres champs sont explorés en fonction de la nature des analyses, y compris le champ maximal (N = 83 045) pour certaines analyses de nature méthodologique (tableaux 7.10 à 7.12).

**Tableau 5.2. L'échantillon de l'enquête TeO retenu pour la reproduction « au plus près »**

	Les 5 sous-échantillons d'enquêtés TeO en métropole en 2008				
	(i)	(ii)	(iii)	(iv)	(v)
	Immigré de 18-60 ans	Personne née dans un DOM de 18-60 ans	Enfant d'immigrés de 18-50 ans	Enfant de personnes nées dans un DOM de 18-50 ans	Personne de la population « majoritaire » de 18-60 ans
L'enfant d'enquête qui est vivant, <b>majeur</b> et habite en métropole en 2008 est-il inclus ?	Non, car l'enfant majeur d'au moins 1 parent immigré (G2) est déjà compris parmi les enquêtés (iii)	Non, car l'enfant majeur d'au moins 1 parent né dans un DOM (G2) est déjà compris parmi les enquêtés (iv)	Oui, car le petit-enfant majeur d'au moins 1 grand-parent immigré (G3) n'est pas compris parmi les enquêtés	Oui, car le petit-enfant majeur d'au moins 1 grand-parent né dans un DOM (G3) n'est pas compris parmi les enquêtés	Non, car l'enfant majeur de 2 « majoritaires » est déjà compris parmi les enquêtés (v), tout comme l'enfant majeur d'1 « majoritaire » et d'1 « minoritaire » (i, ii, iii et iv)
L'enfant d'enquête qui est vivant, <b>mineur</b> et habite en métropole en 2008 est-il inclus ?	Oui, car aucun de ces mineurs n'est déjà compris parmi les enquêtés				Oui pour l'enfant mineur de 2 parents « majoritaires » (et donc lui-même « majoritaire »), car ces mineurs ne sont pas déjà compris parmi les enquêtés ou leurs enfants. Mais <i>non</i> pour l'enfant mineur de l'enquête « majoritaire » et d'un autre parent « minoritaire » (et donc lui-même « minoritaire »), car ce mineur (G2 ou G3) est déjà compris parmi les enfants mineurs des enquêtés « minoritaires »

NB : pour qu'un enquêté fasse partie de la population « majoritaire », aucun de ses 2 parents ne doit être né étranger à l'étranger (sans quoi il serait enfant d'immigré [G2]) ni né dans un DOM (sans quoi il serait enfant de DOMien [G2]), et il ne doit pas lui-même être né étranger à l'étranger (sans quoi il serait immigré [G1]) ni né dans un DOM (sans quoi il serait DOMien [G1]). Et pour qu'un enfant d'enquête fasse partie de la population « majoritaire », aucun de ses 2 parents ni de ses 4 grands-parents ne doit être né étranger à l'étranger (sans quoi il serait enfant [G2] ou petit-enfant [G3] d'immigrés) ni né dans un DOM (sans quoi il serait enfant [G2] ou petit-enfant [G3] de DOMiens).

### ❖ La pondération des individus

L'enquête TeO repose sur un échantillonnage complexe. « Le suréchantillon de l'enquête a surreprésenté les descendants d'immigrés, de rapatriés, d'expatriés ou d'un parent né dans un DOM nés après 1958 » (Insee 2011, p. 6). « Pour les échantillons d'immigrés et de descendants d'immigrés, certaines origines sont surreprésentées : Turquie, Afrique subsaharienne et Asie du Sud-est (Vietnam, Laos et Cambodge) » (Insee 2011, p. 7). Les statistiques descriptives de l'enquête doivent donc être calculées sur un échantillon *pondéré*, afin de tenir compte du suréchantillonnage des descendants d'immigrés (qui exige de les sous-pondérer), mais aussi afin de tenir compte des variations de taux de non-réponse d'une population à l'autre.

Les auteurs m'ont indiqué qu'« il n'y a qu'une pondération à utiliser pour TeO qui s'appelle poidsi » (message de Patrick Simon du 7 mars 2020). Et de fait, les 3 programmes R qu'ils m'ont finalement communiqués (6 janvier 2021) utilisent exclusivement cette pondération POIDSI. Pourtant, il est nécessaire d'utiliser la pondération POIDSE pour pondérer les enfants des enquêtés (Insee 2011, p. 9). Concrètement, POIDSE vaut la moitié de POIDSI, car un enfant d'enquêté a deux fois plus de chances qu'un enquêté d'être échantillonné : par son père, et par sa mère.

La pondération est donc calculée comme suit : les enfants de l'enquêté sont pondérés avec POIDSE (comme dans la base ENFANTS2\_PRENOM dont ils proviennent), et les enquêtés sont pondérés avec POIDSI (comme dans la base INDIV2\_PRENOM dont ils proviennent). Les résultats présentés sont pondérés, sauf certains résultats de nature méthodologique (tableaux 7.10 à 7.12). Notons que lorsque le logiciel Stata pondère des individus statistiques, il rapporte leur effectif total au nombre de lignes analysé, si bien que la pondération ne modifie pas l'effectif total mais seulement la distribution des modalités des variables. Tous les résultats sont présentés arrondis à l'unité près, sauf certains résultats de nature méthodologique (tableau 5.3).

### ❖ Vérification de la validité « interne » de la table REPRO (par rapport à l'enquête TeO)

Pour m'assurer que la table REPRO n'est pas entachée d'erreurs, je procède à plusieurs vérifications :

- la table REPRO contient bien 83 045 individus en lignes et les 24 variables (+ 2 identifiants) en colonnes ;
- les variables de vérification sont cohérentes entre elles, ce qui indique que l'appariement entre bases s'est déroulé comme prévu : pour tout individu les valeurs des variables de sexe sont identiques qu'elles proviennent des tables THL2\_PRENOM (SEXE), INDIV2\_PRENOM (SEXEE) ou ENFANTS2\_PRENOM (E\_SEXE), et il en va de même des valeurs des variables de pondération (POIDSI, POIDSI et POIDSE), d'année de naissance (ANAI, ANAISE et E\_ANAI), de prénom (PRENOM, PRENOME et E\_PRENOM) ou encore de lieu de naissance (REGIONNAIS2 et REGIONNAISENF2) ;
- les variables renseignent le bon nombre d'individus du bon type, ce qui indique que l'appariement entre tables mais aussi les recodages se sont déroulés comme prévu : les variables calculables pour tous les individus ou presque sont bien renseignées pour jusqu'à 83 045 individus (POIDS\_IND, SEXE\_IND, ANAI\_IND, PRENOM\_IND, LIENEGO\_IND, REGIONNAIS\_IND, RESIDENCE\_IND, ORIG\_IND) ; les variables calculables pour les seuls enquêtés sont bien renseignées pour jusqu'à 21 761 enquêtés (NATPNIV2, NATMNIV2, N\_INDPER, N\_INDMER, LNP1\_GR, LNP2\_GR, LNM1\_GR, LNM2\_GR, R\_IMPVIE, X\_VUFRI) ; les variables calculables pour les enquêtés



et seulement certains de leurs apparentés sont bien renseignées pour entre 21 761 et 83 045 individus, de 35 004 pour celle renseignées pour les enquêtés et leurs conjoints cohabitants (GROUP2\_IND) jusqu'à 67 403 pour celles renseignées pour la plupart des apparentés des enquêtés (REGIONNAISPERE\_IND, REGIONNAISMERE\_IND), en passant par 42 651 pour celle renseignée pour les enquêtés et leurs conjoints cohabitants ainsi que les parents des enquêtés (RELIG\_IND) et 59 082 pour celle renseignée pour les enquêtés et leurs enfants ainsi que les frères et sœurs des enquêtés (ORIG\_RELIG\_IND) ; et la variable calculable pour les seuls enfants des enquêtés (AUTREPARENT) est bien renseignée pour moins de 31 476 individus ;

- la table REPRO permet de reproduire à l'identique une trentaine de tris à plat des variables sur chacune des tables appariées (THL2\_PRENOM, INDIV2\_PRENOM et ENFANTS2\_PRENOM), ce qui indique que ni les appariements ni les recodages des variables n'ont introduit d'erreur : par exemple un tri à plat de la variable sexe (SEXE\_IND) permet de retrouver, à l'unité d'effectif non pondéré près et à la première décimale de pourcentage d'effectif pondéré près, les tris à plat — présentés dans le dictionnaire des codes — de la variable SEXE sur le champ de la table THL2\_PRENOM, mais aussi de SEXEE sur le champ de la table INDIV2\_PRENOM, et enfin le tri de E\_SEXE sur le champ de la table ENFANTS2\_PRENOM (Insee 2011, p. 444, 445 et 209) ; il en va de même pour le tri à plat de la variable d'année de naissance (ANAIIS\_IND permet de reproduire les tris de ANAIIS dans THL2\_PRENOM, de ANAISE dans INDIV2\_PRENOM et de E\_ANAIIS dans ENFANTS2\_PRENOM) (Insee 2011, p. 93, 93 et 205) ; il en va de même pour la variable LIENEGO de la table THL2\_PRENOM (Insee 2011, p. 291), pour les variables GROUP2 et GROUP2CA de la table INDIV2\_PRENOM (Insee 2011, p. 241-242), pour les variables N\_INDPER et N\_INDMER de la table INDIV2\_PRENOM (Insee 2011, p. 316), pour la variable R\_IMPVIE de la table INDIV2\_PRENOM (Insee 2011, p. 419) et pour la variable X\_VUFRI de la table INDIV2\_PRENOM (Insee 2011, p. 491) ; ces vérifications permettent au passage de découvrir certaines erreurs mineures dans le dictionnaire des codes de l'enquête TeO<sup>73</sup> ;
- les seules modalités de réponse imprévues concernant certains prénoms (« 0 », « 000000 », « 1 », « 2 », « 3 », « 4 », « 5 », « 6 », « 9999999999999999 », « ? », « A », « AAA », etc.) ; au moins environ 150 individus (sur 83 045) ont un prénom mal renseigné, que je conserve tel quel ;
- quant à la variable d'origine des individus (ORIG\_IND), créée pour les besoins de la reproduction « au plus près » et donc absente du dictionnaire des codes,<sup>74</sup>
  - elle permet de retrouver le tri à plat de la variable GROUP1 sur le champ enquêtés (Insee 2011, p. 239) ; et toujours sur le champ des enquêtés, elle permet de retrouver la taille de l'échantillon de la plupart des origines, telle que présentée dans un document méthodologique relatif à l'enquête TeO (Algava et Lhommeau 2013, p. 32)<sup>75</sup> ;

<sup>73</sup> Par exemple, la moyenne pondérée de la variable ANAISE (Insee 2011, p. 93) est en réalité la moyenne non pondérée.

<sup>74</sup> La plupart des autres variables dont le tri à plat est absent du dictionnaire des codes sont des variables d'accès restreint : d'une part les variables linguistiques LNP1\_GR, LNP2\_GR, LNM1\_GR et LNM2\_GR de la table INDIV2\_PRENOM ; d'autre part les variables religieuses RELEGO1, RELCA1, RELPC1, RELPERE1 et RELMERE1 de la table POLREL2.

<sup>75</sup> C'est le cas pour plusieurs origines (groupe « majoritaire », natifs de DOM, total d'immigrés, immigrés de Turquie) mais pas pour toutes, certaines comprenant des effectifs légèrement différents entre le codage d'Algava et Lhommeau (2013) et le mien (descendants de natifs de DOM, total descendants d'immigrés, descendants

- la distribution de cette variable d'origine sur le champ maximal se croise avec le lien à l'enquêté de façon logique : parmi les immigrés et les descendants d'immigrés on observe que l'enquêté, son conjoint et le frère et la sœur de l'enquêté ne sont que G1 ou G2, jamais G3 ou G4 ; l'enfant de l'enquêté (cohabitant ou non, y compris décédé) et son bel-enfant ne sont que G2 ou G3, jamais G1 ou G4 ; le parent, le grand-parent et le beau-parent de l'enquêté ne sont que G1 (ou G0), jamais G2, G3 ou G4 ; le petit-enfant de l'enquêté n'est que G2, G3 ou G4, jamais G1 ; et les autres liens familiaux et non familiaux de l'enquêté sont tous Non réponse ;
- la distribution de cette variable d'origine sur le champ maximal ne semble pas présenter de caractéristique invraisemblable ; en outre, la pondération accroît fortement la part d'individus qui sont membres du groupe « majoritaire » (de 15 % à 75 %) et elle réduit la part d'individus qui sont « G1 Maghreb » (de 5 % à 2 %), « G2 Maghreb » (de 14 % à 4 %), « G3 Maghreb » (de 4 % à 1 %), « G1 Europe du Sud » (de 3 % à 1 %), « G2 Europe du Sud » (de 9 % à 4 %) ou encore « G3 Europe du Sud » (de 6 % à 2 %), etc., ce qui était attendu puisque l'enquête TeO a délibérément suréchantillonné — et donc, sous-pondère — les enfants d'immigrés.

Au total, la table REPRO ne semble pas contenir d'erreur.

**Tableau 5.3. Distribution de la variable d'origine des individus (ORIG\_IND)  
(enquête TeO)**

	<b>Label</b>	<b>Effectif non pondéré</b>	<b>% pondéré</b>
120	G1 Maghreb	2 083	2,47 %
129	G1 autre Afrique	1 644	1,09 %
130	G1 Asie	1 121	0,72 %
134	G1 Turquie	830	0,45 %
140	G1 Europe du Sud	1 332	1,34 %
149	G1 autre Europe	1 027	1,06 %
150	G1 ailleurs	419	0,50 %
212	G1 DOM	712	0,62 %
320	G2 Maghreb	5 904	4,21 %
329	G2 autre Afrique	3 555	1,25 %
330	G2 Asie	1 862	0,60 %
334	G2 Turquie	1 893	0,52 %
340	G2 Europe du Sud	3 805	3,55 %
349	G2 autre Europe	1 810	1,80 %
350	G2 ailleurs	595	0,41 %
412	G2 DOM	1 181	0,69 %
500	Grand-parent « majoritaire » de l'enquêté		
501	Parent « majoritaire » de l'enquêté		
502	Enquêté « majoritaire »	3 781	57,45 %
503	Enfant « majoritaire » de l'enquêté	2 656	17,83 %
504	Petit-enfant « majoritaire » de l'enquêté		
620	G3 Maghreb	1 490	0,79 %
629	G3 autre Afrique	280	0,08 %
630	G3 Asie	153	0,04 %
634	G3 Turquie	71	0,01 %
640	G3 Europe du Sud	2 626	1,52 %
649	G3 autre Europe	900	0,79 %
650	G3 ailleurs	59	0,05 %
712	G3 DOM	418	0,15 %
820	G4 Maghreb <sup>76</sup>		
840	G4 Europe du Sud <sup>77</sup>		
849	G4 autre Europe <sup>78</sup>		
.	Non réponse	1	0,00 %
	<b>Total</b>	<b>42 208</b>	<b>100 %</b>

Champ : personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008, N = 42 208.

Source : TeO (voir Annexe 5B, p. 183).

### ❖ Analyses préliminaires sur la table REPRO

Examinons maintenant quelques tris à plat et tris croisés des variables de la table REPRO sur l'échantillon pondéré retenu pour la reproduction « au plus près ».

Cet échantillon, représentatif des personnes de 18-50 / 18-60 ans en ménage ordinaire et de leurs enfants de 0-17 ans habitant en France métropolitaine en 2008, comprend à peine moins

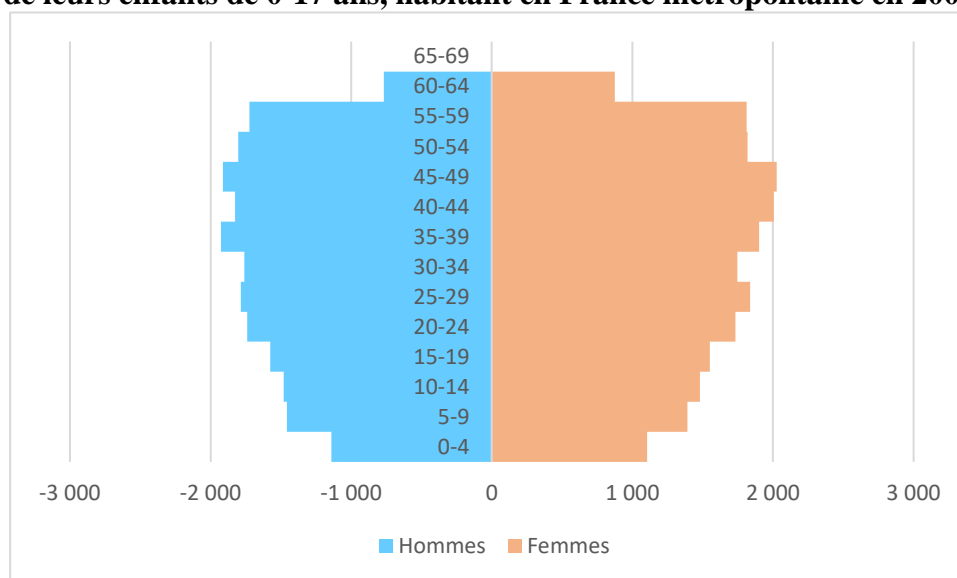
<sup>76</sup> Arrière-petit-enfant d'immigré du Maghreb.

<sup>77</sup> Arrière-petit-enfant d'immigré d'Europe du Sud.

<sup>78</sup> Arrière-petit-enfant d'immigré d'autre Europe.

de 50 % d'hommes et à peine plus de 50 % de femmes (SEXE\_IND), nés de 1948 jusqu'à l'enquête (ANAI\_IND) ; leur pyramide des âges a la forme ci-dessous (figure 5.4). Concernant l'origine (ORIG\_IND) des individus de cet échantillon, 75 % sont du groupe « majoritaire » ; 1 % sont immigrés (G1) et 4 % enfants d'immigrés (G2) et 2 % petits-enfants d'immigrés (G3) d'Europe du Sud ; 2 % sont immigrés (G1) et 4 % enfants d'immigrés (G2) et 1 % petits-enfants d'immigrés (G3) du Maghreb ; etc. (tableau 5.3).

**Figure 5.4. Pyramide des âges des personnes de 18-50 / 18-60 ans en ménage ordinaire et de leurs enfants de 0-17 ans, habitant en France métropolitaine en 2008**



Champ : personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008, N = 42 208.

Source : TeO (voir Annexe 5B, p. 183).

Concernant l'affiliation religieuse (RELIG\_IND) et la religiosité (R\_IMPVE) des enquêtés (représentatifs des personnes de 18-50 / 18-60 ans vivant en ménage ordinaire en France métropolitaine en 2008), 41 % se déclarent sans religion, 49 % se déclarent chrétiens (dont 11 % déclarent accorder « beaucoup » d'importance à la religion dans leur vie), 7 % se déclarent musulmans (dont 50 % déclarent accorder « beaucoup » d'importance à la religion dans leur vie), moins de 0,5 % se déclarent juifs et 1 % se déclarent bouddhistes, hindouistes, autres ou multi-religions. Parmi ces enquêtés, les chrétiens sont en moyenne les plus âgés à l'enquête (42 ans), tandis que les musulmans sont les plus jeunes (35 ans). Ces faits sont déjà bien connus, mais le fait que la table REPRO permette de les retrouver est rassurant.

#### ❖ Analyse préliminaire des prénoms en métropole en 2008

Les 42 208 personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008 portent, avant le nettoyage des prénoms,<sup>79</sup> 11 151 prénoms différents (PRENOM\_IND).<sup>80</sup> Les hommes portent 5 446 prénoms différents, les

<sup>79</sup> Les prénoms de l'enquête TeO font l'objet d'un nettoyage destiné à harmoniser leur format : je supprime diverses fautes de frappe (y compris les chiffres de 0 à 9) ainsi que les accents, trémas et cédilles, je remplace les apostrophes et traits d'union par des espaces, je supprime les doubles espaces et les espaces initiaux et finaux, etc. (Annexe 6B, p. 285). Les prénoms du Fichier INSEE des prénoms ainsi que ceux des nomenclatures font l'objet du même nettoyage harmonisé.

<sup>80</sup> Par contraste, les individus de l'échantillon maximal (N = 83 045) portent 18 161 prénoms différents, un chiffre très proche des « 18 163 » prénoms différents mentionnés par Baptiste Coulmont lors de notre discussion orale du 17 avril 2019 et dans son message du 14 mai 2019 (Annexe 4A, p. 94). Mais cet échantillon maximal ne peut être utilisé tel quel (sauf pour des analyses méthodologiques), puisqu'il comprend de nombreux doubles comptes et

femmes 5 976. Le prénom dont il s'agit n'est que le premier prénom (y compris les prénoms composés), et non pas le deuxième ou troisième prénom. Précisément il s'agit probablement surtout du **premier prénom d'usage** (tel que rapporté par l'enquêté), qui peut éventuellement différer du premier prénom enregistré à l'état civil.

Mais l'échantillon de prénoms de l'enquête TeO est-il fiable, c'est-à-dire produit-il des résultats globalement cohérents avec ceux des données exhaustives de l'état civil ? Pour le savoir, je compare les prénoms portés par les individus de l'échantillon retenu pondéré *qui sont nés en France entière* avec les prénoms attribués en France entière à l'état civil, tels que présentés dans le fichier INSEE des prénoms.

Rappelons au préalable que ces deux sources diffèrent fortement :

- l'enquête TeO est une enquête transversale : elle comporte un *échantillon* d'un *stock* d'individus à une date donnée, et elle décrit la fréquence à laquelle chaque prénom est *porté* en France métropolitaine en 2008 ;
- le fichier INSEE des prénoms, issu de l'état civil, est une source administrative : il comporte la *totalité* des *flux* de naissances entre les deux 1<sup>er</sup> janvier encadrant les années civiles, et il décrit la fréquence à laquelle chaque prénom a été *attribué* en France entière (hors Mayotte) depuis le 1<sup>er</sup> janvier 1900.

On ne saurait s'attendre à ce que des sources aussi différentes produisent des résultats rigoureusement identiques, mais on peut malgré tout s'attendre à ce que l'enquête TeO reproduise globalement les grands faits stylisés issus du fichier INSEE des prénoms, surtout concernant les naissances relativement récentes à la date de l'enquête TeO, encore peu affectées par la mortalité différentielle selon la classe sociale et donc selon le prénom. Est-ce bien le cas ? C'est ce que j'examine ci-dessous pour le Top 3 des prénoms par sexe et par période de naissance, puis pour la distribution complète des prénoms attribués en France. Ces analyses permettent au passage de se faire une idée globale de l'évolution des prénoms attribués en France au cours des dernières décennies, mais aussi au cours des derniers siècles.

### ➤ Évolution du Top 3 des prénoms au fil des décennies

L'échantillon retenu pondéré de l'enquête TeO reflète-t-il bien l'évolution du Top 3 des prénoms les plus attribués en France au fil des ans ? Comme l'indique le tableau 5.5, le Top 3 des prénoms TeO reflète plutôt bien le Top 3 des prénoms du fichier INSEE. Pour chaque décennie, sur les Top 3 des deux sexes, de 2 à 6 prénoms du fichier INSEE (sur 6) figurent dans le Top 3 de l'enquête TeO, et concordent donc bien entre les deux sources. Des « Michel », « Martine » et « Françoise » des années 1950 aux « Lucas », « Enzo », « Manon » et « Léa » des années 2000, en passant par les « Stéphane », « Sandrine », « Nathalie » et « Stéphanie » des années 1970 et par les « Sébastien », « Julien », « Nicolas », « Élodie », « Émilie » et « Aurélie » des années 1980, les prénoms les plus portés par l'échantillon retenu pondéré de l'enquête TeO ressemblent bien à ceux qui ont été les plus attribués à l'état civil lors de leur décennie de naissance. Autrement dit, le « stock » des prénoms des individus TeO en 2008 reflète plutôt bien le cumul des « flux » des prénoms attribués à l'état-civil de 1950 à 2009.

---

n'est donc représentatif d'aucune population habitant en France métropolitaine en 2008.

**Tableau 5.5. Top 3 des prénoms par sexe et par décennie de naissance, selon l'enquête TeO (2008) et selon le fichier INSEE des prénoms (1900-2019)**

	Prénoms portés en 2008 (enquête TeO)		Prénoms attribués depuis 1900 (fichier INSEE des prénoms)		Concordance	
	Homme	Femme	Homme	Femme	Homme	Femme
<b>1900-1909</b>	- - -	- - -	Jean Louis Pierre	Marie Jeanne Marguerite		
<b>1910-1919</b>	- - -	- - -	Jean André Pierre	Marie Jeanne Madeleine		
<b>1920-1929</b>	- - -	- - -	Jean André Pierre	Marie Jeanne Jacqueline		
<b>1930-1939</b>	- - -	- - -	Jean Michel André	Marie Monique Jacqueline		
<b>1940-1949</b>	- - -	- - -	Jean Michel Daniel	Marie Monique Nicole		
<b>1950-1959</b>	Patrick Michel Christian	Martine Françoise Catherine	Jean Michel Alain	Marie Martine Françoise	1/3	2/3
<b>1960-1969</b>	Philippe Éric Thierry	Catherine Isabelle Sylvie	Philippe Éric Pascal	Sylvie Nathalie Isabelle	2/3	2/3
<b>1970-1979</b>	Stéphane Frédéric Sébastien	Sandrine Nathalie Stéphanie	Christophe David Stéphane	Sandrine Nathalie Stéphanie	1/3	3/3
<b>1980-1989</b>	Sébastien Julien Nicolas	Élodie Émilie Aurélie	Nicolas Julien Sébastien	Aurélie Émilie Élodie	3/3	3/3
<b>1990-1999</b>	Nicolas Julien Thomas	Élodie Manon Laura	Thomas Kevin Alexandre	Laura Marie Camille	1/3	1/3
<b>2000-2009</b>	Lucas Enzo Tom	Manon Léa Clara	Lucas Enzo Thomas	Léa Emma Manon	2/3	2/3
<b>2010-2019</b>	- - -	- - -	Gabriel Lucas Léo	Emma Jade Chloé		

Champ enquête TeO : individus du champ retenu (personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008) qui sont nés en France entière des années 1950 aux années 2000, données pondérées, N = 32 022.

Champ [fichier INSEE des prénoms](#) : naissances en France entière (hors Mayotte) de 1900 à 2019, N = 85 047 407. Précisions méthodologiques [ici](#).

### ➤ Évolution de la taille du stock de prénoms attribués

L'enquête TeO reflète-t-elle aussi le fait que « le stock des prénoms couramment attribués s'élargit » (Desplanques 1986, p. 63) ? Des années 1950 aux années 2010, le nombre de prénoms différents attribués au moins une fois à chaque sexe à l'état civil est passé d'environ 1 500 à environ 6 500 par an (figure 5.6b).<sup>81</sup> L'enquête TeO reflète effectivement cet essor de la taille du stock de prénoms attribués, puisque dans l'échantillon retenu pondéré, des années 1950 aux années 2000, le nombre de prénoms différents attribués est passé d'environ 200 à environ 2 000 par décennie (figure 5.6a). Si l'échantillon de l'enquête TeO ici utilisé comprend environ 2 500 fois moins d'individus que la population exhaustive de l'état civil (32 022 contre 85 047 407), et si cela réduit la diversité des prénoms attribués (comme en témoigne la différence entre les ordonnées des figures 5.6a et 5.6b), l'enquête TeO n'en reflète pas moins le mouvement général de hausse de la taille du stock de prénoms attribués en France.

Rappelons le contexte historique de ce mouvement (Annexe 6A, p. 267). En Europe du 12<sup>e</sup> au 18<sup>e</sup> siècles, le nombre de prénoms différents attribués (et en particulier le nombre de prénoms féminins) a chuté, si bien que quelques prénoms en sont venus à suffire pour prénommer la plupart des habitants (Bozon 1987, p. 84-85 ; Leibring 2016, p. 212). En effet la transmission du prénom du grand-parent et du parrain (ou de la marraine) au nouveau-né a réduit le nombre de prénoms différents (Leibring 2016, p. 206). En outre, suite au concile de Trente (1545-1563), le prêtre devait s'assurer que le nouveau-né baptisé portait, dans les registres paroissiaux, le prénom d'un saint (du calendrier) de l'Église catholique romaine, plutôt que, par exemple, un prénom de l'Ancien Testament, selon la pratique alors émergente chez certains protestants (Zgusta 2014). Et dans l'Europe protestante, la plupart des prénoms « païens » (germaniques, y compris vieux-norrois) ont cessé d'être attribués (Leibring 2016, p. 205). Cette réduction du nombre de prénoms et sa concentration sur les prénoms des saints chrétiens a contribué, à partir des environs du 12<sup>e</sup> siècle, à l'attribution en France et ailleurs de deux voire trois prénoms, dans le but de distinguer les individus porteurs du même nom (Leibring 2016, p. 207-208 ; Desplanques 1986, p. 67 et 70 ; pour diverses monographies sur la France du 17<sup>e</sup> au 19<sup>e</sup> siècles, voir Dupâquier et al. 1984).

Puis en Europe à partir des 18<sup>e</sup> et 19<sup>e</sup> siècles, le nombre de prénoms différents attribués a de nouveau crû. C'est le cas dès le 18<sup>e</sup> siècle avec l'introduction de prénoms issus de la mythologie gréco-romaine et de prénoms issus d'œuvres littéraires : en France « Julius » donne « Jules », « Augustus » « Auguste », « Diana » « Diane » et « Flora » « Flore », tandis que se diffusent les prénoms rousseauistes « Julie » et « Héloïse » (Leibring 2016, p. 208-209). Puis au 19<sup>e</sup> siècle le nombre de prénoms augmente par plusieurs biais : par masculinisation de prénoms féminins et surtout par féminisation de prénoms masculins ; par suffixation (« Marie » donne « Maria » mais aussi « Marianne », « Mariette », « Marielle », « Marine » et « Maryse ») ; par simplification (« Élisabeth » donne « Élise ») ; et par l'essor des prénoms composés (Bozon 1987, p. 89-91).<sup>82</sup> Cette multiplication des prénoms attribués est généralement considérée comme un indice du processus d'individualisation à l'œuvre en France depuis le 18<sup>e</sup> siècle, et qui connaît une accélération depuis le milieu du 20<sup>e</sup> siècle : de plus en plus de parents choisissent pour leurs enfants des prénoms relativement originaux, distinctifs et individualisants, qui leur permettent d'apparaître comme unique ou de « sortir du lot » (Twenge et al. 2010 ; Lawson 2016, p. 189). On a pu aussi évoquer l'essor d'un « narcissisme de masse » (Fourquet et Manternach 2019a, p. 60 et 90). Autre indice

---

<sup>81</sup> En réalité, même le Fichier INSEE des prénoms sous-estime le nombre de prénoms différents attribués chaque année à l'état civil, puisqu'il regroupe les prénoms dits « rares » (c'est-à-dire attribués moins de 3 fois par an ou moins de 20 fois par demi-siècle) en une seule catégorie, comme s'ils constituaient ensemble un seul prénom.

<sup>82</sup> De même à Mexico (Mexique) entre le milieu du 16<sup>e</sup> et le milieu du 20<sup>e</sup> siècles, la variété des prénoms attribués a augmenté, ajoutant aux prénoms du Nouveau Testament des prénoms hébraïques, germaniques et païens (mais pas aztèques) (Boyd-Bowman 1970, p. 12-13, 31 et 35).



d'individualisation lié à notre sujet : dans la littérature française du 17<sup>e</sup> au 20<sup>e</sup> siècles, la fréquence d'occurrence de prénoms (chrétiens) ne cesse de croître (Brunet 1988, p. 124, fondé sur la base de données Frantext).

L'enquête TeO reflète aussi le fait que, parmi les prénoms attribués en France au 20<sup>e</sup> siècle, il existe plus de prénoms féminins que masculins (figures 5.6a et 5.6b). « Alors qu'aux âges classiques [16<sup>e</sup>-18<sup>e</sup> siècles], les prénoms féminins avaient toujours été moins nombreux, voilà qu'ils se multiplient à partir du XIX<sup>e</sup> siècle et que le stock masculin est dépassé par le stock féminin » (Bozon 1987, p. 90 ; voir aussi Desplanques 1986, p. 65-66 ; sur le cas du Mexique aux 16<sup>e</sup> et 17<sup>e</sup> siècles, voir Boyd-Bowman 1970, p. 17). Non seulement plus de prénoms sont alors féminisés que masculinisés, mais les diminutifs et hypocoristiques féminins deviennent plus souvent des prénoms que les diminutifs masculins. « À partir du XIX<sup>e</sup> siècle [...], le nombre de prénoms féminins en circulation dépasse le nombre de prénoms masculins en circulation : si *Jeannette* devient un prénom à part entière, *Jeannot* reste un surnom (et n'est pas déclaré à l'état civil) » (Coulmont 2014, p. 34).

Dans la France contemporaine aussi, les parents tendent à attribuer relativement plus de prénoms rares, originaux, nouveaux ou étrangers à leurs filles qu'à leurs fils, qui portent plus souvent des prénoms traditionnels et moins variés. Ce fait n'est d'ailleurs pas spécifique à la France : on l'observe aussi en Allemagne à la fin du 20<sup>e</sup> et au début du 21<sup>e</sup> siècles (Gerhards et Hans 2009, p. 1109-1110 ; voir aussi Gerritzen 2006, p. 180), en Espagne en 2004 (Mateos et Tucker 2008), chez la minorité russophone de Finlande dans les années 2000 et 2010 (Eskola et Hämäläinen 2019, p. 203) ou encore chez les immigrés pakistanais musulmans au Royaume-Uni (Bramwell 2011, p. 36). Mais c'est aussi ce que l'on observe ailleurs qu'en Occident, que ce soit en Turquie de 1950 à 2010 (Sabuncu et Coulmont 2011), en Iran des années 1960 aux années 1980 (60 % des garçons se voient attribuer un prénom islamique contre seulement 25 % des filles) (Habibi 1992, p. 258), chez les juifs israéliens des années 1970 aux années 2010 (Landman 2016, p. 140), chez des collégiens palestiniens en Israël et en Cisjordanie dans les années 2000 ou 2010 (Amara 2014, p. 222-223), ou encore chez les Malais en Malaisie en 2016 (Robustova 2016, p. 216). Pour expliquer ce phénomène et son caractère relativement général, on avance parfois que les fils, chargés de transmettre le patrimoine familial (exploitation agricole, cheptel, entreprise familiale) ainsi que l'identité familiale et le nom de famille, se voient assez souvent attribuer des prénoms de leurs aïeux, héritant ainsi du petit stock de prénoms traditionnels. Par contraste, le choix du prénom aurait, pour les filles, des motifs plus esthétiques, de mode et de distinction, qui conduisent à enrichir le stock de prénoms attribués.

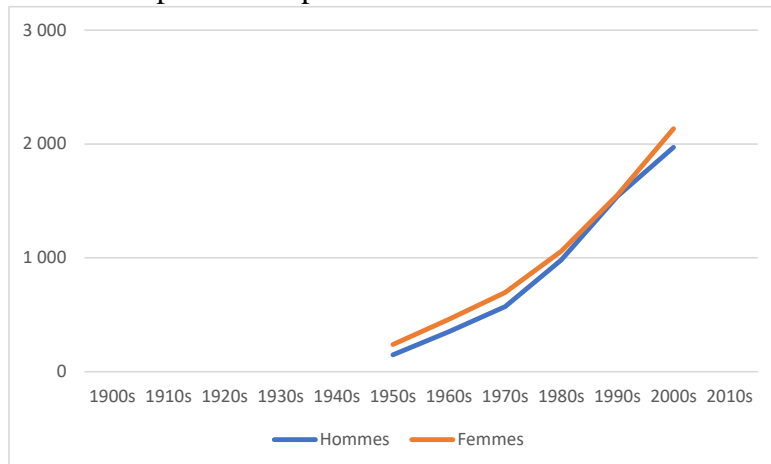
### ➤ Évolution du degré de concentration des prénoms attribués

L'enquête TeO reflète-t-elle la baisse de la part des nouveau-nés qui se voient attribuer l'un des 10 prénoms les plus attribués ? En France depuis les années 1950, de moins en moins de nouveau-nés se voient attribuer l'un des 10 prénoms les plus attribués de la décennie (d'environ 33 % à environ 10 %, comme le montre la figure 5.7b). L'enquête TeO reflète bien ce mouvement de déconcentration (ou de hausse de la dispersion) des prénoms (figure 5.7a), qui est un autre indice du mouvement d'individualisation et qui s'observe aussi bien pour les prénoms masculins que féminins.

## Taille du stock de prénoms en France, selon l'enquête TeO et selon le Fichier INSEE des prénoms

### ENQUÊTE TeO

Figure 5.6a. Nombre de prénoms différents portés en 2008, par sexe et par décennie de naissance

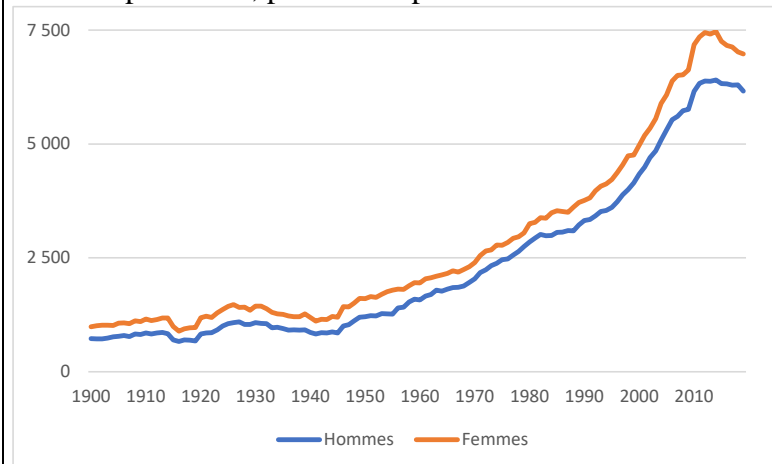


Champ : individus du champ retenu (personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008) qui sont nés en France entière des années 1950 aux années 2000, données pondérées, N = 32 022.

Source : TeO (voir Annexe 5B, p. 183).

### FICHIER INSEE DES PRÉNOMS

Figure 5.6b. Nombre de prénoms différents attribués depuis 1900, par sexe et par année de naissance



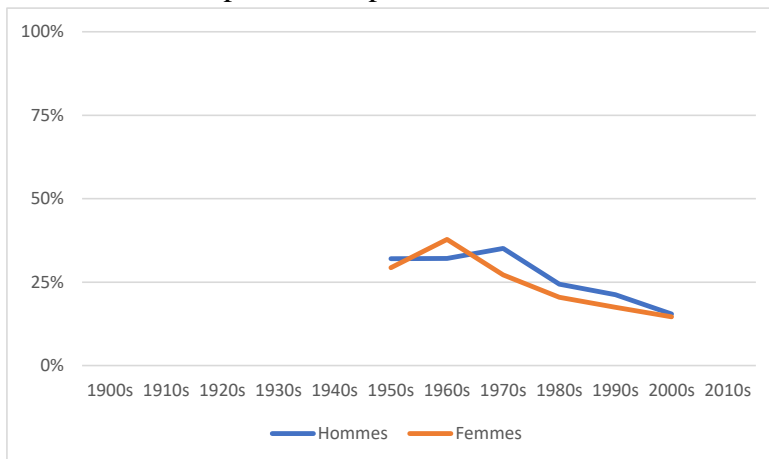
Champ : naissances en France entière (hors Mayotte) de 1900 à 2019, N = 85 047 407.

Source : [Fichier INSEE des prénoms](#).

## Degré de concentration des prénoms en France, selon l'enquête TeO et selon le Fichier INSEE des prénoms

### ENQUÊTE TeO

Figure 5.7a. Part des habitants en 2008 qui portent l'un des 10 prénoms les plus attribués de leur décennie de naissance, par sexe et par décennie de naissance

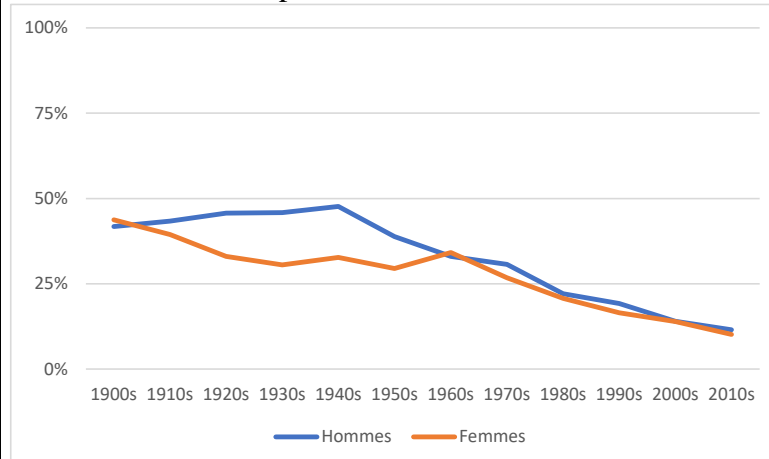


Champ : individus du champ retenu (personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008) qui sont nés en France entière des années 1950 aux années 2000, données pondérées, N = 32 022.

Source : TeO (voir Annexe 5B, p. 183).

### FICHIER INSEE DES PRÉNOMS

Figure 5.7b. Part des nouveau-nés qui se voient attribuer l'un des 10 prénoms les plus attribués de la décennie, par sexe et par décennie de naissance

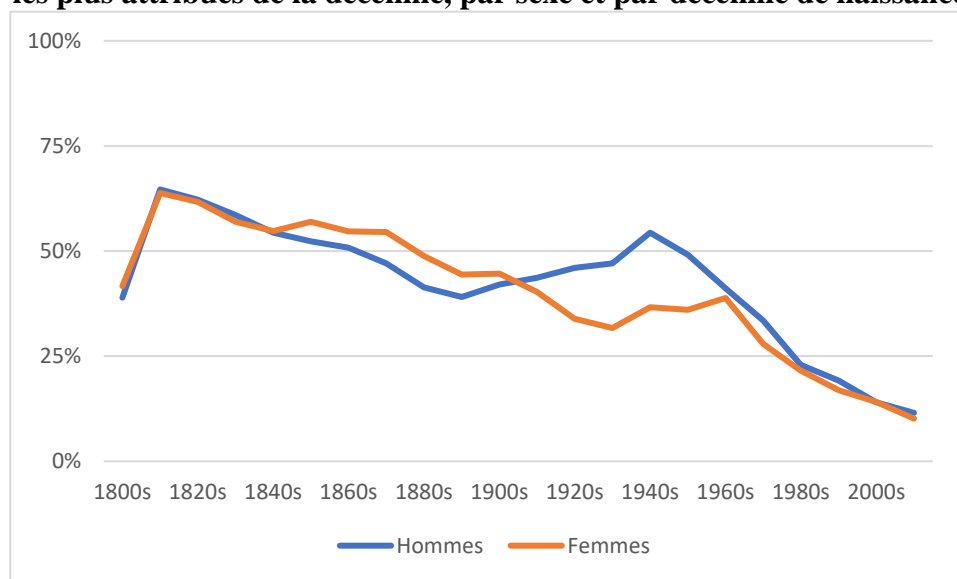


Champ : naissances en France entière (hors Mayotte) de 1900 à 2019, N = 85 047 407.

Source : [Fichier INSEE des prénoms](#).

Des travaux ayant déjà été conduits sur la déconcentration des prénoms attribués en France au 19<sup>e</sup> siècle (Dupâquier et al. 1987),<sup>83</sup> il est possible de retracer ce mouvement sur deux siècles (figure 5.7c). On observe alors que 65 % des nouveau-nés des années 1810 se voyaient attribuer l'un des 10 prénoms les plus attribués de la décennie, contre 10 % des nouveau-nés des années 2010 : il s'agit là d'une évolution d'ampleur majeure, qui témoigne de l'essor de l'individualisme culturel, ou autrement dit d'une massification du souci de n'être pas fondu dans la masse. Si l'on n'examine que le prénom le plus attribué, l'évolution est là encore frappante : alors que dans les années 1810, 14 % des garçons sont prénommés « Jean » et 19 % des filles « Marie », dans les années 2010 seuls 1 % des garçons sont prénommés « Gabriel » et 1 % des filles « Emma ».

**Figure 5.7c. Part des nouveau-nés en France qui se voient attribuer l'un des 10 prénoms les plus attribués de la décennie, par sexe et par décennie de naissance**



Champ : naissances en France entière de 1800 à 2019.

Sources : Dupâquier et al. 1987, p. 106-107 (1800-1899, N = 89 379) ; [Fichier INSEE des prénoms](#) 2019 (1900-2019, N = 85 047 407), avec précisions méthodologiques [ici](#). Conformément à la convention adoptée par Dupâquier et al. 1987, sur ce graphique le prénom masculin « Jean » comprend tous les composés masculins débutant par « Jean- » et le prénom féminin « Marie » comprend tous les composés féminins débutant par « Marie- ».

Les prénoms les plus classiques de ces deux derniers siècles en France se retrouvent plus généralement dans le monde occidental. Le prénom masculin le plus attribué en France des années 1800 jusqu'aux années 1950 est « Jean » (Dupâquier et al. 1987 ; tableau 5.5). Son cognat<sup>84</sup> anglais « John » est aussi le prénom le plus attribué aux États-Unis des années 1880 jusqu'aux années 1910 (Social Security 2020), ainsi qu'en Angleterre-Galles des années 1910 aux années 1940 (Office for National Statistics 2014 et 2020). Leur cognat espagnol « Juan » est bien aussi le prénom masculin le plus attribué à Mexico (Mexique) aux 16<sup>e</sup> et 17<sup>e</sup> siècles, avant d'être dépassé par « Joseph » ou « Josep » puis « José » du 18<sup>e</sup> siècle jusqu'à la seconde moitié du 20<sup>e</sup> siècle (Boyd-Bowman 1970). Le prénom féminin le plus attribué en France des années 1800 jusqu'aux années 1950 est « Marie » (Dupâquier et al. 1987 ; tableau 5.5).<sup>85</sup>

<sup>83</sup> Dupâquier et al. (1987) exploitent les 89 379 actes de naissance du 19<sup>e</sup> siècle des descendants de 3 000 couples représentatifs de la population française mariés entre 1803 et 1832.

<sup>84</sup> Des « cognats » (ou mots apparentés) sont des mots de plusieurs langues qui ont une origine (étymologie) commune.

<sup>85</sup> Sur la baisse de la fréquence d'attribution du prénom « Marie » en France au cours du 20<sup>e</sup> siècle, voir Fourquet et Manternach 2019a, p. 69-72.

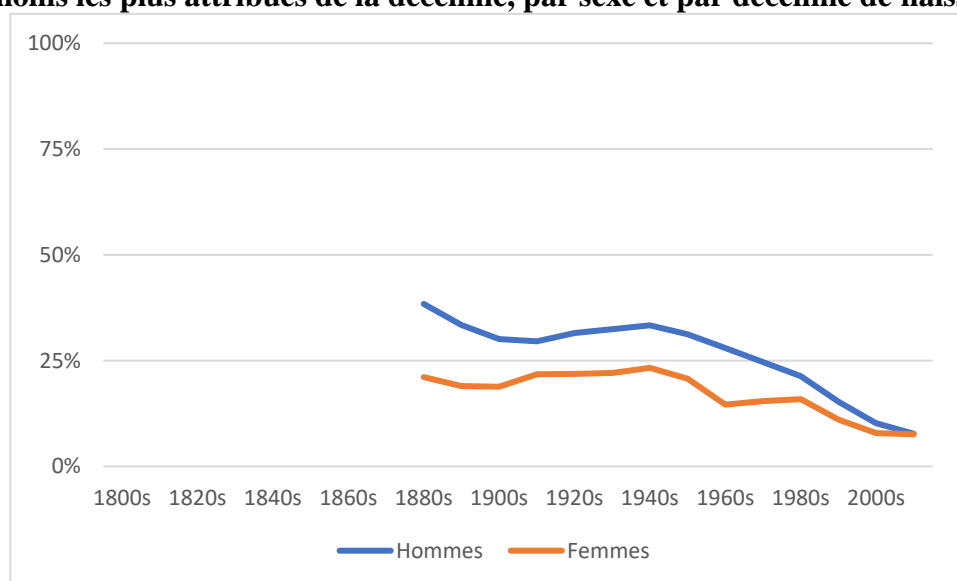
« Mary » est le prénom le plus attribué aux États-Unis des années 1880 jusqu'aux années 1950 (Social Security 2020), ainsi qu'en Angleterre-Galles des années 1900 jusqu'aux années 1910 (Office for National Statistics 2014 et 2020). « Maria » est aussi le prénom féminin le plus attribué à Mexico (Mexique) depuis le 16<sup>e</sup> siècle ; ainsi parmi les filles nées à Mexico en 1800, 92 % se voient attribuer le prénom « María » parmi leurs prénoms (elles en ont en moyenne quatre) (Boyd-Bowman 1970, p. 21 et 24).<sup>86</sup> De même, « Maria » est le prénom le plus attribué au Brésil des années 1930 aux années 2000 (Instituto Brasileiro de Geografia e Estatística 2010). Les prénoms « Jean » et « Marie » se retrouvent aussi largement dans la littérature française du 17<sup>e</sup> au 20<sup>e</sup> siècles (base de données Frantext), où les prénoms chrétiens les plus fréquemment mentionnés sont, par ordre décroissant, « Jean » (qui apparaît dans 1 528 textes sur 3 000), « Louis », « Pierre », « Paul » et « Charles », devant le premier prénom féminin : « Marie » (1 104 textes sur 3 000) (Brunet 1988, p. 131).

La diversification des prénoms au fil du temps et le caractère longtemps plus concentré des prénoms masculins ne sont pas, eux non plus, spécifiques à la France. Il en va de même à Mexico (Mexique), où les prénoms sont de nos jours beaucoup moins concentrés qu'aux 16<sup>e</sup> et 17<sup>e</sup> siècles, lorsque 60 % à 75 % des nouveau-nés se voyaient attribuer l'un des 10 prénoms les plus attribués à leur sexe (Boyd-Bowman 1970, p. 16). On observe les mêmes phénomènes aux États-Unis, où la part des nouveau-nés qui se voient attribuer l'un des 10 prénoms les plus attribués de la décennie a fortement baissé des années 1880 aux années 2010, de 38 % à 8 % pour les garçons et de 21 % à 8 % pour les filles (figure 5.7d ; Twenge et al. 2010, p. 21). Depuis plus d'un siècle les prénoms sont d'ailleurs globalement moins concentrés aux États-Unis qu'en France (figures 5.7c et 5.7d), comme si les États-Unis étaient plus avancés dans le mouvement séculaire d'individualisation. Les prénoms tendent aussi à se déconcentrer (et à être relativement plus concentrés pour les garçons) en Angleterre-Galles depuis au moins les années 1990 (Stripe 2019 ; Office for National Statistics 2020). À Java (Indonésie), la part de nouveau-nés qui se voient attribuer l'un des 20 prénoms les plus attribués baisse depuis les années 1980 (Kuipers et Askuri 2017, p. 39-40).

---

<sup>86</sup> Avec l'essor du dogme de l'Immaculée Conception en Espagne au 17<sup>e</sup> siècle, le prénom « María » au Mexique est le plus souvent accompagné, aux 18<sup>e</sup> et 19<sup>e</sup> siècles, d'invocations mariales (*advocaciones de la Virgen*) comme « Guadalupe », « Dolores », « Concepción », « Soledad », « Luz », « Carmen », « Mercedes », etc. (Boyd-Bowman 1970, p. 17-20).

**Figure 5.7d. Part des nouveau-nés aux États-Unis qui se voient attribuer l'un des 10 prénoms les plus attribués de la décennie, par sexe et par décennie de naissance**



Champ : naissances aux États-Unis de 1880 à 2019, N = 372 948 994. Sont exclus les individus nés avant 1937 et n'ayant jamais détenu de carte de sécurité sociale.

Sources : Social Security 2020. Pour les précisions méthodologiques, voir [ici](#).

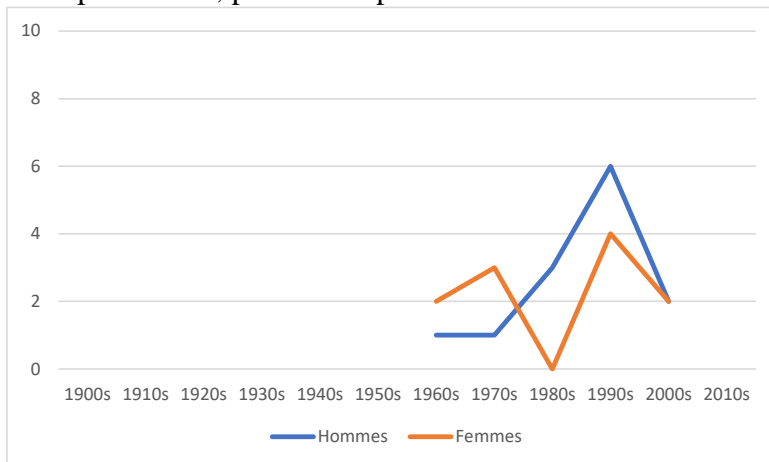
#### ➤ Évolution du rythme de renouvellement des prénoms les plus attribués

L'enquête TeO reflète-t-elle la hausse de la vitesse de rotation — le raccourcissement de la durée de succès — des prénoms les plus attribués ? Alors qu'au début du 20<sup>e</sup> siècle la plupart des prénoms du Top 10 d'une décennie figuraient dans le Top 10 de la décennie précédente, tel n'est plus le cas aujourd'hui (figure 5.8b ; Desplanques 1986, p. 68-69). L'enquête TeO ne reflète qu'imparfaitement ces ordres de grandeur, qui reposent sur un nombre de prénoms relativement restreint (figure 5.8a).

## Rythme de renouvellement des prénoms les plus attribués en France, selon l'enquête TeO et selon le Fichier INSEE des prénoms

### ENQUÊTE TeO

Figure 5.8a. Nombre de prénoms du Top 10 de la décennie de naissance qui figuraient dans le Top 10 de la décennie précédente, par sexe et par décennie de naissance

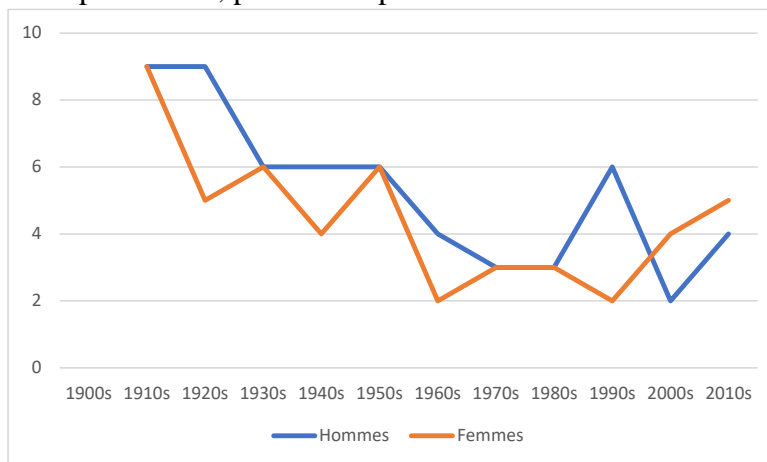


Champ : individus du champ retenu (personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008) qui sont nés en France entière des années 1950 aux années 2000, données pondérées, N = 32 022.

Source : TeO (voir Annexe 5B, p. 183).

### FICHIER INSEE DES PRÉNOMS

Figure 5.8b. Nombre de prénoms du Top 10 de la décennie de naissance qui figuraient dans le Top 10 de la décennie précédente, par sexe et par décennie de naissance



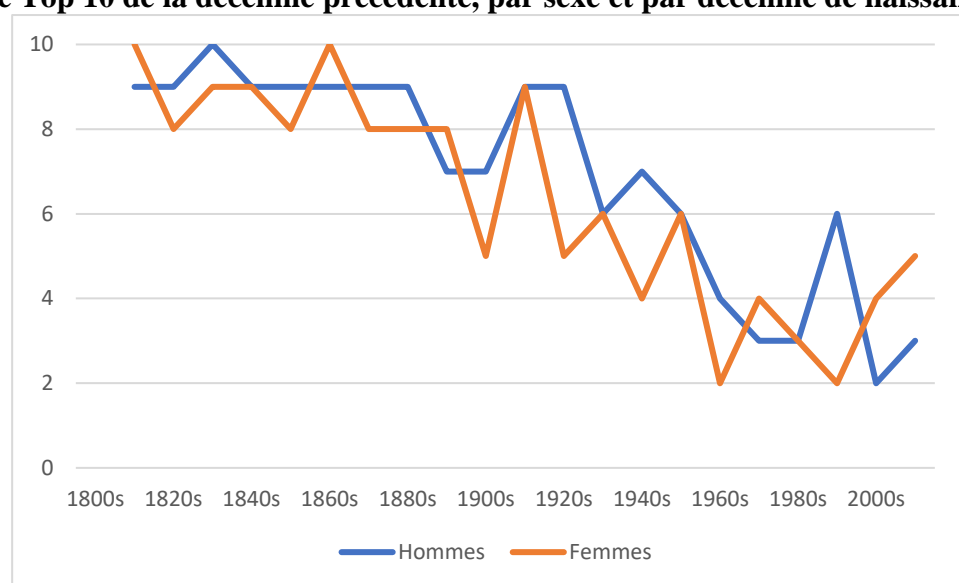
Champ : naissances en France entière (hors Mayotte) de 1900 à 2019, N = 85 047 407.

Source : [Fichier INSEE des prénoms](#).



Alors que jusqu'au 19<sup>e</sup> siècle la grande majorité des prénoms les plus attribués lors de chaque décennie faisaient déjà partie du Top 10 de la décennie précédente — ou qu'autrement dit les prénoms les plus attribués étaient relativement stables au fil du temps —, depuis le début du 20<sup>e</sup> siècle les prénoms les plus attribués se renouvellent de plus en plus rapidement (figure 5.8c). Alors que « Jean », Pierre », « Louis », « François » et « Joseph » auront figuré dans le Top 10 des prénoms masculins les plus attribués tout au long du 19<sup>e</sup> siècle, les « Enzo », « Théo », « Nathan », « Mathis » et « Clément » des années 2000 n'y auront figuré qu'au cours d'une seule décennie. L'évolution est encore plus claire concernant les prénoms féminins : à la stabilité des « Marie », « Jeanne », « Anne » et « Louise » tout au long du 19<sup>e</sup> siècle s'oppose la brièveté d'attribution des prénoms « Clara », « Sarah » et « Océane » des années 2000. Autrement dit, la durée de vie des prénoms se raccourcit, si bien qu'ils en viennent à signaler de façon de plus en plus fiable la génération et l'âge de leur porteur, ce qui n'était pas le cas auparavant (Bozon 1987, p. 89-91). En France on est donc passé, au tournant des 19<sup>e</sup> et 20<sup>e</sup> siècles, « du “nom de baptême” transmis par les parrain et marraine au “prénom” choisi par les parents » (Boyer 2016, p. 12), « du respect de la tradition au souci de distinction » (Boyer 2016, p. 18). Dans la plupart des pays européens depuis le 19<sup>e</sup> siècle, non seulement le stock de prénoms féminins est plus fourni que le stock de prénoms masculins, mais le rythme de renouvellement des prénoms est plus rapide pour les prénoms féminins que masculins (Lawson 1995, p. 1744 ; Leibring 2016, p. 210).

**Figure 5.8c. Nombre de prénoms du Top 10 de la décennie en France qui figuraient dans le Top 10 de la décennie précédente, par sexe et par décennie de naissance**



Champ : naissances décennales en France entière de 1800-1809 à 2010-2019.

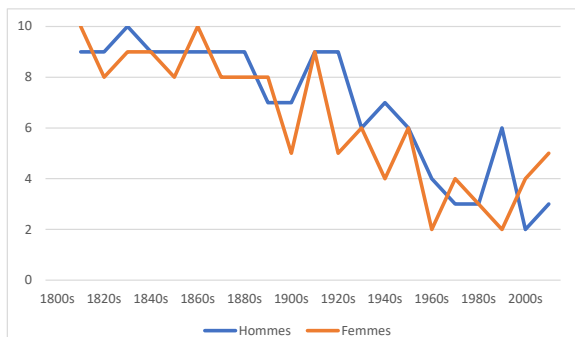
Sources : Dupâquier et al. 1987, p. 106-107 (1800-1899, N = 89 379) ; [Fichier INSEE des prénoms](#) 2019 (1900-2019, N = 85 047 407), avec précisions méthodologiques [ici](#). Conformément à la convention adoptée par Dupâquier et al. 1987, sur ce graphique le prénom masculin « Jean » comprend tous les composés masculins débutant par « Jean- » et le prénom féminin « Marie » comprend tous les composés féminins débutant par « Marie- ».

Le raccourcissement de la durée de succès des prénoms les plus attribués s'observe aussi dans les autres pays occidentaux qui mettent à disposition des données de long terme sur les prénoms attribués. C'est le cas de l'Angleterre-Galles, des États-Unis et du Brésil, selon des degrés, des calendriers et des écarts hommes-femmes variables (figures 5.8d à 5.8f). C'est le cas aussi à Java (Indonésie), où depuis les années 1980 les prénoms les plus attribués changent de plus en plus largement d'une décennie sur l'autre (Kuipers et Askuri 2017, p. 43-45).

## Rythme de renouvellement des prénoms les plus attribués : comparaison internationale

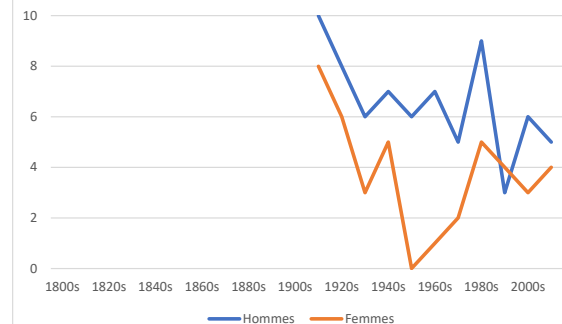
### Nombre de prénoms du Top 10 de la décennie qui figuraient dans le Top 10 de la décennie précédente

**Figure 5.8c. FRANCE**



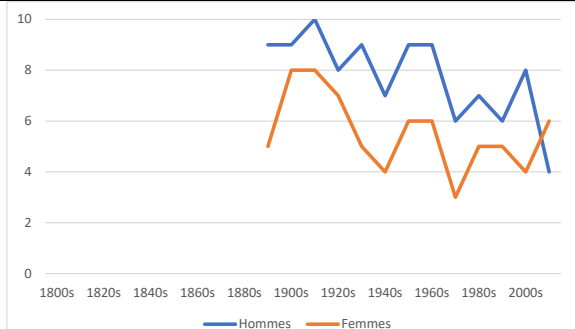
Champ : naissances décennales en France entière de 1800-1809 à 2010-2019.  
Sources : Dupâquier et al. 1987, p. 106-107 (1800-1899, N = 89 379) ; [Fichier INSEE des prénoms](#) 2019 (1900-2019, N = 85 047 407). Précisions méthodologiques ci-dessus.

**Figure 5.8d. ANGLETERRE-GALLES**



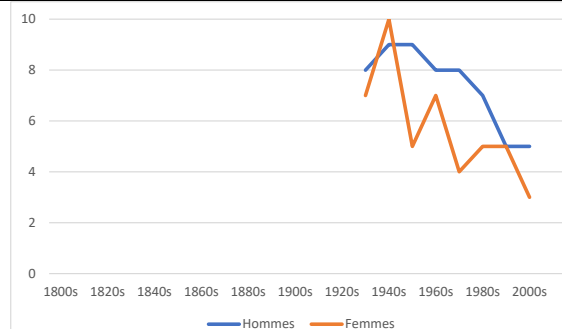
Champ : naissances annuelles tous les 10 ans en Angleterre-Galles de 1904 à 2014, N = inconnu.  
Sources : Office for National Statistics 2014 et 2020. Précisions méthodologiques [ici](#).

**Figure 5.8e. ÉTATS-UNIS**



Champ : naissances décennales aux États-Unis de 1880-1889 à 2010-2019, N = 372 948 994. Sont exclus les individus nés avant 1937 et n'ayant jamais détenu de carte de sécurité sociale.  
Source : Social Security 2020. Précisions méthodologiques [ici](#).

**Figure 5.8f. BRÉSIL**



Champ : décennie de naissance des individus recensés au Brésil en 2010, N ~ 200 000 000.  
Source: Instituto Brasileiro de Geografia e Estatística 2010. Précisions méthodologiques [ici](#).

### ➤ Évolution de la morphologie et de la phonologie des prénoms attribués

L'enquête TeO reflète globalement bien l'évolution de la taille du stock de prénoms ainsi que l'évolution de sa concentration et de son rythme de renouvellement. Mais reflète-t-elle aussi l'évolution de la morphologie des prénoms, qui a été moins étudiée ?<sup>87</sup>

La part de prénoms composés, très faible parmi les prénoms attribués au début du 20<sup>e</sup> siècle comme aujourd'hui, a connu son maximum au milieu du 20<sup>e</sup> siècle et notamment dans les années 1950, avec des prénoms composés masculins (« Jean-Pierre », « Jean-Claude », « Jean-Luc », « Jean-Paul » ou « Jean-Louis ») plus souvent que féminins (« Marie-Christine », « Anne-Marie » ou « Marie-Claude ») (figure 5.9b). Là encore l'enquête TeO reflète bien les faits observés depuis les années 1950 (figure 5.9a), tout en surestimant légèrement la proportion de prénoms composés, peut-être parce que certains enquêtés TeO ont indiqué, pour eux-mêmes et/ou leurs enfants, le premier prénom accompagné du deuxième, et parce que le nettoyage des prénoms de l'enquête TeO conduit à considérer les prénoms comprenant une apostrophe comme des prénoms composés. En outre, les prénoms féminins comme masculins qui sont attribués sont de plus en plus courts (figure 5.9d). Alors que jusqu'aux années 1970 les prénoms féminins attribués étaient en moyenne plus longs que les prénoms masculins, depuis les années 1980 tel n'est plus le cas. La signification de ces évolutions mériterait d'être analysée dans le détail, mais quoi qu'il en soit l'enquête TeO les reflète bien (figure 5.9c).

L'enquête TeO décrit-elle bien, enfin, l'évolution de la phonologie des prénoms attribués en France ? Depuis les années 1950, la part de prénoms attribués débutant par une voyelle a légèrement augmenté (figure 5.10b), tandis que la part de prénoms féminins terminant par une voyelle baissait légèrement (figure 5.10d). L'enquête TeO décrit bien ces faits (figures 5.10a et 5.10c), ce qui achève de me convaincre que **l'échantillon de l'enquête TeO et la variable « prénom » sont de bonne qualité**. Le fait le plus saillant reste l'écart de terminaison entre prénoms masculins et féminins : de nos jours environ 85 % des prénoms féminins attribués terminent par une voyelle, contre seulement 34 % des prénoms masculins (figures 5.10c et 5.10d). De même, parmi les prénoms les plus attribués en Pennsylvanie en 1990, la plupart des prénoms féminins terminaient par l'une des trois voyelles -a, -e ou -i, alors que la plupart des prénoms masculins terminaient par les autres lettres (sauf -h et -y, de fréquence égale entre les sexes) (Barry et Harper 2000). Si en français la terminaison féminine est en -e, en latin et dans la plupart des langues romanes ainsi que des langues slaves elle est en -a, et elle est aussi en -a dans des langues sémitiques comme l'arabe et l'hébreu. Comme le cerveau humain tend à associer à des sonorités comme « bouba » une forme ronde et à des sonorités comme « kiki » une forme pointue (c'est [l'effet bouba-kiki](#)), plusieurs recherches visent à savoir si les prénoms masculins et féminins auraient des sonorités différentes et, si oui, si ces différences de sonorités reflètent des différences d'attentes vis-à-vis des garçons et des filles. Parmi les prénoms de langue anglaise, il semblerait ainsi que les prénoms féminins présentent relativement souvent des sonorités « rondes » (« Amelia ») ou évoquant une petite taille (« Emily »), tandis que les prénoms masculins présentent assez souvent des sonorités « pointues » (« Jack ») ou évoquant une grande taille (« Thomas ») (Pitcher et al. 2013 ; Sidhu et Pexman 2015 ; voir aussi Whissell 2001).

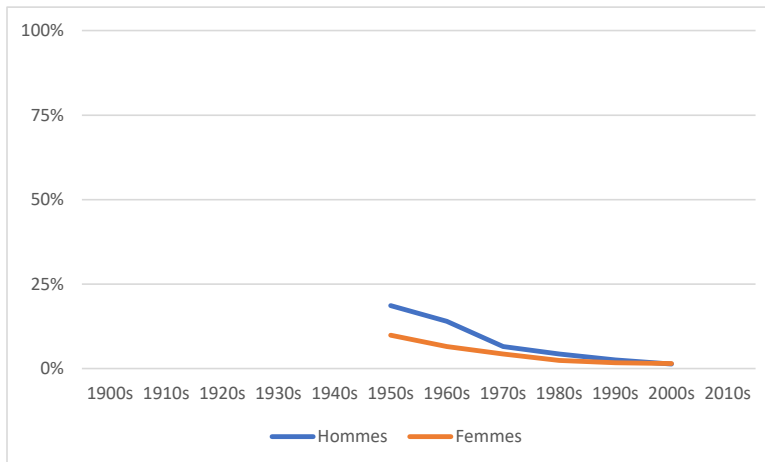
---

<sup>87</sup> Je propose ici quelques analyses élémentaires, que je serais heureux de prolonger avec l'aide — indispensable — d'un linguiste. Si des linguistes étaient intéressés à travailler sur la morphologie et la phonologie des prénoms masculins et féminins attribués en France depuis le 19<sup>e</sup> siècle, qu'ils n'hésitent pas à me contacter.

## Morphologie des prénoms en France, selon l'enquête TeO et selon le Fichier INSEE des prénoms

### ENQUÊTE TeO

Figure 5.9a. Part de prénoms portés en 2008 qui sont des prénoms composés, par sexe et par décennie de naissance

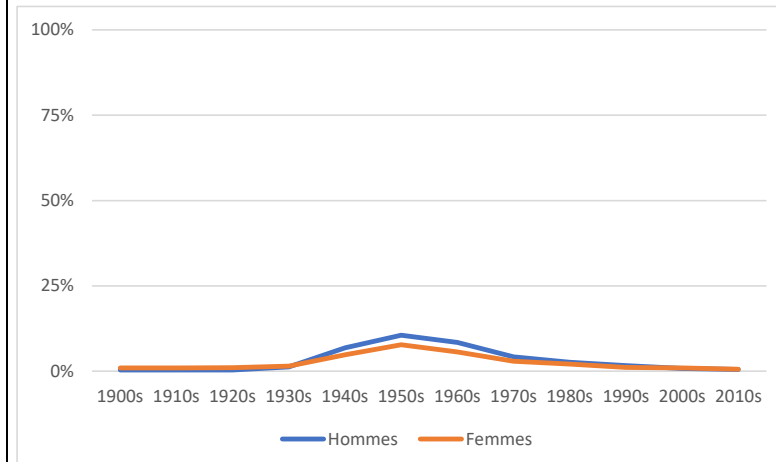


Champ : individus du champ retenu (personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008) qui sont nés en France entière des années 1950 aux années 2000, N = 32 022.

Source : TeO (voir Annexe 5B, p. 183).

### FICHIER INSEE DES PRÉNOMS

Figure 5.9b. Part de prénoms attribués depuis 1900 qui sont des prénoms composés, par sexe et par décennie de naissance



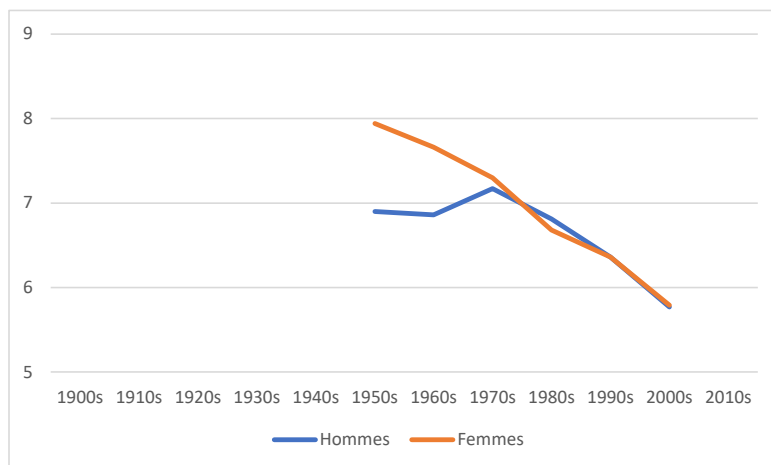
Champ : naissances en France entière (hors Mayotte) de 1900 à 2019, prénoms « rares » exclus, N = 84 264 960.

Source : [Fichier INSEE des prénoms](#).

## Morphologie des prénoms en France, selon l'enquête TeO et selon le Fichier INSEE des prénoms

### ENQUÊTE TeO

Figure 5.9c. Nombre moyen de lettres des prénoms portés en 2008, par sexe et par décennie de naissance

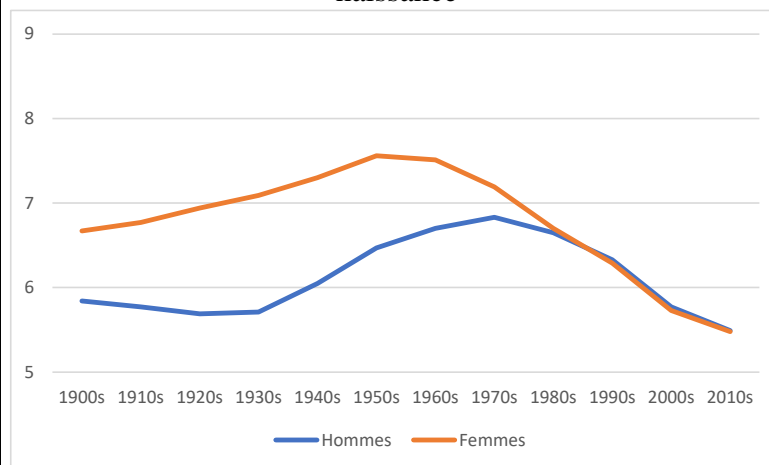


Champ : individus du champ retenu (personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008) qui sont nés en France entière des années 1950 aux années 2000, N = 32 022.

Source : TeO (voir Annexe 5B, p. 183).

### FICHIER INSEE DES PRÉNOMS

Figure 5.9d. Nombre moyen de lettres des prénoms attribués depuis 1900, par sexe et par décennie de naissance



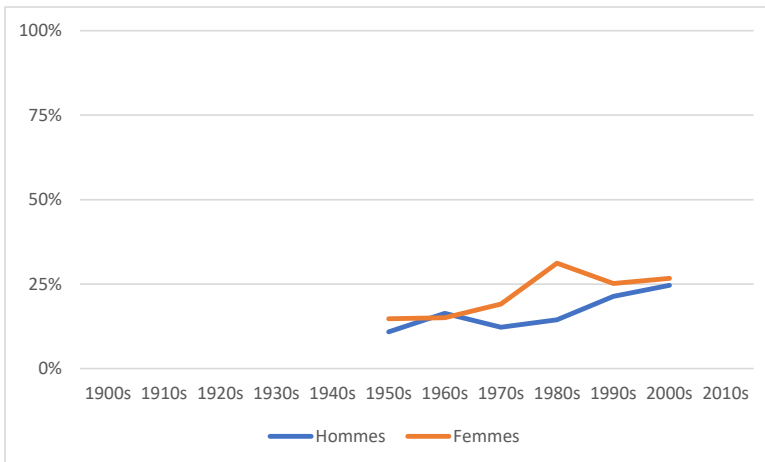
Champ : naissances en France entière (hors Mayotte) de 1900 à 2019, prénoms « rares » exclus, N = 84 264 960.

Source : [Fichier INSEE des prénoms](#).

## Phonologie des prénoms en France, selon l'enquête TeO et selon le Fichier INSEE des prénoms

### ENQUÊTE TeO

Figure 5.10a. Part de prénoms portés en 2008 qui débutent par une voyelle, par sexe et par décennie de naissance

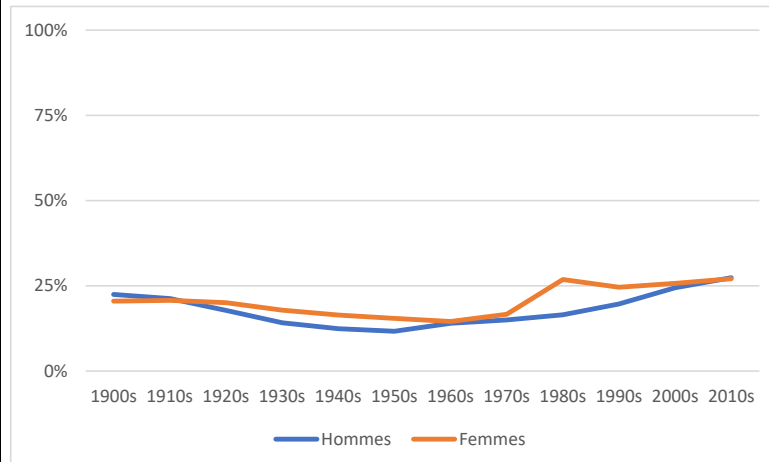


Champ : individus du champ retenu (personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008) qui sont nés en France entière des années 1950 aux années 2000, N = 32 022.

Source : TeO (voir Annexe 5B, p. 183).

### FICHIER INSEE DES PRÉNOMS

Figure 5.10b. Part de prénoms attribués depuis 1900 qui débutent par une voyelle, par sexe et par décennie de naissance



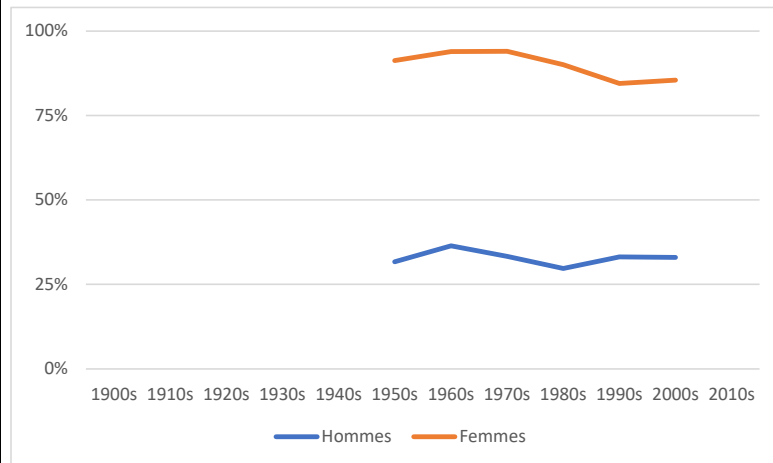
Champ : naissances en France entière (hors Mayotte) de 1900 à 2019, prénoms « rares » exclus, N = 84 264 960.

Source : [Fichier INSEE des prénoms](#).

## Phonologie des prénoms en France, selon l'enquête TeO et selon le Fichier INSEE des prénoms

### ENQUÊTE TeO

Figure 5.10c. Part de prénoms portés en 2008 qui terminent par une voyelle, par sexe et par décennie de naissance

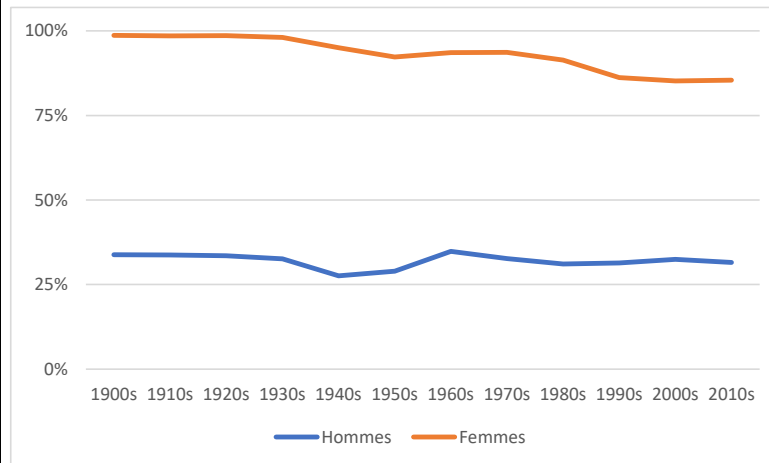


Champ : individus du champ retenu (personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008) qui sont nés en France entière des années 1950 aux années 2000, N = 32 022.

Source : TeO (voir Annexe 5B, p. 183).

### FICHIER INSEE DES PRÉNOMS

Figure 5.10d. Part de prénoms attribués depuis 1900 qui terminent par une voyelle, par sexe et par décennie de naissance



Champ : naissances en France entière (hors Mayotte) de 1900 à 2019, prénoms « rares » exclus, N = 84 264 960.

Source : [Fichier INSEE des prénoms](#).



### ❖ Résumé

De plus en plus de prénoms différents sont attribués aux nouveau-nés en France : moins de 2 000 prénoms différents par an dans les années 1900, contre plus de 13 000 par an depuis 2010. En outre, de moins en moins de nouveau-nés se voient attribuer l'un des 10 prénoms les plus fréquemment attribués de la décennie, de 65 % des nouveau-nés en 1810-1819 à 10 % en 2010-2019. Les prénoms les plus attribués ont aussi une durée de succès de plus en plus brève : alors que les prénoms « Jean » et « Marie » auront figuré dans le Top 10 des prénoms les plus attribués tout au long du 19<sup>e</sup> siècle, les « Enzo » et « Clara » de 2000-2009 n'y auront figuré qu'au cours d'une seule décennie. C'est ce raccourcissement de la durée de succès des prénoms qui fait que de nos jours les prénoms sont un indicateur assez fiable de l'âge de leur porteur, ce qu'ils n'étaient pas autant jusqu'au 19<sup>e</sup> siècle. En France comme ailleurs en Occident et dans le monde, de plus en plus de parents choisissent pour leurs enfants — et particulièrement pour leurs filles — des prénoms relativement originaux, distinctifs et individualisants.

## Annexe 5B : Le programme informatique de la reproduction « au plus près »

Cette Annexe 5B, associée au Chapitre 5, p. 24, comprend la totalité du programme informatique Stata que j'ai utilisé pour reproduire « au plus près » les résultats de l'article de Coulmont et Simon (2019). Tout chercheur qui dispose des bases TeO enrichies de la variable « prénom » (et, pour certaines analyses méthodologiques, des trois nomenclatures de prénoms par origine) pourra en 1 clic exécuter le programme informatique Stata ci-dessous (version 15.1) pour obtenir à l'identique la totalité des résultats présentés dans ce travail. Je tiens aussi à la disposition de tout un chacun ce programme informatique sous format Stata (.do).

### ❖ Le programme informatique de la reproduction « au plus près » (PROG\_REPRO)

```
/*
log using C:\Users\Public\Documents\results_repro.log, text
*/

*****
*****
/***** PRENOMS DES DESCENDANTS D'IMMIGRES EN FRANCE: *****/
***** ESSAI DE REPRODUCTION D'UN ARTICLE SCIENTIFIQUE *****/
*****
*****

/*
Dans le répertoire du CASD "C:\Users\Public\Documents", j'ai déposé:
- le présent programme, appelé PROG_REPRO;
- 4 des tables TeO nominatives telles que communiquées par le CASD:
il s'agit des tables THL2_PRENOM, INDIV2_PRENOM, ENFANTS2_PRENOM et POLREL2;
- le fichier INSEE des prénoms (1900-2019), appelé INSEE;
- "mes" 3 nomenclatures de prénoms par origine, respectivement appelées
Nomenclature1, Nomenclature2 et Nomenclature3; et
- les nomenclatures de prénoms par origine de l'INED, contenues dans les
3 tables THL2_CAT1, INDIV2_CAT1 et ENFANTS2_CAT1;
- soit au total 1 programme et 11 tables, le tout sous format Stata.
*/

cd "C:\Users\Public\Documents"

*****
/*CHAPITRE 5. Méthodologie de la reproduction "au plus près":
la base de données*/
*****

/*Chapitre 5: Des programmes informatiques insuffisants*/

/*Chapitre 5: L'échantillon maximal*/

*J'examine la table THL2_PRENOM
use thl2_prenom
tab sexe, missing
tab lienego, missing
codebook prenom /*la variable PRENOM est bien renseignée pour les 75 121
individus: ils portent 16 938 prénoms différents*/

*J'examine la table INDIV2_PRENOM
use indiv2_prenom
tab sexee, missing
codebook prenome /*la variable PRENOME est bien renseignée pour les 21 761
individus: ils portent 6 699 prénoms différents*/

*J'examine la table ENFANTS2_PRENOM
use enfants2_prenom
tab e_sexe, missing
tab e_presmen, missing /*23 552 enfants de cette table habitent dans le logement
```

```

de l'enquêté, 7 924 enfants n'y habitent pas (y compris 314 qui sont décédés)*/
tab e_presmen e_vie, missing
codebook e_prenom /*la variable E_PRENOM est bien renseignée pour les 31 476
individus: ils portent 8 844 prénoms différents*/

/*Chapitre 5: Les variables*/

*Les 8 variables (dont 2 identifiants) issues de la table THL2_PRENOM
use thl2_prenom

*IDENT (identifiant de l'enquêté)
summarize ident /*Cette variable est bien renseignée pour les 75 121
habitants du logement et va de 1 à 21 761.*/

*NOI (numéro d'ordre de l'individu dans le logement)
/*tab noi, missing*/ /*Cette variable est bien renseignée pour les 75 121
habitants du logement et va de 1 à 19.*/

*POIDSI (pondération de l'habitant du logement)
summarize poidsi /*Cette variable est bien renseignée pour les 75 121
habitants du logement et va de 13 à 59 700.*/
tab sexe [aweight=poidsi], missing /*La somme des habitants du logement pondérés est de 75 121
individus, tout comme la somme des individus non pondérés, parce que la pondération Stata
ramène l'effectif total au nombre de lignes pris pour champ*/

*SEXE (sexe de l'habitant du logement)
tab sexe, missing /*Cette variable est bien renseignée pour les 75 121
habitants du logement et va de 1 Homme à 2 Femme.*/

*ANAIIS (année de naissance de l'habitant du logement)
/*tab anais, missing*/ /*Cette variable est bien renseignée pour les 75 121
habitants du logement et va de 1900 à 2009.*/

*REGIONNAIS2 (lieu de naissance de l'habitant du logement)
tab regionnais2, missing /*Cette variable est bien renseignée pour les 75 121
habitants du logement et va de 1101 à 9999*/
destring regionnais2, replace
recode regionnais2 (1101=11/*11=France métro*/) (1201/1205=12/*12=DOM*/)
(2101/2301=20/*20=Maghreb*/) (2401/2601=29/*29=autre Afrique*/) (3101/3301=30/*30=Asie*/)
(3401=34/*34=Turquie*/) (3501=30/*30=Asie*/) (4101/4301=40/*40=Europe du Sud*/)
(4401/4801=49/*49=autre Europe*/) (5102/5107=50/*50=Ailleurs dans le monde*/)
(9999=90/*90=NSP*/) (.=/*.=NR*/)
tab regionnais2, missing

*PRENOM (prénom de l'habitant du logement)
codebook prenom /*Cette variable est bien renseignée pour les 75 121
habitants du logement: ils portent 16 938 prénoms différents*/

*LIENEGO (lien avec l'enquêté)
tab lienego, missing /*Cette variable est bien renseignée pour les 75 121
habitants du logement et va de 00 à 90.*/

*J'enregistre
save thl2_prenom, replace

*Les 41 variables (dont 2 identifiants) issues de la table INDIV2_PRENOM
use indiv2_prenom

*IDENT (identifiant de l'enquêté)
summarize ident /*Cette variable est bien renseignée pour les 21 761
enquêtés et va de 1 à 21 761.*/

*NOI (numéro d'ordre de l'enquêté dans le logement); cette variable nous permettra ci-dessous
d'apparier chaque enquêté de la table INDIV2_PRENOM à chaque enquêté de la table THL2_PRENOM
/*tab ne, missing*/ /*Cette variable est bien renseignée pour les 21 761
enquêtés et va de 1 à 9.*/
generate noi = ne
/*tab noi, missing*/

*POIDSI (pondération de l'enquêté)
summarize poidsi /*Cette variable est bien renseignée pour les 21 761
enquêtés et va de 13 à 59 700.*/
tab sexee [aw=poidsi], missing /*La somme des enquêtés pondérés est de 21 761 individus,
tout comme la somme des individus non pondérés*/

```

```

*SEXEE (sexe de l'enquêté)
tab sexee, missing /*Cette variable est bien renseignée pour les 21 761
enquêtés et va de 1 Homme à 2 Femme.*//

*ANAISE (année de naissance de l'enquêté)
tab anaise, missing /*Cette variable est bien renseignée pour les 21 761
enquêtés et va de 1948 à 1990, ce qui correspond au champ de l'enquête.*//

*PRENOME (prénom de l'enquêté)
codebook prenome /*Cette variable est bien renseignée pour les 21 761
enquêtés; ils portent 6 699 prénoms différents*/

*3 VARIABLES de lien à la migration

*GROUP2 (statut de l'enquêté par rapport à la migration en 14 positions)
tab group2, missing /*Cette variable est bien renseignée pour les 21 761
enquêtés et va de 11 à 56*/

*GROUP2CA (statut du conjoint actuel de l'enquêté par rapport à la migration)
tab group2ca, missing /*Cette variable est bien renseignée pour 15 775
enquêtés et va de 10 à 99*/

*GROUP2PC (statut du 1er conjoint de l'enquêté par rapport à la migration)
tab group2pc, missing /*Cette variable est bien renseignée pour 5 048
enquêtés et va de 10 à 99*/

*9 VARIABLES de lieu de naissance

*REGIONNAISE2 (lieu de naissance de l'enquêté)
/*tab regionnaise2, missing*/ /*Cette variable est renseignée pour 21 761 enquêtés et va
de 1101 à 5107*/

*REGIONNAISP2 (lieu de naissance du père de l'enquêté ou de l'homme qui l'a élevé)
/*tab regionnaisp2, missing*/ /*Cette variable est renseignée pour 21 664 enquêtés et va
de 1101 à 9999*/

*REGIONNAISM2 (lieu de naissance de la mère de l'enquêté ou de la femme qui l'a élevé)
/*tab regionnaism2, missing*/ /*Cette variable est renseignée pour 21 752 enquêtés et va
de 1101 à 9999*/

*REGIONNAISCA2 (lieu de naissance du conjoint actuel de l'enquêté)
/*tab regionnaisca2, missing*/ /*Cette variable est renseignée pour 15 775 enquêtés et va
de 1101 à 9999*/

*REGIONNAISCAP2 (lieu de naissance du père du conjoint actuel de l'enquêté)
/*tab regionnaiscap2, missing*/ /*Cette variable est renseignée pour 10 461 enquêtés et va
de 1101 à 9999*/

*REGIONNAISCAM2 (lieu de naissance de la mère du conjoint actuel de l'enquêté)
/*tab regionnaiscam2, missing*/ /*Cette variable est renseignée pour 10 461 enquêtés et va
de 1101 à 9999*/

*REGIONNAISPC2 (lieu de naissance du 1er conjoint de l'enquêté)
/*tab regionnaisp2, missing*/ /*Cette variable est renseignée pour 5 048 enquêtés et va
de 1101 à 9999*/

*REGIONNAISPCP2 (lieu de naissance du père du 1er conjoint de l'enquêté)
/*tab regionnaisp2, missing*/ /*Cette variable est renseignée pour 3 544 enquêtés et va
de 1101 à 9999*/

*REGIONNAISPCM2 (lieu de naissance de la mère du 1er conjoint de l'enquêté)
/*tab regionnaisp2, missing*/ /*Cette variable est renseignée pour 3 544 enquêtés et va
de 1101 à 9999*/

destring regionnaise2 regionnaism2 regionnaisp2 regionnaisca2 regionnaiscam2 regionnaiscap2
regionnaisp2 regionnaisp2 regionnaisp2, replace
recode regionnaise2 regionnaism2 regionnaisp2 regionnaisca2 regionnaiscam2 regionnaiscap2
regionnaisp2 regionnaisp2 regionnaisp2 (1101=11/*11=France métro*/)
(1201/1205=12/*12=DOM*/) (2101/2301=20/*20=Maghreb*/) (2401/2601=29/*29=autre Afrique*/)
(3101/3301=30/*30=Asie*/) (3401=34/*34=Turquie*/) (3501=30/*30=Asie*/)
(4101/4301=40/*40=Europe du Sud*/) (4401/4801=49/*49=autre Europe*/)
(5102/5107=50/*50=Ailleurs dans le monde*/) (9999=90/*90=NSP*/) (./=NR*/)
tab regionnaise2, missing /*Cette variable (lieu de naissance de l'enquêté) est bien
renseignée pour les 21 761 enquêtés et elle va de 11 à 50*/
tab regionnaisp2, missing /*Cette variable (lieu de naissance du père de l'enquêté)
va de 11 à 90 et contient 97 "." (non réponse)*/
tab regionnaism2, missing /*Cette variable (lieu de naissance de la mère de

```

```

l'enquêté)                                va de 11 à 90 et contient      9 "." (non
réponse)*/
tab regionnaisca2, missing                /*Cette variable (lieu de naissance du conjoint actuel de
l'enquêté)                                va de 11 à 90 et contient  5 986 "." (non
réponse)*/
tab regionnaiscap2, missing                /*Cette variable (lieu de naissance du père du conjoint
actuel de l'enquêté)                       va de 11 à 90 et contient 11 300 "." (non
réponse)*/
tab regionnaiscam2, missing                /*Cette variable (lieu de naissance de la mère du conjoint
actuel de l'enquêté)                       va de 11 à 90 et contient 11 300 "." (non
réponse)*/
tab regionnaispc2, missing                 /*Cette variable (lieu de naissance du 1er conjoint de
l'enquêté)                                va de 11 à 90 et contient 16 713 "." (non
réponse)*/
tab regionnaispcp2, missing                /*Cette variable (lieu de naissance du père du 1er conjoint
de l'enquêté)                              va de 11 à 90 et contient 18 217 "." (non
réponse)*/
tab regionnaispcm2, missing                /*Cette variable (lieu de naissance de la mère du 1er
conjoint de l'enquêté)                     va de 11 à 90 et contient 18 217 "." (non
réponse)*/

*2 VARIABLES de nationalité à la naissance des parents de l'enquêté

*NATPNIV2 (nationalité à la naissance du père de l'enquêté ou de l'homme qui l'a élevé)
/*tab natpniv2, missing*/                 /*Cette variable est renseignée pour 21 664 enquêtés et va
de 1101 à 9999*/

*NATMNIV2 (nationalité à la naissance de la mère de l'enquêté ou de la femme qui l'a élevé)
/*tab natmniv2, missing*/                 /*Cette variable est renseignée pour 21 752 enquêtés et va
de 1101 à 9999*/

*2 VARIABLES de choix de nationalité des parents de l'enquêté nés dans une ancienne colonie
française

*N_INDPER (choix de nationalité du père de l'enquêté si né dans une ancienne colonie
française)
tab n_indper, missing                     /*Cette variable est renseignée pour 814 enquêtés et va de
1 à 9*/

*N_INDMER (choix de nationalité de la mère de l'enquêté si née dans une ancienne colonie
française)
tab n_indmer, missing                     /*Cette variable est renseignée pour 806 enquêtés et va de
1 à 9*/

*13 VARIABLES d'origine "dominante" des individus
/*Ces variables commençant par ORIG_ ont toutes le même codage à 3 chiffres:
Le premier chiffre indique le lien de l'individu à la migration
1=immigré (G1)
2=natif d'un DOM
3=enfant d'1 ou 2 immigrés (G2)
4=enfant d'1 ou 2 natifs d'un DOM
5=groupe "majoritaire", sans ascendant immigré
6=petit-enfant d'1 à 4 immigrés (G3)
7=petit-enfant d'1 à 4 natifs d'un DOM
8=arrière-petit-enfant d'1 à 8 immigrés (G4) ou natifs d'un DOM
9=Ne sait pas
Les deux derniers chiffres indiquent le lieu de naissance de l'ascendant immigré de l'individu
(ou de l'individu lui-même, s'il est immigré)
11=France métropolitaine
12=DOM
20=Maghreb
29=autre Afrique
30=Asie (sauf Turquie)
34=Turquie
40=Europe du Sud
49=autre Europe
50=Ailleurs dans le monde
90=Ne sait pas
. =Non réponse
*/

tab group1, missing                       /*Cette variable (lien à la migration de l'enquêté)
est bien renseignée pour les 21 761 enquêtés et elle va de 1 à 5.*/
tab group1ca, missing                     /*Cette variable (lien à la migration du conjoint actuel de
l'enquêté) est bien renseignée pour les 21 761 enquêtés (sauf les 5 986 qui n'ont pas de
conjoint actuel) et elle va de 1 à 9.*/
tab group1pc, missing                     /*Cette variable (lien à la migration du 1er conjoint de

```

```

l'enquêté) est bien renseignée pour les 21 761 enquêtés (sauf les 16 713 qui n'ont pas de
ler conjoint distinct du conjoint actuel) et elle va de 1 à 9.*/
tab group2, missing /*idem que group1, en plus détaillé*/
tab group2ca, missing /*idem que group1ca, en plus détaillé*/
tab group2pc, missing /*idem que group1pc, en plus détaillé*/
destring group1 group1ca group1pc group2 group2ca group2pc, replace

*ORIG_ENQ (origine "dominante" de l'enquêté)
generate orig_enq = group1*100 + regionnaise2 if (group1==1)
/*si l'enquêté est immigré, on lui associe
son lieu de naissance*/
replace orig_enq = group1*100 + 12 if (group1==2)
/*si l'enquêté est natif d'un DOM, on lui associe
le DOM*/
replace orig_enq = group1*100 + regionnaism2 if (group2==32)
/*si l'enquêté est enfant d'une mère immigrée uniquement, on lui associe
le lieu de naissance de sa mère*/
replace orig_enq = group1*100 + regionnaisp2 if (group2==33)
/*si l'enquêté est enfant d'un père immigré uniquement, on lui associe
le lieu de naissance de son père*/
replace orig_enq = group1*100 + regionnaism2 if (group2==31 & regionnaism2==regionnaisp2)
/*si l'enquêté est enfant de 2 parents immigrés nés dans le même lieu, on lui associe
le lieu de naissance de ses parents*/
replace orig_enq = group1*100 + regionnaisp2 if (group2==31 & regionnaism2!=regionnaisp2)
/*si l'enquêté est enfant de 2 parents immigrés nés dans des lieux différents, on lui associe
- conformément à la convention TeO - le lieu de naissance de son père*/
replace orig_enq = group1*100 + 12 if (group1==4)
/*si l'enquêté est enfant d'un ou 2 natifs d'un DOM, on lui associe
le DOM*/
replace orig_enq = group1*100 + 2 if (group1==5)
/*si l'enquêté est membre du groupe "majoritaire", on lui associe
ce groupe "majoritaire" G2*/
/*tab orig_enq, missing*/
/*Cette variable est bien renseignée pour les 21 761 enquêtés et elle va de 120 (immigré
originaire du Maghreb) à 502 (membre du groupe "majoritaire" G2)*/

*ORIG_PEREENQ (origine "dominante" du père de l'enquêté)
generate orig_pereenq = 100 + regionnaisp2 if (group2==31 | group2==33) /*si l'enquêté est
enfant de 2 parents immigrés ou d'un père immigré uniquement, on associe au père de
l'enquêté son lieu de naissance*/
replace orig_pereenq = 501 if (group2==32) /*si l'enquêté est
enfant d'une mère immigrée uniquement, on associe au père de
l'enquêté le groupe "majoritaire" G1*/
replace orig_pereenq = 212 if (group2==41 | group2==43) /*si l'enquêté est
enfant de 2 parents natifs des DOM ou d'un père né dans un DOM, on associe au père de
l'enquêté le DOM*/
replace orig_pereenq = 501 if (group2==42) /*si l'enquêté est
enfant d'une mère née dans un DOM uniquement, on associe au père de
l'enquêté le groupe "majoritaire" G1*/
replace orig_pereenq = 501 if (group1==5) /*si l'enquêté est
membre du groupe "majoritaire, on associe au père de
l'enquêté le groupe "majoritaire" G1*/
replace orig_pereenq = 100 + regionnaisp2 if (group1==1 | group1==2) /*si l'enquêté est
immigré ou natif d'un DOM, on associe au père de
l'enquêté son lieu de naissance*/
replace orig_pereenq = 501 if (orig_pereenq==111)
replace orig_pereenq = 212 if (orig_pereenq==112)
replace orig_pereenq = 190 if (orig_pereenq==.)
tab orig_pereenq, missing /*Cette variable est
bien renseignée pour les 21 761 enquêtés et elle va de 120 (immigré originaire du Maghreb) à
501 (membre du groupe "majoritaire" G1)*/

*ORIG_MEREENQ (origine "dominante" de la mère de l'enquêté)
generate orig_mereenq = 100 + regionnaism2 if (group2==31 | group2==32) /*si l'enquêté est
enfant de 2 parents immigrés ou d'une mère immigrée uniquement, on associe à la mère de
l'enquêté son lieu de naissance*/
replace orig_mereenq = 501 if (group2==33) /*si l'enquêté est
enfant d'un père immigré uniquement, on associe à la mère de
l'enquêté le groupe "majoritaire" G1*/
replace orig_mereenq = 212 if (group2==41 | group2==42) /*si l'enquêté est
enfant de 2 parents natifs des DOM ou d'une mère née dans un DOM, on associe à la mère de
l'enquêté le DOM*/
replace orig_mereenq = 501 if (group2==43) /*si l'enquêté est
enfant d'un père né dans un DOM uniquement, on associe à la mère de
l'enquêté le groupe "majoritaire" G1*/
replace orig_mereenq = 501 if (group1==5) /*si l'enquêté est
membre du groupe "majoritaire, on associe à la mère de

```

```

l'enquêté le groupe "majoritaire" G1*/
  replace orig_mereenq = 100 + regionnaism2 if (group1==1 | group1==2) /*si l'enquêté est
immigré ou natif d'un DOM, on associe à la mère de
l'enquêté son lieu de naissance*/
  replace orig_mereenq = 501 if (orig_mereenq==111)
  replace orig_mereenq = 212 if (orig_mereenq==112)
  replace orig_mereenq = 190 if (orig_mereenq==.)
  tab orig_mereenq, missing /*Cette variable est
bien renseignée pour les 21 761 enquêtés et elle va de 120 (immigré originaire du Maghreb) à
501 (membre du groupe "majoritaire" G1)*/

*ORIG_FRERESOEURENQ (origine "dominante" du frère ou de la soeur de l'enquêté)
generate orig_freresoeurenq = orig_enq
/*tab orig_freresoeurenq, missing*/ /*Cette variable est
bien renseignée pour les 21 761 enquêtés et elle va de 120 (immigré originaire du Maghreb) à
502 (membre du groupe "majoritaire" G2)*/

*ORIG_CA (origine "dominante" du conjoint actuel de l'enquêté)
generate orig_ca = group1ca*100 + regionnaiscam2 if (group1ca==1)
/*si le conjoint actuel de l'enquêté est immigré,
on lui associe son lieu de naissance*/
  replace orig_ca = group1ca*100 + 12 if (group1ca==2)
/*si le conjoint actuel de l'enquêté est natif d'un DOM,
on lui associe le DOM*/
  replace orig_ca = group1ca*100 + regionnaiscam2 if (group2ca==32)
/*si le conjoint actuel de l'enquêté est enfant d'1 mère immigrée uniquement,
on lui associe le lieu de naissance de sa mère*/
  replace orig_ca = group1ca*100 + regionnaiscap2 if (group2ca==33)
/*si le conjoint actuel de l'enquêté est enfant d'1 père immigré uniquement,
on lui associe le lieu de naissance de son père*/
  replace orig_ca = group1ca*100 + regionnaiscam2 if (group2ca==31 &
regionnaism2==regionnaisp2) /*si le conjoint actuel de l'enquêté est enfant de 2 parents
immigrés nés dans le même lieu, on lui associe le lieu de naissance de ses parents*/
  replace orig_ca = group1ca*100 + regionnaiscap2 if (group2ca==31 &
regionnaism2!=regionnaisp2) /*si le conjoint actuel de l'enquêté est enfant de 2 parents
immigrés nés dans des lieux différents, on lui associe - conformément à la convention TeO - le
lieu de naissance de son père*/
  replace orig_ca = group1ca*100 + 12 if (group1ca==4)
/*si le conjoint actuel de l'enquêté est enfant d'1 ou 2 natifs d'un DOM,
on lui associe le DOM*/
  replace orig_ca = group1ca*100 + 2 if (group1ca==5)
/*si le conjoint actuel de l'enquêté est membre du groupe "majoritaire",
on lui associe ce groupe "majoritaire" G2*/
  tab orig_ca, missing
/*Cette variable contient 5 989 "." (non réponse) et elle va de 120 (immigré originaire du
Maghreb) à 502 (membre du groupe "majoritaire" G2)*/

*ORIG_PERECA (origine "dominante" du père du conjoint actuel de l'enquêté)
generate orig_pereca = 100 + regionnaiscap2 if (group2ca==31 | group2ca==33) /*si le conjoint
actuel de l'enquêté est enfant de 2 parents immigrés ou d'un père immigré uniquement, on
associe au père du conjoint actuel de l'enquêté son lieu de naissance*/
  replace orig_pereca = 501 if (group2ca==32) /*si le conjoint
actuel de l'enquêté est enfant d'une mère immigrée uniquement, on
associe au père du conjoint actuel de l'enquêté le groupe "majoritaire" G1*/
  replace orig_pereca = 212 if (group2ca==41 | group2ca==43) /*si le conjoint
actuel de l'enquêté est enfant de 2 parents natifs des DOM ou d'un père né dans un DOM, on
associe au père du conjoint actuel de l'enquêté le DOM*/
  replace orig_pereca = 501 if (group2ca==42) /*si le conjoint
actuel de l'enquêté est enfant d'une mère née dans un DOM uniquement, on
associe au père du conjoint actuel de l'enquêté le groupe "majoritaire" G1*/
  replace orig_pereca = 501 if (group1ca==5) /*si le conjoint
actuel de l'enquêté est membre du groupe "majoritaire, on
associe au père du conjoint actuel de l'enquêté le groupe "majoritaire" G1*/
  replace orig_pereca = 100 + regionnaiscap2 if (group1ca==1 | group1ca==2) /*si le conjoint
actuel de l'enquêté est immigré ou natif d'un DOM, on
associe au père du conjoint actuel de l'enquêté son lieu de naissance*/
  replace orig_pereca = 501 if (orig_pereca==111)
  replace orig_pereca = 212 if (orig_pereca==112)
  tab orig_pereca, missing /*Cette variable
contient 11 296 "." (non réponse) et elle va de 120 (immigré originaire du Maghreb) à 501
(membre du groupe "majoritaire" G1)*/

*ORIG_MERECA (origine "dominante" de la mère du conjoint actuel de l'enquêté)
generate orig_mereca = 100 + regionnaiscam2 if (group2ca==31 | group2ca==32) /*si le conjoint
actuel de l'enquêté est enfant de 2 parents immigrés ou d'une mère immigrée uniquement, on
associe à la mère du conjoint actuel de l'enquêté son lieu de naissance*/
  replace orig_mereca = 501 if (group2ca==33) /*si le conjoint

```



```

actuel de l'enquêté est enfant d'un père immigré uniquement, on
associe à la mère du conjoint actuel de l'enquêté le groupe "majoritaire" G1*/
  replace orig_mereca = 212 if (group2ca==41 | group2ca==42) /*si le conjoint
actuel de l'enquêté est enfant de 2 parents natifs des DOM ou d'une mère née dans un DOM, on
associe à la mère du conjoint actuel de l'enquêté le DOM*/
  replace orig_mereca = 501 if (group2ca==43) /*si le conjoint
actuel de l'enquêté est enfant d'un père né dans un DOM uniquement, on
associe à la mère du conjoint actuel de l'enquêté le groupe "majoritaire" G1*/
  replace orig_mereca = 501 if (group1ca==5) /*si le conjoint
actuel de l'enquêté est membre du groupe "majoritaire", on
associe à la mère du conjoint actuel de l'enquêté le groupe "majoritaire" G1*/
  replace orig_mereca = 100 + regionnaiscam2 if (group1ca==1 | group1ca==2) /*si le conjoint
actuel de l'enquêté est immigré ou natif d'un DOM, on
associe à la mère du conjoint actuel de l'enquêté son lieu de naissance*/
  replace orig_mereca = 501 if (orig_mereca==111)
  replace orig_mereca = 212 if (orig_mereca==112)
  tab orig_mereca, missing /*Cette variable
contient 11 297 "." (non réponse) et elle va de 120 (immigrée originaire du Maghreb) à 501
(membre du groupe "majoritaire" G1)*/

*ORIG_ENFCA (origine "dominante" de l'enfant du conjoint actuel de l'enquêté)
*On ne connaît les origines que d'1 des 2 parents de cet enfant, mais on lui attribue quand
même une origine.
generate orig_enfca = orig_ca + 200 if (orig_ca>=120 & orig_ca<=190) /*si le conjoint actuel
de l'enquêté est immigré et d'une origine donnée, on catégorise son enfant comme enfant
d'immigrés de la même origine*/
  replace orig_enfca = orig_ca + 200 if (orig_ca==212) /*si le conjoint actuel
de l'enquêté est natif d'un DOM, on catégorise son enfant comme enfant
de natifs d'un DOM*/
  replace orig_enfca = orig_ca + 300 if (orig_ca>=320 & orig_ca<=390) /*si le conjoint actuel
de l'enquêté est enfant d'immigré d'une origine donnée, on catégorise son enfant comme petit-
enfant d'immigrés de la même origine*/
  replace orig_enfca = orig_ca + 300 if (orig_ca==412) /*si le conjoint actuel
de l'enquêté est enfant d'1 ou 2 natifs d'un DOM, on catégorise son enfant comme petit-
enfant de natifs d'un DOM*/
  replace orig_enfca = 503 if (orig_ca==502) /*si le conjoint actuel
de l'enquêté est membre du groupe "majoritaire" G2, on catégorise son enfant comme membre
du groupe "majoritaire" G3*/
  tab orig_enfca, missing /*Cette variable contient
5 989 "." (non réponse) et elle va de 320 (enfant d'immigrés originaire du Maghreb) à 712
(petit-enfant de natifs d'un DOM)*/

*ORIG_PC (origine "dominante" du 1er conjoint de l'enquêté)
generate orig_pc = group1pc*100 + regionnaispc2 if (group1pc==1)
/*si le 1er conjoint de l'enquêté est immigré,
on lui associe son lieu de naissance*/
  replace orig_pc = group1pc*100 + 12 if (group1pc==2)
/*si le 1er conjoint de l'enquêté est natif d'un DOM,
on lui associe le DOM*/
  replace orig_pc = group1pc*100 + regionnaispcm2 if (group2pc==32)
/*si le 1er conjoint de l'enquêté est enfant d'1 mère immigrée uniquement,
on lui associe le lieu de naissance de sa mère*/
  replace orig_pc = group1pc*100 + regionnaispcp2 if (group2pc==33)
/*si le 1er conjoint de l'enquêté est enfant d'1 père immigré uniquement,
on lui associe le lieu de naissance de son père*/
  replace orig_pc = group1pc*100 + regionnaispcm2 if (group2pc==31 &
regionnaispcm2==regionnaispcp2) /*si le 1er conjoint de l'enquêté est enfant de 2 parents
immigrés nés dans le même lieu, on lui associe le lieu de naissance de ses parents*/
  replace orig_pc = group1pc*100 + regionnaispcp2 if (group2pc==31 &
regionnaispcm2!=regionnaispcp2) /*si le 1er conjoint de l'enquêté est enfant de 2 parents
immigrés nés dans des lieux différents, on lui associe - conformément à la convention TeO - le
lieu de naissance de son père*/
  replace orig_pc = group1pc*100 + 12 if (group1pc==4)
/*si le 1er conjoint de l'enquêté est enfant d'1 ou 2 natifs d'un DOM,
on lui associe le DOM*/
  replace orig_pc = 502 if (group1pc==5)
/*si le 1er conjoint de l'enquêté est membre du groupe "majoritaire",
on lui associe ce groupe "majoritaire" G2*/
  /*tab orig_pc, missing*/
/*Cette variable contient 16 789 "." (non réponse) et elle va de 120 (immigré originaire du
Maghreb) à 502 (membre du groupe "majoritaire" G2)*/

*ORIG_PEREPC (origine "dominante" du père du 1er conjoint de l'enquêté)
generate orig_perepc = 100 + regionnaispcp2 if (group2pc==31 | group2pc==33) /*si le 1er
conjoint de l'enquêté est enfant de 2 parents immigrés ou d'un père immigré uniquement, on
associe au père du 1er conjoint de l'enquêté son lieu de naissance*/
  replace orig_perepc = 501 if (group2pc==32) /*si le 1er

```

```

conjoint de l'enquêté est enfant d'une mère immigrée uniquement, on
associe au père du 1er conjoint de l'enquêté le groupe "majoritaire" G1*/
  replace orig_perepc = 212 if (group2pc==41 | group2pc==43) /*si le 1er
conjoint de l'enquêté est enfant de 2 parents natifs des DOM ou d'un père né dans un DOM, on
associe au père du 1er conjoint de l'enquêté le DOM*/
  replace orig_perepc = 501 if (group2pc==42) /*si le 1er
conjoint de l'enquêté est enfant d'une mère née dans un DOM uniquement, on
associe au père du 1er conjoint de l'enquêté le groupe "majoritaire" G1*/
  replace orig_perepc = 501 if (group1pc==5) /*si le 1er
conjoint de l'enquêté est membre du groupe "majoritaire", on
associe au père du 1er conjoint de l'enquêté le groupe "majoritaire" G1*/
  replace orig_perepc = 100 + regionnaispcp2 if (group1pc==1 | group1pc==2) /*si le 1er
conjoint de l'enquêté est immigré ou natif d'un DOM, on
associe au père du 1er conjoint de l'enquêté son lieu de naissance*/
  replace orig_perepc = 501 if (orig_perepc==111)
  replace orig_perepc = 212 if (orig_perepc==112)
  tab orig_perepc, missing /*Cette variable
contient 18 293 "." (non réponse) et elle va de 120 (immigré originaire du Maghreb) à 501
(membre du groupe "majoritaire" G1)*/

*ORIG_MEREPC (origine "dominante" de la mère du 1er conjoint de l'enquêté)
generate orig_merepc = 100 + regionnaispcm2 if (group2pc==31 | group2pc==32) /*si le 1er
conjoint de l'enquêté est enfant de 2 parents immigrés ou d'une mère immigrée uniquement, on
associe à la mère du 1er conjoint de l'enquêté son lieu de naissance*/
  replace orig_merepc = 501 if (group2pc==33) /*si le 1er
conjoint de l'enquêté est enfant d'un père immigré uniquement, on
associe à la mère du 1er conjoint de l'enquêté le groupe "majoritaire" G1*/
  replace orig_merepc = 212 if (group2pc==41 | group2pc==42) /*si le 1er
conjoint de l'enquêté est enfant de 2 parents natifs des DOM ou d'une mère née dans un DOM, on
associe à la mère du 1er conjoint de l'enquêté le DOM*/
  replace orig_merepc = 501 if (group2pc==43) /*si le 1er
conjoint de l'enquêté est enfant d'un père né dans un DOM uniquement, on
associe à la mère du 1er conjoint de l'enquêté le groupe "majoritaire" G1*/
  replace orig_merepc = 501 if (group1pc==5) /*si le 1er
conjoint de l'enquêté est membre du groupe "majoritaire", on
associe à la mère du 1er conjoint de l'enquêté le groupe "majoritaire" G1*/
  replace orig_merepc = 100 + regionnaispcm2 if (group1pc==1 | group1pc==2) /*si le 1er
conjoint de l'enquêté est immigré ou natif d'un DOM, on
associe à la mère du 1er conjoint de l'enquêté son lieu de naissance*/
  replace orig_merepc = 501 if (orig_merepc==111)
  replace orig_merepc = 212 if (orig_merepc==112)
  tab orig_merepc, missing /*Cette variable
contient 18 293 "." (non réponse) et elle va de 120 (immigré originaire du Maghreb) à 501
(membre du groupe "majoritaire" G1)*/

*ORIG_GRDPARENTPATERNELENQ (origine "dominante" des grands-parents paternels de l'enquêté)
*Ce codage est grossier, mais on leur attribue quand même une origine
generate orig_grdparentpaternelenq = orig_pereenq if (orig_pereenq>=120 & orig_pereenq<=190)
/*si le père de l'enquêté est immigré, on admet qu'il en va de même
pour les grands-parents paternels de l'enquêté (qui auraient migré suite à leur fils) et on
leur associe le lieu de naissance de leur fils*/
  replace orig_grdparentpaternelenq = orig_pereenq if (orig_pereenq==212)
/*si le père de l'enquêté est natif d'un DOM, on admet qu'il en va de même
pour les grands-parents paternels de l'enquêté et on
leur associe le DOM*/
  replace orig_grdparentpaternelenq = 500 if (orig_pereenq==501)
/*si le père de l'enquêté est membre du groupe "majoritaire", on admet qu'il en va de même
pour les grands-parents paternels de l'enquêté et on
les catégorise comme membres du groupe "majoritaire" G0*/
  tab orig_grdparentpaternelenq, missing
/*Cette variable est bien renseignée pour les 21 761 enquêtés et elle va de 120 (immigré
originaire du Maghreb) à 500 (membre du groupe "majoritaire" G0)*/

*ORIG_GRDPARENTMATERNELENQ (origine "dominante" des grands-parents maternels de l'enquêté)
*Ce codage est grossier, mais on leur attribue quand même une origine
generate orig_grdparentmaternelenq = orig_mereenq if (orig_mereenq>=120 & orig_mereenq<=190)
/*si la mère de l'enquêté est immigrée, on admet qu'il en va de même
pour les grands-parents maternels de l'enquêté (qui auraient migré suite à leur fille) et on
leur associe le lieu de naissance de leur fille*/
  replace orig_grdparentmaternelenq = orig_mereenq if (orig_mereenq==212)
/*si la mère de l'enquêté est native d'un DOM, on admet qu'il en va de même
pour les grands-parents maternels de l'enquêté et on
leur associe le DOM*/
  replace orig_grdparentmaternelenq = 500 if (orig_mereenq==501)
/*si la mère de l'enquêté est membre du groupe "majoritaire", on admet qu'il en va de même
pour les grands-parents maternels de l'enquêté et on
les catégorise comme membres du groupe "majoritaire" G0*/

```

```

    tab orig_grdparentmaternelenq, missing
/*Cette variable est bien renseignée pour les 21 761 enquêtés et elle va de 120 (immigré
originaire du Maghreb) à 500 (membre du groupe "majoritaire" G0)*/

*On ne peut malheureusement pas attribuer d'origine aux habitants du logement qui ont avec
l'enquêté d'"autres" liens familiaux (LIENEGO=40) ou des liens familiaux "indéterminés"
(LIENEGO=50) ou des liens d'amitié (LIENEGO=60) ou d'autres liens non familiaux (LIENEGO=90)

*4 VARIABLES de langues parlées par les parents de l'enquêté dans son enfance

*LNP1_GR (1e langue parlée par le père de l'enquêté dans son enfance)
tab lnp1_gr, missing /*Cette variable est renseignée pour 21 664 enquêtés et elle va de 10
(français) à 90 (NSP)*/

*LNP2_GR (2e langue parlée par le père de l'enquêté dans son enfance)
tab lnp2_gr, missing /*Cette variable est renseignée pour 6 583 enquêtés et elle va de 10
(français) à 50 (autres langues)*/

*LNM1_GR (1e langue parlée par la mère de l'enquêté dans son enfance)
tab lnm1_gr, missing /*Cette variable est renseignée pour 21 752 enquêtés et elle va de 10
(français) à 90 (NSP)*/

*LNM2_GR (2e langue parlée par la mère de l'enquêté dans son enfance)
tab lnm2_gr, missing /*Cette variable est renseignée pour 6 867 enquêtés et elle va de 10
(français) à 50 (autres langues)*/

*R_IMPVIIE (importance qu'accorde aujourd'hui l'enquêté à la religion dans sa vie)
tab r_impvie, missing /*Cette variable n'est renseignée que pour les 15 239 enquêtés (sur 21
761) qui ont déclaré avoir aujourd'hui une religion, et elle va de 1 à 9.*/

*X_VUFRI (opinion de l'enquêté sur: "On me voit comme un Français")
tab x_vufri, missing /*Cette variable est bien renseignée pour les 21 761 enquêtés et va de 1
à 9.*/

*J'enregistre
save indiv2_prenom, replace

*Les 11 variables (dont 2 identifiants) issues de la table ENFANTS2_PRENOM
use enfants2_prenom

*IDENT (identifiant de l'enquêté, pour chaque enfant de l'enquêté)
summarize ident /*Cette variable est bien renseignée pour les 31 476 enfants
et va de 3 à 21 760.*/

*NOI (numéro d'ordre de l'enfant de l'enquêté dans le logement, pour chaque enfant cohabitant
de l'enquêté)
*Cette variable nous permettra ci-dessous d'apparier chaque enfant de l'enquêté de la table
ENFANTS2_PRENOM à chaque enfant de l'enquêté de la table THL2_PRENOM
sort ident e_noi_thl e_vie e_sexe e_anais e_nat e_lvmet e_situa e_diplo
tab e_noi_thl, missing /*Cette variable n'est pas renseignée pour les 7 924 enfants
non cohabitants ou décédés, et elle va de 1 à 13*/
generate noi = e_noi_thl
tab noi, missing
generate noi_noncohab = 100 + _n /*Pour les 7 924 enfants non cohabitants ou décédés et qui
logiquement n'ont pas de NOI, on attribue un NOI arbitraire (égal à 100 + le numéro de leur
ligne dans la table ENFANTS2_PRENOM) afin de pouvoir les identifier uniquement, sans quoi on
ne pourrait pas ajouter ci-dessous ces enfants non cohabitants de la table ENFANTS2_PRENOM aux
habitants du logement de l'enquêté de la table THL2_PRENOM*/
tostring noi_noncohab, replace
replace noi = noi_noncohab if (noi=="")
sort ident e_noi_thl e_anais e_sexe

*POIDSE (pondération de l'enfant de l'enquêté)
tab e_sexe[aw=poidse], missing /*La somme des enfants pondérés est de 31 476 individus,
tout comme la somme des enfants non pondérés*/

*E_SEXE (sexe de l'enfant de l'enquêté)
tab e_sexe, missing /*Cette variable n'est pas renseignée pour 10 enfants, dont
on ignore le sexe; sinon elle va de 1 Homme à 2 Femme.*/

*E_ANAIS (année de naissance de l'enfant de l'enquêté)
/*tab e_anais, missing*/ /*Cette variable compte 84 non réponses; sinon elle va de
1962 à 2009.*/

*REGIONNAISENF2 (lieu de naissance de l'enfant de l'enquêté)
/*tab regionnaisenf2, missing*/ /*Cette variable est bien renseignée pour les 31 476 enfants

```

```

et va de 1101 à 9999*/
destring regionnaisenf2, replace
  recode regionnaisenf2 (1101=11/*11=France métro*/) (1201/1205=12/*12=DOM*/)
(2101/2301=20/*20=Maghreb*/) (2401/2601=29/*29=autre Afrique*/) (3101/3301=30/*30=Asie*/)
(3401=34/*34=Turquie*/) (3501=30/*30=Asie*/) (4101/4301=40/*40=Europe du Sud*/)
(4401/4801=49/*49=autre Europe*/) (5102/5107=50/*50=Ailleurs dans le monde*/)
(9999=90/*90=NSP*/) (.=/*.=NR*/)
  tab regionnaisenf2, missing

*E_PRENOM (prénom de l'enfant de l'enquêté)
codebook e_prenom /*Cete variable est renseignée pour les 31 476 enfants: ils
portent 8 844 prénoms différents*/

*E_VIE (est-ce que l'enfant de l'enquêté est toujours en vie)
/*tab e_vie, missing*/ /*Cette variable compte 31 142 réponses positives, et
plusieurs types de réponses négatives (non, refus, nsp) ou sans objet*/

*E_LVIE (lieu où vit l'enfant)
tab e_lvie, missing /*Cette variable compte 334 non réponses et va de 1 à 9*/

*ORIG_ENFENQ (origine "dominante" de l'enfant de l'enquêté, y compris non cohabitant et
décédé)
*Cette variable a le même codage à 3 chiffres que les autres variables commençant par ORIG_

merge m:1 ident using indiv2_prenom.dta, keepusing(orig_enq orig_ca orig_pc c_cohac c_cohap
c_finap sexee) keep(match)
  tab orig_enq, missing
  tab orig_ca, missing
/*tab orig_pc, missing*/
  tab c_cohac, missing /*année du début de cohabitation avec le conjoint actuel*/
/*tab c_cohap, missing*/ /*année du début de cohabitation avec le 1er conjoint*/
  tab c_finap, missing /*année de la fin de la relation avec le 1er conjoint*/
  tab sexee, missing

generate orig_autreparent = orig_ca if (e_anais>=1962 & e_anais<=2009 & e_anais>=c_cohac &
c_cohac!=.)
  replace orig_autreparent = orig_pc if (e_anais>=1962 & e_anais<=2009 & e_anais>=c_cohap &
e_anais<=c_finap & c_cohap!=. & c_finap!=.)
  /*tab orig_autreparent, missing*/ /*Cette variable contient 3 787 "." (non réponse) sur
les 31 476 enfants, et elle va de 120 (immigré originaire du Maghreb) à 500 (membre du groupe
"majoritaire")*/

generate orig_enfeng = orig_enq + 300 if (orig_enq>=320 & orig_enq<=390) &
(orig_autreparent<320 | orig_autreparent>390 | orig_autreparent==.)
/*si l'enfant
n'a qu'1 parent enfant d'immigrés (l'enquêté),
on le catégorise comme petit-enfant d'immigrés de même
origine que son parent enfant d'immigrés (l'enquêté)*/
  replace orig_enfeng = orig_autreparent + 300 if (orig_autreparent>=320 &
orig_autreparent<=390) & (orig_enq<320 | orig_enq>390 | orig_enq==.)
/*si l'enfant
n'a qu'1 parent enfant d'immigrés (pas l'enquêté, l'autre parent),
on le catégorise comme petit-enfant d'immigrés de même
origine que son parent enfant d'immigrés (pas l'enquêté, l'autre parent)*/
  replace orig_enfeng = orig_enq + 300 if (orig_enq>=320 & orig_enq<=390) &
(orig_autreparent>=320 & orig_autreparent<=390) & (orig_enq==orig_autreparent)
/*si l'enfant
a 2 parents tous deux enfant d'immigrés du même lieu,
on le catégorise comme petit-enfant d'immigrés de même
origine que ses parents*/
  replace orig_enfeng = orig_enq + 300 if (orig_enq>=320 & orig_enq<=390) &
(orig_autreparent>=320 & orig_autreparent<=390) & (orig_enq!=orig_autreparent) & (sexee=="1")
/*si l'enfant
a 2 parents tous deux enfant d'immigrés mais pas du même lieu,
on le catégorise - conformément à la convention TeO - comme petit-enfant d'immigrés originaire
du lieu de naissance de son père*/
  replace orig_enfeng = orig_autreparent + 300 if (orig_enq>=320 & orig_enq<=390) &
(orig_autreparent>=320 & orig_autreparent<=390) & (orig_enq!=orig_autreparent) & (sexee=="2")
/*si l'enfant
a 2 parents tous deux enfant d'immigrés mais pas du même lieu,
on le catégorise - conformément à la convention TeO - comme petit-enfant d'immigrés originaire
du lieu de naissance de son père*/
  replace orig_enfeng = orig_enq + 200 if (orig_enq>=120 & orig_enq<=150) &
(orig_autreparent<120 | orig_autreparent>150 | orig_autreparent==.)
/*si l'enfant (sans aucun parent enfant d'immigrés)
n'a qu'1 parent immigré (l'enquêté),
on le catégorise comme enfant d'immigrés de même origine

```

```

que son parent immigré (l'enquêté)*/
  replace orig_enfeng = orig_autreparent + 200 if (orig_autreparent>=120 &
orig_autreparent<=150) & (orig_enq<120 | orig_enq>150 | orig_enq==.)
/*si l'enfant (sans aucun parent enfant d'immigrés)
n'a qu'1 parent immigré (pas l'enquêté, l'autre parent),
on le catégorise                                     comme enfant d'immigrés de même origine
que son parent immigré (pas l'enquêté, l'autre parent)*/
  replace orig_enfeng = orig_enq + 200 if (orig_enq>=120 & orig_enq<=150) &
(orig_autreparent>=120 & orig_autreparent<=150) & (orig_enq==orig_autreparent)
/*si l'enfant (sans aucun parent enfant d'immigrés)
a 2 parents tous deux immigrés du même lieu,
on le catégorise                                     comme enfant d'immigrés de même origine
que ses parents*/
  replace orig_enfeng = orig_enq + 200 if (orig_enq>=120 & orig_enq<=150) &
(orig_autreparent>=120 & orig_autreparent<=150) & (orig_enq!=orig_autreparent) & (sexee=="1")
/*si l'enfant (sans aucun parent enfant d'immigrés)
a 2 parents tous deux immigrés mais pas du même lieu,
on le catégorise - conformément à la convention TeO - comme enfant d'immigrés originaire du
lieu de naissance de son père*/
  replace orig_enfeng = orig_autreparent + 200 if (orig_enq>=120 & orig_enq<=150) &
(orig_autreparent>=120 & orig_autreparent<=150) & (orig_enq!=orig_autreparent) & (sexee=="2")
/*si l'enfant (sans aucun parent enfant d'immigrés)
a 2 parents tous deux immigrés mais pas du même lieu,
on le catégorise - conformément à la convention TeO - comme enfant d'immigrés originaire du
lieu de naissance de son père*/
  replace orig_enfeng = 712 if (orig_enq==412 | orig_autreparent==412) &
(orig_enfeng==.)
/*si l'enfant (sans aucun parent enfant d'immigrés ni aucun parent immigré)
a 1 ou 2 parents enfant de natifs d'un DOM,
on le catégorise                                     comme petit-enfant de natifs d'un DOM*/
  replace orig_enfeng = 412 if (orig_enq==212 | orig_autreparent==212) &
(orig_enfeng==.)
/*si l'enfant (sans aucun parent enfant d'immigrés ni aucun parent immigré ni aucun parent
enfant de natif d'un DOM) a 1 ou 2 parents natifs d'un DOM,
on le catégorise                                     comme enfant de natifs d'un DOM*/
  replace orig_enfeng = 503 if ((orig_enq==502 & orig_autreparent==502) |
(orig_enq==. & orig_autreparent==502) | (orig_enq==502 & orig_autreparent==.)) &
(orig_enfeng==.) /*si l'enfant (sans aucun parent enfant d'immigrés ni aucun parent immigré ni
aucun parent enfant de natif d'un DOM ni aucun parent natif d'un DOM) a 2 parents membres du
groupe "majoritaire" G2, ou 1 parent du groupe "majoritaire" G2 et l'autre sans réponse, on le
catégorise                                     comme membre du groupe "majoritaire" G3*/
  tab orig_enfeng, missing
/*Cette variable est bien renseignée pour les 31 476 enfants et elle va de 320 (enfant d'1 ou
2 immigrés du Maghreb) à 712 (petit-enfant d'1 à 4 natifs d'un DOM*/

*AUTREPARENT (outre l'enquêté, est-ce que l'autre parent de l'enfant est le conjoint actuel de
l'enquêté [autreparent=1], ou son 1er conjoint [autreparent=0])
generate autreparent = 1 if (e_anais>=1962 & e_anais<=2009 & e_anais>=c_cohac & c_cohac!=.)
  replace autreparent = 0 if (e_anais>=1962 & e_anais<=2009 & e_anais>=c_cohap &
e_anais<=c_finap & c_cohap!=. & c_finap!=.)
  tab autreparent, missing /*Cette variable contient 3 784 "." (non réponse) sur les 31 476
enfants, et elle va de 0 (l'autre parent est le 1er [ex-] conjoint de l'enquêté) à 1 (l'autre
parent est le conjoint actuel de l'enquêté)*/

*J'enregistre
save enfants2_prenom, replace

*Les 6 variables (dont 1 identifiant) issues de la table POLREL2
use polrel2

*IDENT (identifiant de l'enquêté)
summarize ident /*Cette variable est bien renseignée pour les 21 761 enquêtés et va
de 1 à 21 761.*/

*5 VARIABLES d'affiliation religieuse

*RELEGO1 (religion de l'enquêté)
tab relego1, missing /*Cette variable est bien renseignée pour les 21 761 enquêtés et va
de 1 à 99.*/
destring relego1, replace
  recode relego1 (01=01/*1=sans religion*/) (10/14=10/*10=chrétiens*/)
(20/21=20/*20=musulmans*/) (30=30/*30=juifs*/) (40/44=40/*40=bouddhistes, hindouistes, autres,
multi-religions*/) (88/99=88/*88=Refus de répondre, NSP*/)
  tab relego1, missing /*Cette variable est bien renseignée pour les 21 761 enquêtés (y
compris 318 Refus de répondre ou NSP)*/

```

```

*RELCA1 (religion du conjoint actuel de l'enquêté)
tab relcal, missing /*Cette variable est renseignée pour seulement 15 788 enquêtés et
va de 1 à .*/
destring relcal, replace
  recode relcal (01=01/*1=sans religion*/) (10/14=10/*10=chrétiens*/)
(20/21=20/*20=musulmans*/) (30=30/*30=juifs*/) (40/44=40/*40=bouddhistes, hindouistes, autres,
multi-religions*/) (88/99=88/*88=Refus de répondre, NSP*/)
  tab relcal, missing /*Cette variable est bien renseignée pour 15 788 enquêtés (y
compris 560 Refus de répondre ou NSP)*/

*RELPC1 (religion du 1er conjoint de l'enquêté)
/*tab relpcl, missing*/ /*Cette variable est renseignée pour seulement 5 048 enquêtés et va
de 1 à 99.*/
destring relpcl, replace
  recode relpcl (01=01/*1=sans religion*/) (10/14=10/*10=chrétiens*/)
(20/21=20/*20=musulmans*/) (30=30/*30=juifs*/) (40/44=40/*40=bouddhistes, hindouistes, autres,
multi-religions*/) (88/99=88/*88=Refus de répondre, NSP*/)
  tab relpcl, missing /*Cette variable est renseignée pour seulement 5 048 enquêtés (y
compris 526 Refus de répondre ou NSP)*/

*RELPERE1 (religion du père de l'enquêté)
tab relperel, missing /*Cette variable est renseignée pour 21 664 enquêtés et va de 1 à
99.*/
destring relperel, replace
  recode relperel (01=01/*1=sans religion*/) (10/14=10/*10=chrétiens*/)
(20/21=20/*20=musulmans*/) (30=30/*30=juifs*/) (40/44=40/*40=bouddhistes, hindouistes, autres,
multi-religions*/) (88/99=88/*88=Refus de répondre, NSP*/)
  tab relperel, missing /*Cette variable est renseignée pour 21 664 enquêtés (y compris 887
Refus de répondre ou NSP)*/

*RELMERE1 (religion de la mère de l'enquêté)
tab relmerel, missing /*Cette variable est renseignée pour 21 752 enquêtés et va de 1 à
99.*/
destring relmerel, replace
  recode relmerel (01=01/*1=sans religion*/) (10/14=10/*10=chrétiens*/)
(20/21=20/*20=musulmans*/) (30=30/*30=juifs*/) (40/44=40/*40=bouddhistes, hindouistes, autres,
multi-religions*/) (88/99=88/*88=Refus de répondre, NSP*/)
  tab relmerel, missing /*Cette variable est renseignée pour 21 752 enquêtés (y compris 511
Refus de répondre ou NSP)*/

/*Les 5 VARIABLES d'affiliation religieuse suivantes sont codées comme suit:
1=sans religion
10=chrétien
20=musulman
30=juif
40=bouddhiste, hindouiste, autre, multi-religions
88=Refus de répondre, NSP
*/

*J'enregistre
save polrel2, replace

/*Chapitre 5: La construction de la table REPRO*/

*Je pars des 75 121 habitants du logement et des (8-2-6) variables de la table THL2_PRENOM
(poidsi, sexe, anais, regionnais2, prenom et lienego)
use thl2_prenom

*J'y ajoute les 7 924 enfants des enquêtés de la table ENFANTS2_PRENOM

*J'y ajoute les (11-2=9) variables issues de la table ENFANTS2_PRENOM, appariées aux seuls
enfants des enquêtés de la table THL2_PRENOM (poidse, e_sexe, e_anais, regionnaisenf2, e_vie,
e_lvie, e_prenom, orig_enfeng et autreparent)
merge 1:1 ident noi using enfants2_prenom, keepusing(poidse e_sexe e_anais regionnaisenf2
e_vie e_lvie e_prenom orig_enfeng autreparent)
/*Je vérifie que dans la table THL2_PRENOM j'ai bien
- 7 924 lignes de plus (de la ligne 75122 à la ligne 83045), avec des IDENT divers et tous
les NOI > 100 et seules 8 autres variables renseignées: c'est bon
- 9 colonnes de plus, qui sont renseignées d'une part pour les 23 552 enfants cohabitants
(LIENEGO=02), d'autre part pour les 7 924 enfants non cohabitants (LIENEGO=non renseigné),
mais pour aucun autre individu: c'est bon*/
drop _merge

*J'y ajoute ensuite les (41-2=39) variables issues de la table INDIV2_PRENOM

*Pour cela j'apparie tout d'abord 6 variables de la table INDIV2_PRENOM aux seuls enquêtés de

```

```

la table THL2_PRENOM (poids, sexee, anaise, prenome, r_impvie et x_vufri)
merge 1:1 ident noi using indiv2_prenom, keepusing(poids sexee anaise prenome r_impvie
x_vufri)
/*Je vérifie que dans la table THL2_PRENOM j'ai bien
- 0 ligne de plus: c'est bon
- 6 colonnes de plus, renseignées pour les 21 761 enquêtés (LIENEGO=00) mais pour aucun autre
individu: c'est bon*/
drop _merge

*Pour cela j'apparie ensuite les 33 autres variables de la table INDIV2_PRENOM à tous les
habitants du logement de la table THL2_PRENOM (regionnaise2, regionnaisp2, regionnaism2,
regionnaisca2, regionnaiscap2, regionnaiscam2, regionnaisp2, regionnaisp2, regionnaisp2,
group2, group2ca, group2pc, orig_enq, orig_pereenq, orig_mereenq, orig_freresoeurenq, orig_ca,
orig_pereca, orig_mereca, orig_enfca, orig_pc, orig_perepc, orig_merepc,
orig_grdparentpaternelenq, orig_grdparentmaternelenq, inat, natca, natpc, natpniv2, natmniv2,
n_indper, n_indmer, lnp1_gr, lnp2_gr, lnm1_gr et lnm2_gr)
merge m:1 ident using indiv2_prenom, keepusing(regionnaise2 regionnaisp2 regionnaism2
regionnaisca2 regionnaiscap2 regionnaiscam2 regionnaisp2 regionnaisp2 regionnaisp2 group2
group2ca group2pc orig_enq orig_pereenq orig_mereenq orig_freresoeurenq orig_ca orig_pereca
orig_mereca orig_enfca orig_pc orig_perepc orig_merepc orig_grdparentpaternelenq
orig_grdparentmaternelenq natpniv2 natmniv2 n_indper n_indmer lnp1_gr lnp2_gr lnm1_gr lnm2_gr)
/*Je vérifie que dans la table THL2_PRENOM j'ai bien
- 0 ligne de plus: c'est bon
- 33 colonnes de plus, renseignées pour les 83 045 individus: c'est bon*/
drop _merge

*J'y ajoute enfin les (6-1=5) variables issue de la table POLREL2, appariées à tous les
habitants du logement de la table THL2_PRENOM (relegol, relcal, relpcl, relperel et relmerel)
merge m:1 ident using polrel2, keepusing(relegol relcal relpcl relperel relmerel)
/*Je vérifie que dans la table THL2_PRENOM j'ai bien
- 0 ligne de plus: c'est bon
- 5 colonnes de plus, renseignées pour les 83 045 individus: c'est bon*/
drop _merge

*J'obtiens ainsi la table REPRO, que j'enregistre
save repro, replace

/*Chapitre 5: Le recodage des variables de la table REPRO*/
use repro
sort ident noi

*LIENEGO_IND (lien de l'individu avec l'enquêté)
tab lienego, missing /*Cette variable est bien renseignée pour les 83 045 individus, sauf
les 7 924 enfants non cohabitants ou décédés à la date de l'enquête*/
generate lienego_ind = lienego
replace lienego_ind = "201" if (lienego_ind=="" & e_vie=="1") /*201=enfant non cohabitant en
vie*/
replace lienego_ind = "200" if (lienego_ind=="" & e_vie!="1") /*200=enfant non cohabitant
décédé*/
tab lienego_ind, missing /*Cette variable est bien renseignée pour les 83 045
individus et va de 00 à 90*/
/*LIENEGO_IND est codée comme suit:
00= enquêté
01= conjoint
02= enfant cohabitant
200=enfant non cohabitant décédé
201=enfant non cohabitant en vie
03= parent
10= frère, soeur
21= petit-enfant
22= grand-parent
31= beau-fils, belle-fille
32= beau-parent
40= autre lien familial
50= lien familial indéterminé
60= ami
90= autre lien non familial*/

*POIDS_IND (pondération de l'individu)
summarize poids /*Cette variable est bien renseignée pour les 75 121
habitants du logement de l'enquêté*/
summarize poidse /*Cette variable est bien renseignée pour les 31 476
enfants des enquêtés, y compris non cohabitants ou décédés à la date de l'enquête*/
*Je vérifie qu'il n'y a pas d'incohérences entre ces pondérations
generate incoherencepoids = 1 if (poids!=. & poidse!=. & poids!=2*poidse)
tab incoherencepoids, missing /*il n'y a pas d'incohérences: poidse vaut bien la

```

```

moitié de poids*/
    drop incoherencepoids
*Je crée POIDS_IND
generate poids_ind = poids_i if (lienego_ind!="02" & lienego_ind!="200" & lienego_ind!="201")
/*si l'individu n'est pas un enfant de l'enquêté, sa pondération vaut poids_i*/
replace poids_ind = poids_e if (lienego_ind=="02" | lienego_ind=="200" | lienego_ind=="201")
/*si l'individu est un enfant de l'enquêté, sa pondération vaut poids_e*/
summarize poids_ind /*Cette variable est bien renseignée pour les 83 045
individus et va de 8 à 59 701*/

*J'applique des labels à la variable POIDS_IND
label variable poids_ind "pondération"

*SEXE_IND (sexe de l'individu)
tab lienego_ind sexe, missing /*Cette variable est bien renseignée pour les 75 121
habitants du logement de l'enquêté, mais pas pour les 7 924 enfants de l'enquêté non
cohabitants ou décédés à la date de l'enquête*/
tab lienego_ind sexee, missing /*Cette variable est bien renseignée pour les 21 761
enquêtés*/
tab lienego_ind e_sexe, missing /*Cette variable est bien renseignée pour les 31 476
enfants des enquêtés, y compris non cohabitants ou décédés à la date de l'enquête*/
*Je vérifie qu'il n'y a pas d'incohérences entre ces variables de sexe
generate incoherencesexe = 1 if (sexe!="" & sexee!="" & sexe!=sexee) | (sexe!="" & e_sexe!=""
& sexe!=e_sexe)
tab incoherencesexe, missing /*il n'y a pas d'incohérences: le sexe renseigné dans
une variable est le même qui est renseigné dans l'autre*/
drop incoherencesexe
*Je crée SEXE_IND
generate sexe_ind = sexe if (lienego_ind!="02" & lienego_ind!="200" & lienego_ind!="201")
/*si l'individu n'est pas un enfant de l'enquêté, son sexe vaut sexe*/
replace sexe_ind = e_sexe if (lienego_ind=="02" | lienego_ind=="200" | lienego_ind=="201")
/*si l'individu est un enfant de l'enquêté, son sexe vaut e_sexe*/
tab sexe_ind, missing /*Cette variable est bien renseignée pour 83 035
individus (on compte 10 Refus de répondre) et va de 1 (Homme) à 2 (Femme)*/
/*SEXE_IND est codée comme suit:
1=homme
2=femme
8=refus de répondre*/

*J'applique des labels à la variable SEXE_IND
destring sexe_ind, replace
label variable sexe_ind "sexe"
label define sexe 1 "Homme" 2 "Femme" 8 "Non réponse"
label values sexe_ind sexe
tab sexe_ind, missing

*ANAIIS_IND (année de naissance de l'individu)
/*tab anais, missing*/ /*Cette variable est bien renseignée pour les 75 121
habitants du logement de l'enquêté, mais pas pour les 7 924 enfants de l'enquêté non
cohabitants ou décédés à la date de l'enquête*/
tab anaise, missing /*Cette variable est bien renseignée pour les 21 761
enquêtés*/
/*tab e_anais, missing*/ /*Cette variable est bien renseignée pour 31 392
enfants des enquêtés*/
*Je vérifie qu'il n'y a pas d'incohérences entre ces variables d'année de naissance
generate incoherenceanais = 1 if (anais!=. & anaise!=. & anais!=anaise) | (anais!=. &
e_anais!=. & anais!=e_anais)
tab incoherenceanais, missing /*il n'y a pas d'incohérences: l'année de naissance
renseignée dans une variable est la même qui est renseignée dans l'autre*/
drop incoherenceanais
*Je crée ANAIS_IND
generate anais_ind = anais if (lienego_ind!="02" & lienego_ind!="200" & lienego_ind!="201")
/*si l'individu n'est pas un enfant de l'enquêté, son année de naissance vaut anais*/
replace anais_ind = e_anais if (lienego_ind=="02" | lienego_ind=="200" | lienego_ind=="201")
/*si l'individu est un enfant de l'enquêté, son année de naissance vaut e_anais*/
/*tab anais_ind, missing*/ /*Cette variable est bien renseignée pour 82 961
individus (on compte 84 Non réponse) et va de 1900 à 2009*/

*J'applique des labels à la variable ANAIS_IND
label variable anais_ind "année de naissance"

*PRENOM_IND (prénom de l'individu)
codebook prenom
codebook prenom
codebook e_prenom
/*
- la variable PRENOM est bien renseignée pour les 75 121 habitants du logement des

```



```

enquêtés: ils portent 16 938 prénoms différents.
- la variable PRENOME est bien renseignée pour les 21 761 enquêtés: ils portent 6 699
prénoms différents.
- la variable E_PRENOM est bien renseignée pour les 31 476 enfants des enquêtés, y compris
non cohabitants ou décédés à la date de l'enquête: ils portent 8 844 prénoms différents.*/
*Je vérifie qu'il n'y a pas d'incohérences entre ces variables de prénom */
generate incoherenceprenom = 1 if (prenom!=" " & prenome!=" " & prenom!=prenome) | (prenom!=" " &
e_prenom!=" " & prenom!=e_prenom)
    tab incoherenceprenom, missing /*il n'y a pas d'incohérences: le prénom renseigné
dans une variable est le même qui est renseigné dans l'autre*/
    drop incoherenceprenom
*Je crée PRENOM_IND
generate prenom_ind = prenom if (lienego_ind!="02" & lienego_ind!="200" &
lienego_ind!="201") /*si l'individu n'est pas un enfant de l'enquêté, son prénom vaut prenom*/
    replace prenom_ind = e_prenom if (lienego_ind=="02" | lienego_ind=="200" |
lienego_ind=="201") /*si l'individu est un enfant de l'enquêté, son prénom vaut e_prenom*/
codebook prenom_ind /*Cette variable est bien renseignée pour les 83 045
individus: ils portent 18 161 prénoms différents.*/

*J'applique des labels à la variable PRENOM_IND
label variable prenom_ind "prénom"

*REGIONNAIS_IND (lieu de naissance de l'individu)
tab regionnais2, missing /*Cette variable est bien renseignée pour les 75 121
habitants du logement de l'enquêté, mais pas pour les 7 924 enfants de l'enquêté non
cohabitants ou décédés à la date de l'enquête*/
tab regionnaisenf2, missing /*Cette variable est bien renseignée pour les 31 476
enfants des enquêtés*/
*Je vérifie qu'il n'y a pas d'incohérences entre ces variables de lieu de naissance
generate incoherenceregionnais = 1 if (regionnais2!=. & regionnaisenf2!=. &
regionnais2!=regionnaisenf2)
    tab incoherenceregionnais, missing /*il n'y a pas d'incohérences: le lieu de naissance
renseigné dans une variable est le même qui est renseigné dans l'autre*/
    drop incoherenceregionnais
*Je crée REGIONNAIS_IND
generate regionnais_ind = regionnais2 if (lienego_ind!="02" & lienego_ind!="200" &
lienego_ind!="201") /*si l'individu n'est pas un enfant de l'enquêté, son lieu de naissance
vaut regionnais2*/
    replace regionnais_ind = regionnaisenf2 if (lienego_ind=="02" | lienego_ind=="200" |
lienego_ind=="201") /*si l'individu est un enfant de l'enquêté, son lieu de naissance vaut
regionnaisenf2*/
    tab regionnais_ind, missing /*Cette variable est bien renseignée pour 83 027
individus (on compte 18 non réponse) et va de 11 à 50*/
/*REGIONNAIS_IND est recodée comme suit:
11=France métropolitaine
12=DOM
20=Maghreb
29=autre Afrique
30=Asie
34=Turquie
40=Europe du Sud
49=autre Europe
50=Ailleurs dans le monde
90=NSP*/

*J'applique des labels à la variable REGIONNAIS_IND
label variable regionnais_ind "lieu de naissance"
    label define regionnais 11 "France métropolitaine" 12 "DOM" 20 "Maghreb" 29 "autre Afrique"
30 "Asie" 34 "Turquie" 40 "Europe du Sud" 49 "autre Europe" 50 "Ailleurs dans le monde" 90
"NSP"
    label values regionnais_ind regionnais
    tab regionnais_ind, missing

*RESIDENCE_IND (lieu de résidence de l'individu)
generate residence_ind = 1 if (lienego_ind!="02" & lienego_ind!="200" & lienego_ind!="201")
    replace residence_ind = 1 if (lienego_ind=="02" | lienego_ind=="200" | lienego_ind=="201") &
(e_lvie=="1")
    tab residence_ind, missing /*seuls 1 611 enfants d'enquêté ne résident pas en France
métropolitaine (y compris parce qu'ils sont décédés)*/

*GROUP2_IND (lien à la migration de l'enquêté et de son conjoint actuel)
*J'attribue group2ca à toute la famille
    tab group2ca, missing
egen group2ca_famille = max(group2ca), by(ident)
    tab group2ca_famille, missing
*Je crée GROUP2_IND
generate group2_ind = group2 if (lienego_ind=="00")

```

```

    replace group2_ind = group2ca_famille if (lienego_ind=="01")
        tab group2_ind, missing /*Cette variable est bien renseignée pour 35 004
individus et va de 10 à 56*/
        tab lienego_ind group2_ind, missing /*Cette variable n'est renseignée que pour des
enquêtés et leurs conjoints actuels, ce qui est normal*/

*NATPNIV2 et NATMNIV2 (nationalité à la naissance du père / de la mère de l'enquêté)
*Je vérifie qu'il n'y a pas d'invéraisemblances entre ces variables
/*tab natpniv2 natmniv2 if (lienego_ind=="00"), missing*/ /*les diagonales sont bien les plus
remplies*/

*N_INDPER et N_INDMER (choix de nationalité du père / de la mère si né(e) dans une ancienne
colonie française)
*Je vérifie qu'il n'y a pas d'invéraisemblances entre ces variables
/*tab n_indper n_indmer if (lienego_ind=="00"), missing*/ /*les diagonales sont bien les plus
remplies*/

*LNP1_GR, LNP2_GR, LNM1_GR et LNM2_GR (langues parlées par les parents de l'enquêté dans son
enfance)
*Je vérifie qu'il n'y a pas d'invéraisemblances entre ces variables
/*tab lnp1_gr lnm1_gr if (lienego_ind=="00"), missing*/ /*les diagonales sont bien les plus
remplies*/
/*tab lnp2_gr lnm2_gr if (lienego_ind=="00"), missing*/ /*les diagonales sont bien les plus
remplies*/

*RELIG_IND (religion de l'individu)
*Je vérifie qu'il n'y a pas d'invéraisemblances entre ces variables de religion*/
tab relegol relcal, missing /*les diagonales sont bien les plus remplies*/
tab relegol relpcl, missing /*les diagonales sont bien les plus remplies*/
/*tab relegol relperel, missing*/ /*les diagonales sont bien les plus remplies*/
tab relegol relmerel, missing /*les diagonales sont bien les plus remplies*/
*Je crée RELIG_IND
generate relig_ind = relegol if (lienego_ind=="00") /*si l'individu est un
enquêté, sa religion vaut relegol*/
replace relig_ind = relcal if (lienego_ind=="01") /*si l'individu est un
conjoint actuel de l'enquêté, sa religion vaut relcal*/
/*replace relig_ind = relpcl if (lienego_ind=="02")*/
replace relig_ind = relperel if (lienego_ind=="03" & sexe_ind==1) /*si l'individu est le
père de l'enquêté, sa religion vaut relperel*/
replace relig_ind = relmerel if (lienego_ind=="03" & sexe_ind==2) /*si l'individu est la
mère de l'enquêté, sa religion vaut relmerel*/
tab relig_ind, missing /*Cette variable est
renseignée pour 42 651 individus (on compte 819 Non réponse) et va de 1 à 40*/
tab lienego_ind relig_ind, missing /*Cette variable n'est
renseignée que pour les enquêtés et leurs conjoints actuels et les parents des enquêtés, ce
qui est normal*/
/*RELIG_IND est (re)codée comme suit:
1=sans religion
10=chrétien
20=musulman
30=juif
40=bouddhiste, hindouiste, autre, multi-religions
88=Refus de répondre, NSP*/

*J'applique des labels à la variable RELIG_IND
label variable relig_ind "religion"
label define religion 1 "Sans religion" 10 "Chrétien" 20 "Musulman" 30 "Juif" 40 "Bouddhiste
etc." 88 "Non réponse"
label values relig_ind religion
tab relig_ind, missing

*R_IMPVE (religiosité, i.e. importance qu'accorde aujourd'hui l'enquêté à la religion dans sa
vie)
tab r_impvie, missing /*Cette variable n'est renseignée que pour les 15 239
enquêtés (sur 21 761) ayant déclaré avoir aujourd'hui une religion, et elle va de 1 à 8*/
/*tab relegol r_impvie, missing*/
tab lienego_ind r_impvie, missing /*Cette variable n'est renseignée que pour les enquêtés,
ce qui est normal*/
/*R_IMPVE est codée comme suit:
1=pas du tout d'importance
2=un peu d'importance
3=assez d'importance
4=beaucoup d'importance
8=refus de répondre
9=ne sait pas*/

*J'applique des labels à la variable R_IMPVE

```

```

    deststring r_impvie, replace
label variable r_impvie "religiosite"
    label define religiosite 1 "Pas du tout d'importance" 2 "Un peu d'importance" 3 "Assez
d'importance" 4 "Beaucoup d'importance" 8 "Refus de répondre" 9 "Ne sait pas"
    label values r_impvie religiosite
    tab r_impvie, missing

*ORIG_RELIG_IND (origine religieuse de l'individu, i.e. religions de ses père et mère)
tostring relegol relcal relpcl relperel relmerel, replace
    replace relegol = "01" if (relegol=="1")
    replace relcal = "01" if (relcal=="1")
    replace relpcl = "01" if (relpcl=="1")
    replace relperel = "01" if (relperel=="1")
    replace relmerel = "01" if (relmerel=="1")
    replace relegol = "88" if (relegol=="")
    replace relcal = "88" if (relcal=="")
    replace relpcl = "88" if (relpcl=="")
    replace relperel = "88" if (relperel=="")
    replace relmerel = "88" if (relmerel=="")
*Pour créer ORIG_RELIG_IND je dois d'abord créer, pour chaque ménage (y compris les enfants
non cohabitants), plusieurs variables
**le sexe de l'enquêté
generate sexe_ind_enquete = sexe_ind if (lienego_ind == "00")
    egen sexe_enquete = mode(sexe_ind_enquete), by(ident)
tab sexe_enquete, missing
*Je crée ensuite relig_pere_ind (religion du père de l'individu)
generate relig_pere_ind = relegol if (lienego_ind=="02" | lienego_ind=="200" |
lienego_ind=="201") & (sexe_enquete==1) /*si l'individu est un enfant de
l'enquêté (vivant ou non, cohabitant ou non) et que l'enquêté est un homme, la religion de son
père vaut relegol*/
    replace relig_pere_ind = relcal if (lienego_ind=="02" | lienego_ind=="200" |
lienego_ind=="201") & (sexe_enquete==2) & (autreparent==1) /*si l'individu est un enfant de
l'enquêté (vivant ou non, cohabitant ou non) et que l'enquêté est une femme, et que son père
est le conjoint actuel de cette dernière, la religion de son père vaut relcal*/
    replace relig_pere_ind = relpcl if (lienego_ind=="02" | lienego_ind=="200" |
lienego_ind=="201") & (sexe_enquete==2) & (autreparent==0) /*si l'individu est un enfant de
l'enquêté (vivant ou non, cohabitant ou non) et que l'enquêté est une femme, et que son père
est le conjoint passé de cette dernière, la religion de son père vaut relpcl*/
    replace relig_pere_ind = relcal if (lienego_ind=="31")
& (sexe_enquete==2) & (autreparent==1) /*si l'individu est un bel-enfant de l'enquêté
et que l'enquêté est une femme, et que son père est le conjoint actuel de cette dernière, la
religion de son père vaut relcal*/
    replace relig_pere_ind = relpcl if (lienego_ind=="31")
& (sexe_enquete==2) & (autreparent==0) /*si l'individu est un bel-enfant de l'enquêté
et que l'enquêté est une femme, et que son père est le conjoint passé de cette dernière, la
religion de son père vaut relpcl*/
    replace relig_pere_ind = "88" if (lienego_ind=="31")
& (sexe_enquete==1) /*si l'individu est un bel-enfant de l'enquêté
et que l'enquêté est un homme, la religion de son père n'est pas renseignée*/
    replace relig_pere_ind = relperel if (lienego_ind=="00" | lienego_ind=="10")
/*si l'individu est un enquêté, ou un frère ou soeur de l'enquêté, la religion de son père
vaut relperel*/
    replace relig_pere_ind = "88" if (lienego_ind=="01" | lienego_ind=="03" |
lienego_ind=="21" | lienego_ind=="22" | lienego_ind=="32" | lienego_ind=="40" |
lienego_ind=="50" | lienego_ind=="60" | lienego_ind=="90") /*si l'individu est un conjoint
actuel de l'enquêté, ou un parent de l'enquêté, ou un petit-enfant de l'enquêté, ou un grand-
parent de l'enquêté, ou un beau-parent de l'enquêté, ou un autre lien familial, ou un lien
familial indéterminé, ou un ami de l'enquêté, ou un autre lien non familial, la religion de
son père n'est pas renseignée*/
    replace relig_pere_ind = "88" if ((lienego_ind=="02" & relig_pere_ind=="") |
(lienego_ind=="200" & relig_pere_ind=="") | (lienego_ind=="201" & relig_pere_ind=="") |
(lienego_ind=="31" & relig_pere_ind=="")) /*si l'individu est un enfant ou bel-enfant de
l'enquêté dont on ignore la religion du père, la religion de son père n'est pas renseignée*/
    tab lienego_ind relig_pere_ind, missing
*Je crée ensuite relig_mere_ind (religion de la mère de l'individu)
generate relig_mere_ind = relegol if (lienego_ind=="02" | lienego_ind=="200" |
lienego_ind=="201") & (sexe_enquete==2) /*si l'individu est un enfant de
l'enquêté (vivant ou non, cohabitant ou non) et que l'enquêté est une femme, la religion de sa
mère vaut relegol*/
    replace relig_mere_ind = relcal if (lienego_ind=="02" | lienego_ind=="200" |
lienego_ind=="201") & (sexe_enquete==1) & (autreparent==1) /*si l'individu est un enfant de
l'enquêté (vivant ou non, cohabitant ou non) et que l'enquêté est un homme, et que sa mère est
le conjoint actuel de ce dernier, la religion de sa mère vaut relcal*/
    replace relig_mere_ind = relpcl if (lienego_ind=="02" | lienego_ind=="200" |
lienego_ind=="201") & (sexe_enquete==1) & (autreparent==0) /*si l'individu est un enfant de
l'enquêté (vivant ou non, cohabitant ou non) et que l'enquêté est un homme, et que sa mère est
le conjoint passé de ce dernier, la religion de sa mère vaut relpcl*/

```

```

    replace relig_mere_ind = relcal if (lienego_ind=="31")
    & (sexe_enquete==1) & (autreparent==1) /*si l'individu est un bel-enfant de l'enquêté
    et que l'enquêté est un homme, et que sa mère est le conjoint actuel de ce dernier, la
    religion de sa mère vaut relcal*/
    replace relig_mere_ind = relpcl if (lienego_ind=="31")
    & (sexe_enquete==1) & (autreparent==0) /*si l'individu est un bel-enfant de l'enquêté
    et que l'enquêté est un homme, et que sa mère est le conjoint passé de ce dernier, la
    religion de sa mère vaut relpcl*/
    replace relig_mere_ind = "88" if (lienego_ind=="31")
    & (sexe_enquete==2) /*si l'individu est un bel-enfant de l'enquêté
    et que l'enquêté est une femme, la religion de sa mère n'est pas renseignée*/
    replace relig_mere_ind = relmerel if (lienego_ind=="00" | lienego_ind=="10")
    /*si l'individu est un enquêté, ou un frère ou soeur de l'enquêté, la religion de sa mère vaut
    relmerel*/
    replace relig_mere_ind = "88" if (lienego_ind=="01" | lienego_ind=="03" |
    lienego_ind=="21" | lienego_ind=="22" | lienego_ind=="32" | lienego_ind=="40" |
    lienego_ind=="50" | lienego_ind=="60" | lienego_ind=="90") /*si l'individu est un conjoint
    actuel de l'enquêté, ou un parent de l'enquêté, ou un petit-enfant de l'enquêté, ou un grand-
    parent de l'enquêté, ou un beau-parent de l'enquêté, ou un autre lien familial, ou un lien
    familial indéterminé, ou un ami de l'enquêté, ou un autre lien non familial, la religion de sa
    mère n'est pas renseignée*/
    replace relig_mere_ind = "88" if ((lienego_ind=="02" & relig_mere_ind=="") |
    (lienego_ind=="200" & relig_mere_ind=="") | (lienego_ind=="201" & relig_mere_ind=="") |
    (lienego_ind=="31" & relig_mere_ind=="")) /*si l'individu est un enfant ou bel-enfant de
    l'enquêté dont on ignore la religion de la mère, la religion de sa mère n'est pas renseignée*/
    tab lienego_ind relig_mere_ind, missing
*Je peux enfin créer ORIG_RELIG_IND
generate orig_relig_ind = relig_pere_ind + " " + relig_mere_ind
    replace orig_relig_ind = "10" if (orig_relig_ind=="01 10" | orig_relig_ind=="10 01" |
    orig_relig_ind=="10 10" | orig_relig_ind=="10 20" | orig_relig_ind=="10 30" |
    orig_relig_ind=="10 40" | orig_relig_ind=="10 88" | orig_relig_ind=="88 10")
    replace orig_relig_ind = "20" if (orig_relig_ind=="01 20" | orig_relig_ind=="20 01" |
    orig_relig_ind=="20 10" | orig_relig_ind=="20 20" | orig_relig_ind=="20 30" |
    orig_relig_ind=="20 40" | orig_relig_ind=="20 88" | orig_relig_ind=="88 20")
    replace orig_relig_ind = "30" if (orig_relig_ind=="01 30" | orig_relig_ind=="30 01" |
    orig_relig_ind=="30 10" | orig_relig_ind=="30 20" | orig_relig_ind=="30 30" |
    orig_relig_ind=="30 40" | orig_relig_ind=="30 88" | orig_relig_ind=="88 30")
    replace orig_relig_ind = "40" if (orig_relig_ind=="01 40" | orig_relig_ind=="40 01" |
    orig_relig_ind=="40 10" | orig_relig_ind=="40 20" | orig_relig_ind=="40 30" |
    orig_relig_ind=="40 40" | orig_relig_ind=="40 88" | orig_relig_ind=="88 40")
    replace orig_relig_ind = "100" if (orig_relig_ind=="01 01" | orig_relig_ind=="01 88" |
    orig_relig_ind=="88 01")
    replace orig_relig_ind = "." if (orig_relig_ind=="88 88")
destring orig_relig_ind, replace
    tab orig_relig_ind, missing /*Cette variable est renseignée pour 59 082
individus et va de 10 (origine chrétienne) à 100 (sans origine religieuse)*/
    tab orig_relig_ind lienego_ind, missing /*Cette variable n'est renseignée que pour les
enquêtés et leurs enfants, ainsi que pour les frères et soeurs des enquêtés, ce qui est
normal*/
/*ORIG_RELIG_IND est (re)codée comme suit:
10=origine chrétienne
20=origine musulmane
30=origine juive
40=origine bouddhiste, hindouiste, autre, multi-religions
100=sans origine religieuse
. =Non réponse (structurelle), Refus de répondre, NSP*/

*J'applique des labels à la variable ORIG_RELIG_IND
label variable orig_relig_ind "origine religieuse"
label define origrelig 10 "chrétienne" 20 "musulmane" 30 "juive" 40 "bouddhiste, etc." 100
"sans origine religieuse"
label values orig_relig_ind origrelig
    tab orig_relig_ind, missing

*X_VUFRI (opinion de l'enquêté sur: "On me voit comme un Français")
tab x_vufri, missing /*Cette variable est bien renseignée pour les 21 761 enquêtés
et va de 1 à 9.*/
tab lienego_ind x_vufri, missing /*Cette variable n'est renseignée que pour les enquêtés, ce
qui est normal*/

*J'applique des labels à la variable X_VUFRI
destring x_vufri, replace
label variable x_vufri "on me voit comme un Français"
label define vucomme 1 "tout à fait d'accord" 2 "plutôt d'accord" 3 "plutôt pas d'accord" 4
"pas du tout d'accord" 8 "ne sait pas" 9 "ne souhaite pas répondre"
label values x_vufri vucomme
    tab x_vufri, missing

```

```

*ORIG_IND (origine "dominante" de l'individu)
*Cette variable étant relativement complexe à construire, je procède étape par étape

*Dans ORIG_IND, je crée l'origine des enquêtés
  tab orig_enq, missing /*Cette variable est
bien renseignée pour les 83 045 individus, et elle va de 120 à 502*/
generate orig_ind = orig_enq if (lienego_ind=="00") /*si l'individu est un
enquêté, son origine vaut orig_enq*/

*Dans ORIG_IND, je crée l'origine des conjoints actuels des enquêtés
  tab orig_ca, missing /*Cette variable n'est
renseignée que pour 62 814 individus, et elle va de 120 à 502*/
  replace orig_ind = orig_ca if (lienego_ind=="01") /*si l'individu est le
conjoint d'un enquêté, son origine vaut orig_ca; en revanche orig_pc n'est plus utile à ce
stade*/

*Dans ORIG_IND, je crée l'origine des enfants des enquêtés
  tab orig_enfeng, missing /*Cette variable est
bien renseignée pour les 31 476 enfants des enquêtés, et elle va de 320 à 712*/
  replace orig_ind = orig_enfeng if (lienego_ind=="02" | lienego_ind=="200" |
lienego_ind=="201") /*si l'individu est l'enfant cohabitant d'un enquêté, ou l'enfant non
cohabitant d'un enquêté, son origine vaut orig_enfeng*/

*Dans ORIG_IND, je crée l'origine des parents des enquêtés
  tab orig_mereeng, missing /*Cette variable est
bien renseignée pour les 83 045 individus, et elle va de 120 à 501*/
  tab orig_pereeng, missing /*Cette variable est
bien renseignée pour les 83 045 individus, et elle va de 120 à 501*/
  replace orig_ind = orig_mereeng if (lienego_ind=="03" & sexe_ind==2) /*si l'individu est un
parent féminin (mère) d'un enquêté, son origine vaut orig_mereeng*/
  replace orig_ind = orig_pereeng if (lienego_ind=="03" & sexe_ind==1) /*si l'individu est un
parent masculin (père) d'un enquêté, son origine vaut orig_pereeng*/

*Dans ORIG_IND, je crée l'origine des frères et soeurs des enquêtés
  tab orig_freresoeureng, missing /*Cette variable est bien renseignée pour
les 83 045 individus, et elle va de 120 à 502*/
  replace orig_ind = orig_freresoeureng if (lienego_ind=="10") /*si l'individu est un
frère ou une soeur d'un enquêté, son origine vaut orig_freresoeureng*/

*Dans ORIG_IND, je crée l'origine des "beaux-enfants" des enquêtés
*Attention: comme le terme "beaux-enfants" est ambigu, il y a 2 possibilités: si l'individu
est l'enfant du conjoint actuel de l'enquêté, on connaît son origine (c'est orig_enfca); mais
si l'individu est le conjoint de l'enfant de l'enquêté, on ignore son origine
  tab orig_enfca, missing /*Cette variable n'est
renseignée que pour 62 814 individus, et elle va de 320 à 712*/
*Je crée donc d'abord une variable ENFCA=1, indiquant que l'individu est le "bel-enfant" de
l'enquêté en ce sens qu'il est l'enfant du conjoint actuel de l'enquêté (mais pas en cet autre
sens qu'il serait le conjoint de l'enfant de l'enquêté)
sort ident noi
generate noi_ca = noi if lienego_ind=="01" /*si l'individu est
conjoint actuel de l'enquêté, on lui attribue son noi*/
ssc inst mipolate
by ident: stripolate noi_ca ident, gen(noi_ca_group) groupwise /*si le logement
comprend un conjoint actuel de l'enquêté, on attribue son noi à tous ses habitants - et même
aux apparentés non cohabitants */
generate enfca = 1 if lienego_ind=="31" & ((lien_01=="02" &
noi_ca_group=="01") | (lien_02=="02" & noi_ca_group=="02") | (lien_03=="02" &
noi_ca_group=="03") | (lien_04=="02" & noi_ca_group=="04") | (lien_05=="02" &
noi_ca_group=="05") | (lien_06=="02" & noi_ca_group=="06") | (lien_07=="02" &
noi_ca_group=="07") | (lien_08=="02" & noi_ca_group=="08") | (lien_09=="02" &
noi_ca_group=="09") | (lien_10=="02" & noi_ca_group=="10") | (lien_11=="02" &
noi_ca_group=="11") | (lien_12=="02" & noi_ca_group=="12") | (lien_13=="02" &
noi_ca_group=="13") | (lien_14=="02" & noi_ca_group=="14") | (lien_15=="02" &
noi_ca_group=="15") | (lien_16=="02" & noi_ca_group=="16") | (lien_17=="02" &
noi_ca_group=="17") | (lien_18=="02" & noi_ca_group=="18") | (lien_19=="02" &
noi_ca_group=="19") | (lien_20=="02" & noi_ca_group=="20")) /*si le "bel-enfant" de l'enquêté
est l'enfant (lien_xx=="02") d'un individu qui est le conjoint actuel de l'enquêté
(noi_ca_group=="xx"), ce "bel-enfant" de l'enquêté est bien l'enfant du conjoint actuel de
l'enquêté (plutôt que le conjoint de l'enfant de l'enquêté) */
  tab enfca, missing /*sur les 639 "beaux-
enfants" de l'enquêté (LIENEGO=31), seuls 592 sont enfants du conjoint actuel de l'enquêté
(ENFCA=1), les autres devant être conjoints de l'enfant de l'enquêté */
*Je peux enfin créer l'origine des "beaux-enfants" des enquêtés
  replace orig_ind = orig_enfca if (lienego_ind=="31" & enfca==1) /*si l'individu est un
enfant du conjoint actuel d'un enquêté, son origine vaut orig_enfca*/
  replace orig_ind = . if (lienego_ind=="31" & enfca!=1) /*si l'individu est un

```

```

"bel-enfant" mais pas un enfant du conjoint actuel d'un enquêté, son origine est inconnue*/

*Dans ORIG_IND, je crée l'origine des "beaux-parents" des enquêtés
*Attention: comme le terme "beaux-parents" est ambigu, il y a 2 possibilités: si l'individu
est un parent du conjoint actuel de l'enquêté, on connaît son origine (c'est orig_mereca ou
orig_pereca); mais si l'individu est le conjoint d'un parent de l'enquêté, on ignore son
origine
    tab orig_mereca, missing /*Cette variable n'est renseignée que pour 62 814 individus, et
elle va de 120 à 501*/
    tab orig_pereca, missing /*Cette variable n'est renseignée que pour 62 814 individus, et
elle va de 120 à 501*/
*Je crée donc d'abord une variable MERECA=1, indiquant que l'individu est la "belle-mère" de
l'enquêté en ce sens qu'elle est la mère du conjoint actuel de l'enquêté (mais pas en cet
autre sens qu'elle serait la conjointe du père de l'enquêté)
tab lienego sexe_ind, missing
generate mereca = 1 if lienego_ind=="32" & sexe_ind==2 & ((lien_01=="03" & noi_ca_group=="01")
| (lien_02=="03" & noi_ca_group=="02") | (lien_03=="03" & noi_ca_group=="03") | (lien_04=="03"
& noi_ca_group=="04") | (lien_05=="03" & noi_ca_group=="05") | (lien_06=="03" &
noi_ca_group=="06") | (lien_07=="03" & noi_ca_group=="07") | (lien_08=="03" &
noi_ca_group=="08") | (lien_09=="03" & noi_ca_group=="09") | (lien_10=="03" &
noi_ca_group=="10") | (lien_11=="03" & noi_ca_group=="11") | (lien_12=="03" &
noi_ca_group=="12") | (lien_13=="03" & noi_ca_group=="13") | (lien_14=="03" &
noi_ca_group=="14") | (lien_15=="03" & noi_ca_group=="15") | (lien_16=="03" &
noi_ca_group=="16") | (lien_17=="03" & noi_ca_group=="17") | (lien_18=="03" &
noi_ca_group=="18") | (lien_19=="03" & noi_ca_group=="19") | (lien_20=="03" &
noi_ca_group=="20")) /*si la "belle-mère" de l'enquêté est le parent (lien_xx=="03") d'un
individu qui est le conjoint actuel de l'enquêté (noi_ca_group="xx"), cette "belle-mère" de
l'enquêté est bien la mère du conjoint actuel de l'enquêté (plutôt que la conjointe du père de
l'enquêté) */
    tab mereca, missing /*sur les 184 "belles-mères" de l'enquêté (LIENEGO=32 &
SEXE_IND=2), seules 106 sont mères du conjoint actuel de l'enquêté (MERECA=1), les autres
devant être conjointes du père de l'enquêté */
*Je crée donc ensuite une variable PERECA=1, indiquant que l'individu est le "beau-père" de
l'enquêté en ce sens qu'il est le père du conjoint actuel de l'enquêté (mais pas en cet autre
sens qu'il serait le conjoint de la mère de l'enquêté)
generate pereca = 1 if lienego_ind=="32" & sexe_ind==1 & ((lien_01=="03" & noi_ca_group=="01")
| (lien_02=="03" & noi_ca_group=="02") | (lien_03=="03" & noi_ca_group=="03") | (lien_04=="03"
& noi_ca_group=="04") | (lien_05=="03" & noi_ca_group=="05") | (lien_06=="03" &
noi_ca_group=="06") | (lien_07=="03" & noi_ca_group=="07") | (lien_08=="03" &
noi_ca_group=="08") | (lien_09=="03" & noi_ca_group=="09") | (lien_10=="03" &
noi_ca_group=="10") | (lien_11=="03" & noi_ca_group=="11") | (lien_12=="03" &
noi_ca_group=="12") | (lien_13=="03" & noi_ca_group=="13") | (lien_14=="03" &
noi_ca_group=="14") | (lien_15=="03" & noi_ca_group=="15") | (lien_16=="03" &
noi_ca_group=="16") | (lien_17=="03" & noi_ca_group=="17") | (lien_18=="03" &
noi_ca_group=="18") | (lien_19=="03" & noi_ca_group=="19") | (lien_20=="03" &
noi_ca_group=="20")) /*si le "beau-père" de l'enquêté est le parent (lien_xx=="03") d'un
individu qui est le conjoint actuel de l'enquêté (noi_ca_group="xx"), ce "beau-père" de
l'enquêté est bien le père du conjoint actuel de l'enquêté (plutôt que le conjoint de la mère
de l'enquêté) */
    tab pereca, missing /*sur les 207 "beaux-pères" de l'enquêté (LIENEGO=32 &
SEXE_IND=1), seuls 44 sont pères du conjoint actuel de l'enquêté (PERECA=1), les autres devant
être conjoints de la mère de l'enquêté */
*Je peux enfin créer l'origine des "beaux-parents" des enquêtés
    replace orig_ind = orig_mereca if (lienego_ind=="32" & mereca==1 & sexe_ind==2) /*si
l'individu est la mère du conjoint actuel d'un enquêté, son origine vaut mereca; en revanche
orig_merepc n'est plus utile à ce stade*/
    replace orig_ind = . if (lienego_ind=="32" & mereca!=1 & sexe_ind==2) /*si
l'individu est une "belle-mère" d'un enquêté mais pas la mère du conjoint actuel d'un enquêté,
son origine est inconnue*/
    replace orig_ind = orig_pereca if (lienego_ind=="32" & pereca==1 & sexe_ind==1) /*si
l'individu est le père du conjoint actuel d'un enquêté, son origine vaut pereca; en revanche
orig_perepc n'est plus utile à ce stade*/
    replace orig_ind = . if (lienego_ind=="32" & pereca!=1 & sexe_ind==1) /*si
l'individu est un "beau-père" d'un enquêté mais pas le père du conjoint actuel d'un enquêté,
son origine est inconnue*/

*Dans ORIG_IND, je crée l'origine des grands-parents des enquêtés
    tab orig_grdparentmaternelnq, missing /*Cette variable est
bien renseignée pour les 83 045 individus, et elle va de 120 à 500*/
    tab orig_grdparentpaternelnq, missing /*Cette variable est
bien renseignée pour les 83 045 individus, et elle va de 120 à 500*/
*Je crée d'abord une variable GRDPARENTMATERNELENG=1, indiquant que l'individu est le grand-
parent maternel de l'enquêté
sort ident noi
generate noi_mere = noi if (lienego_ind=="03" & sexe_ind==2) /*si l'individu est la
mère de l'enquêté, on lui attribue son noi*/
by ident: stripolate noi_mere ident, gen(noi_mere_group) groupwise /*si le logement

```

```

comprend la mère de l'enquêté, on attribue son noi à tous ses habitants - et même aux enfants
non cohabitants */
generate grdparentmaternelnq = 1 if lienego_ind=="22" & ((lien_01=="03" &
noi_mere_group=="01") | (lien_02=="03" & noi_mere_group=="02") | (lien_03=="03" &
noi_mere_group=="03") | (lien_04=="03" & noi_mere_group=="04") | (lien_05=="03" &
noi_mere_group=="05") | (lien_06=="03" & noi_mere_group=="06") | (lien_07=="03" &
noi_mere_group=="07") | (lien_08=="03" & noi_mere_group=="08") | (lien_09=="03" &
noi_mere_group=="09") | (lien_10=="03" & noi_mere_group=="10") | (lien_11=="03" &
noi_mere_group=="11") | (lien_12=="03" & noi_mere_group=="12") | (lien_13=="03" &
noi_mere_group=="13") | (lien_14=="03" & noi_mere_group=="14") | (lien_15=="03" &
noi_mere_group=="15") | (lien_16=="03" & noi_mere_group=="16") | (lien_17=="03" &
noi_mere_group=="17") | (lien_18=="03" & noi_mere_group=="18") | (lien_19=="03" &
noi_mere_group=="19") | (lien_20=="03" & noi_mere_group=="20")) /*si l'individu est grand-
parent de l'enquêté et parent (lien_xx=="03") de la mère de l'enquêté (noi_mere_group=="xx"),
cet individu est bien grand-parent maternel de l'enquêté*/

tab grdparentmaternelnq, missing /*sur les 81 grands-
parents de l'enquêté (LIENEGO=22), seuls 43 sont maternels*/
*Je crée ensuite une variable GRDPARENTPATERNELENQ=1, indiquant que l'individu est le grand-
parent paternel de l'enquêté
sort ident noi
generate noi_pere = noi if (lienego_ind=="03" & sexe_ind==1) /*si l'individu est le
père de l'enquêté, on lui attribue son noi*/
by ident: stripolate noi_pere ident, gen(noi_pere_group) groupwise /*si le logement
comprend le père de l'enquêté, on attribue son noi à tous ses habitants - et même aux enfants
non cohabitants */
generate grdparentpaternelnq = 1 if lienego_ind=="22" & ((lien_01=="03" &
noi_pere_group=="01") | (lien_02=="03" & noi_pere_group=="02") | (lien_03=="03" &
noi_pere_group=="03") | (lien_04=="03" & noi_pere_group=="04") | (lien_05=="03" &
noi_pere_group=="05") | (lien_06=="03" & noi_pere_group=="06") | (lien_07=="03" &
noi_pere_group=="07") | (lien_08=="03" & noi_pere_group=="08") | (lien_09=="03" &
noi_pere_group=="09") | (lien_10=="03" & noi_pere_group=="10") | (lien_11=="03" &
noi_pere_group=="11") | (lien_12=="03" & noi_pere_group=="12") | (lien_13=="03" &
noi_pere_group=="13") | (lien_14=="03" & noi_pere_group=="14") | (lien_15=="03" &
noi_pere_group=="15") | (lien_16=="03" & noi_pere_group=="16") | (lien_17=="03" &
noi_pere_group=="17") | (lien_18=="03" & noi_pere_group=="18") | (lien_19=="03" &
noi_pere_group=="19") | (lien_20=="03" & noi_pere_group=="20")) /*si l'individu est grand-
parent de l'enquêté et parent (lien_xx=="03") du père de l'enquêté (noi_pere_group=="xx"), cet
individu est bien grand-parent paternel de l'enquêté*/

tab grdparentpaternelnq, missing /*sur les 81 grands-
parents de l'enquêté (LIENEGO=22), seuls 18 sont paternels*/

/*sur les 81 grands-
parents de l'enquêté (LIENEGO=22), 20 sont inattribuables à une lignée (maternelle ou
paternelle)*/
*Je peux enfin créer l'origine des grands-parents des enquêtés
replace orig_ind = orig_grdparentmaternelnq if (lienego_ind=="22" &
orig_grdparentmaternelnq==orig_grdparentpaternelnq) /*si l'individu est grand-parent d'un
enquêté et que son origine est la même qu'il soit du côté maternel ou paternel, on lui
attribue cette origine*/
replace orig_ind = orig_grdparentmaternelnq if (lienego_ind=="22" &
orig_grdparentmaternelnq!=orig_grdparentpaternelnq & grdparentmaternelnq==1) /*si
l'individu est grand-parent d'un enquêté et que son origine diffère qu'il soit du côté
maternel ou paternel, mais qu'on sait qu'il est grand-parent maternel, son origine vaut
orig_grdparentmaternelnq*/
replace orig_ind = orig_grdparentpaternelnq if (lienego_ind=="22" &
orig_grdparentmaternelnq!=orig_grdparentpaternelnq & grdparentpaternelnq==1) /*si
l'individu est grand-parent d'un enquêté et que son origine diffère qu'il soit du côté
maternel ou paternel, mais qu'on sait qu'il est grand-parent paternel, son origine vaut
orig_grdparentpaternelnq*/

*Dans ORIG_IND, je crée l'origine des autres liens familiaux des enquêtés
replace orig_ind = . if (lienego_ind=="40") /*si l'individu est un
autre lien familial d'un enquêté, son origine est inconnue*/

*Dans ORIG_IND, je crée l'origine des liens familiaux indéterminés des enquêtés
replace orig_ind = . if (lienego_ind=="50") /*si l'individu est un
lien familial indéterminé d'un enquêté, son origine est inconnue*/

*Dans ORIG_IND, je crée l'origine des amis des enquêtés
replace orig_ind = . if (lienego_ind=="60") /*si l'individu est un
ami d'un enquêté, son origine est inconnue*/

*Dans ORIG_IND, je crée l'origine des autres liens non familiaux des enquêtés
replace orig_ind = . if (lienego_ind=="90") /*si l'individu est un
autre lien non familial d'un enquêté, son origine est inconnue*/

*Dans ORIG_IND, je recatégorise les individus originaire(s) d'"on ne sait où" (ORIG_IND=390)
en "Non réponse"

```

```

replace orig_ind = . if (orig_ind==190 | orig_ind==390)

*Je ne peux pas à ce stade créer l'origine des petits-enfants des enquêtés
(lienego_ind=="21"), j'y reviens ci-dessous

*REGIONNAISPHERE_IND (lieu de naissance du père de l'individu) et REGIONNAISMERE_IND (lieu de
naissance de la mère de l'individu)
tab sexee, missing
egen sexe_enq = mode(sexee), by(ident)
tab sexe_enq, missing
destring sexe_enq, replace
*REGIONNAISPHERE_IND (lieu de naissance du père de l'individu)
generate regionnaispere_ind = regionnaisp2 if (lienego_ind=="00" | lienego_ind=="10")
replace regionnaispere_ind = regionnaiscap2 if (lienego_ind=="01")
replace regionnaispere_ind = regionnaise2 if (lienego_ind=="02" | lienego_ind=="200" |
lienego_ind=="201") & (sexe_enq==1)
replace regionnaispere_ind = regionnaisca2 if (lienego_ind=="02" | lienego_ind=="200" |
lienego_ind=="201") & (sexe_enq==2 & autreparent==1)
replace regionnaispere_ind = regionnaisp2 if (lienego_ind=="02" | lienego_ind=="200" |
lienego_ind=="201") & (sexe_enq==2 & autreparent==0)
replace regionnaispere_ind = . if (lienego_ind=="03" | lienego_ind=="22" |
lienego_ind=="32" | lienego_ind=="40" | lienego_ind=="50" | lienego_ind=="60" |
lienego_ind=="90")
replace regionnaispere_ind = regionnaisca2 if (lienego_ind=="31") & (sexe_enq==2)
replace regionnaispere_ind = . if (lienego_ind=="31") & (sexe_enq==1)
*REGIONNAISMERE_IND (lieu de naissance de la mère de l'individu)
generate regionnaismere_ind = regionnaism2 if (lienego_ind=="00" | lienego_ind=="10")
replace regionnaismere_ind = regionnaiscam2 if (lienego_ind=="01")
replace regionnaismere_ind = regionnaise2 if (lienego_ind=="02" | lienego_ind=="200" |
lienego_ind=="201") & (sexe_enq==2)
replace regionnaismere_ind = regionnaisca2 if (lienego_ind=="02" | lienego_ind=="200" |
lienego_ind=="201") & (sexe_enq==1 & autreparent==1)
replace regionnaismere_ind = regionnaisp2 if (lienego_ind=="02" | lienego_ind=="200" |
lienego_ind=="201") & (sexe_enq==1 & autreparent==0)
replace regionnaismere_ind = . if (lienego_ind=="03" | lienego_ind=="22" |
lienego_ind=="32" | lienego_ind=="40" | lienego_ind=="50" | lienego_ind=="60" |
lienego_ind=="90")
replace regionnaismere_ind = regionnaisca2 if (lienego_ind=="31") & (sexe_enq==1)
replace regionnaismere_ind = . if (lienego_ind=="31") & (sexe_enq==2)

*Je ne peux pas à ce stade créer les REGIONNAISPHERE_IND ni REGIONNAISMERE_IND des petits-
enfants des enquêtés (lienego_ind=="21"), j'y reviens ci-dessous

*NOUS RESTENT 2 CHOSES A FAIRE:
*attribuer à chaque petit-enfant des enquêtés (il y en a 127) le REGIONNAIS_IND et le SEXE_IND
de son ou ses parent(s), pour calculer son REGIONNAISPHERE_IND et son REGIONNAISMERE_IND
*attribuer à chaque petit-enfant des enquêtés (il y en a 127) le ORIG_IND et le SEXE_IND
de son ou ses parent(s), pour calculer son ORIG_IND à lui
*cela concerne peu de monde, mais vu la structure des données, c'est un peu complexe
*Pour cela, je commence par attribuer à chaque petit-enfant des enquêtés le NOI de son ou ses
parents
generate noi_parent1_petitenf = "01" if (lienego=="21") & (lien_01=="02")
replace noi_parent1_petitenf = "02" if (lienego=="21") & (lien_01!="02" & lien_02=="02")
replace noi_parent1_petitenf = "03" if (lienego=="21") & (lien_01!="02" & lien_02!="02" &
lien_03=="02")
replace noi_parent1_petitenf = "04" if (lienego=="21") & (lien_01!="02" & lien_02!="02" &
lien_03!="02" & lien_04=="02")
replace noi_parent1_petitenf = "05" if (lienego=="21") & (lien_01!="02" & lien_02!="02" &
lien_03!="02" & lien_04!="02" & lien_05=="02")
replace noi_parent1_petitenf = "06" if (lienego=="21") & (lien_01!="02" & lien_02!="02" &
lien_03!="02" & lien_04!="02" & lien_05!="02" & lien_06=="02")
replace noi_parent1_petitenf = "07" if (lienego=="21") & (lien_01!="02" & lien_02!="02" &
lien_03!="02" & lien_04!="02" & lien_05!="02" & lien_06!="02" & lien_07=="02")
replace noi_parent1_petitenf = "08" if (lienego=="21") & (lien_01!="02" & lien_02!="02" &
lien_03!="02" & lien_04!="02" & lien_05!="02" & lien_06!="02" & lien_07!="02" & lien_08=="02")
replace noi_parent1_petitenf = "09" if (lienego=="21") & (lien_01!="02" & lien_02!="02" &
lien_03!="02" & lien_04!="02" & lien_05!="02" & lien_06!="02" & lien_07!="02" & lien_08!="02"
& lien_09=="02")
replace noi_parent1_petitenf = "10" if (lienego=="21") & (lien_01!="02" & lien_02!="02" &
lien_03!="02" & lien_04!="02" & lien_05!="02" & lien_06!="02" & lien_07!="02" & lien_08!="02"
& lien_09!="02" & lien_10=="02")
/*tab noi_parent1_petitenf if (lienego=="21"), missing*/ /*sur 127 petits-enfants, 105 ont
le NOI d'au moins 1 parent renseigné*/
generate noi_parent2_petitenf = "02" if (lienego=="21") & (noi_parent1_petitenf=="01") &
(lien_02=="02")
replace noi_parent2_petitenf = "03" if (lienego=="21") & (noi_parent1_petitenf=="01" |

```



```

noi_parent1_petitenf=="02") & (lien_03=="02")
  replace noi_parent2_petitenf = "04" if (lienego=="21") & (noi_parent1_petitenf=="01" |
noi_parent1_petitenf=="02" | noi_parent1_petitenf=="03") & (lien_04=="02")
  replace noi_parent2_petitenf = "05" if (lienego=="21") & (noi_parent1_petitenf=="01" |
noi_parent1_petitenf=="02" | noi_parent1_petitenf=="03" | noi_parent1_petitenf=="04") &
(lien_05=="02")
  replace noi_parent2_petitenf = "06" if (lienego=="21") & (noi_parent1_petitenf=="01" |
noi_parent1_petitenf=="02" | noi_parent1_petitenf=="03" | noi_parent1_petitenf=="04" |
noi_parent1_petitenf=="05") & (lien_06=="02")
  replace noi_parent2_petitenf = "07" if (lienego=="21") & (noi_parent1_petitenf=="01" |
noi_parent1_petitenf=="02" | noi_parent1_petitenf=="03" | noi_parent1_petitenf=="04" |
noi_parent1_petitenf=="05" | noi_parent1_petitenf=="06") & (lien_07=="02")
  replace noi_parent2_petitenf = "08" if (lienego=="21") & (noi_parent1_petitenf=="01" |
noi_parent1_petitenf=="02" | noi_parent1_petitenf=="03" | noi_parent1_petitenf=="04" |
noi_parent1_petitenf=="05" | noi_parent1_petitenf=="06" | noi_parent1_petitenf=="07") &
(lien_08=="02")
  replace noi_parent2_petitenf = "09" if (lienego=="21") & (noi_parent1_petitenf=="01" |
noi_parent1_petitenf=="02" | noi_parent1_petitenf=="03" | noi_parent1_petitenf=="04" |
noi_parent1_petitenf=="05" | noi_parent1_petitenf=="06" | noi_parent1_petitenf=="07" |
noi_parent1_petitenf=="08") & (lien_09=="02")
  replace noi_parent2_petitenf = "10" if (lienego=="21") & (noi_parent1_petitenf=="01" |
noi_parent1_petitenf=="02" | noi_parent1_petitenf=="03" | noi_parent1_petitenf=="04" |
noi_parent1_petitenf=="05" | noi_parent1_petitenf=="06" | noi_parent1_petitenf=="07" |
noi_parent1_petitenf=="08" | noi_parent1_petitenf=="08") & (lien_10=="02")
  tab noi_parent2_petitenf if (lienego=="21"), missing /*sur 127 petits-enfants et 105 qui
ont le NOI d'au moins 1 parent renseigné, 44 ont aussi le NOI du 2e parent renseigné*/
*Je crée ensuite une indicatrice, pour chaque famille (logement + enfants non cohabitants), de
présence d'au moins 1 petit-enfant de l'enquête
generate petitenf = 1 if lienego=="21"
egen presence_petitenf = max(petitenf), by(ident)
save repro, replace
*Je crée maintenant une base de données séparée, ne contenant que les petits-enfants des
enquêtés pour lesquels on a des infos sur au moins l'un de ses parents
use repro, clear
keep if (lienego=="21" & noi_parent1_petitenf!="")
*noi_parent1_petitenf et noi_parent2_petitenf
order ident noi noi_parent1_petitenf noi_parent2_petitenf poids_ind sexe_ind anais_ind
prenom_ind lienego_ind orig_ind regionnais_ind
save petitsenf
*Je crée maintenant une autre base de données séparée, contenant les apparentés des petits-
enfants des enquêtés
use repro
keep if presence_petitenf==1
drop if lienego=="21"
replace noi_parent1_petitenf = noi
replace noi_parent2_petitenf = noi
generate sexe_parent1_petitenf = sexe_ind
generate sexe_parent2_petitenf = sexe_ind
generate regionnais_parent1_petitenf = regionnais_ind
generate regionnais_parent2_petitenf = regionnais_ind
generate orig_parent1_petitenf = orig_ind
generate orig_parent2_petitenf = orig_ind
order ident noi noi_parent1_petitenf noi_parent2_petitenf poids_ind sexe_ind anais_ind
prenom_ind lienego_ind orig_ind regionnais_ind
save famille_petitsenf
*J'attribue les caractéristiques du 1er parent des petits-enfants des enquêtés aux petits-
enfants des enquêtés
use petitsenf
merge m:1 ident noi_parent1_petitenf using famille_petitsenf, keepusing(sexe_parent1_petitenf
regionnais_parent1_petitenf orig_parent1_petitenf) keep(match master)
/*il y a bien 105 appariements, c'est bon*/
drop _merge
*J'attribue les caractéristiques du 2e parent des petits-enfants des enquêtés aux petits-
enfants des enquêtés
merge m:1 ident noi_parent2_petitenf using famille_petitsenf, keepusing(sexe_parent2_petitenf
regionnais_parent2_petitenf orig_parent2_petitenf) keep(match master)
/*il y a bien 44 appariements, c'est bon*/
drop _merge
*Je calcule maintenant le ORIG_IND des petits-enfants de l'enquête
generate orig_pere_petitenf = orig_parent1_petitenf if (sexe_parent1_petitenf==1)
replace orig_pere_petitenf = orig_parent2_petitenf if (sexe_parent1_petitenf==2)
/*tab orig_pere_petitenf, missing*/ /**/
generate orig_mere_petitenf = orig_parent2_petitenf if (sexe_parent1_petitenf==1)
replace orig_mere_petitenf = orig_parent1_petitenf if (sexe_parent1_petitenf==2)
/*tab orig_mere_petitenf, missing*/ /**/
generate orig_petitenf = orig_pere_petitenf + 200 if (orig_pere_petitenf!=. &
orig_mere_petitenf==. & orig_pere_petitenf>120 & orig_pere_petitenf<=212) /*si seule

```

```

l'origine du père du petit-enfant de l'enquêté est renseignée et qu'il est
immigré (y compris originaire des DOM) (G1), on catégorise le petit-enfant de l'enquêté
comme enfant d'immigrés (G2) de même origine que son père*/
replace orig_petitenfenq = orig_pere_petitenfenq + 300 if (orig_pere_petitenfenq!=. &
orig_mere_petitenfenq==. & orig_pere_petitenfenq>=320 & orig_pere_petitenfenq<=412) /*si seule
l'origine du père du petit-enfant de l'enquêté est renseignée et qu'il est
enfant d'immigrés (y compris originaire des DOM) (G2), on catégorise le petit-enfant de
l'enquêté comme petit-enfant d'immigrés (G3) de même origine que son père*/
replace orig_petitenfenq = 504 if (orig_pere_petitenfenq!=. &
orig_mere_petitenfenq==. & orig_pere_petitenfenq==503) /*si seule
l'origine du père du petit-enfant de l'enquêté est renseignée et qu'il est petit-
enfant du groupe "majoritaire" , on catégorise le petit-enfant de
l'enquêté comme arrière-petit-enfant (G4) de même origine que son père*/
replace orig_petitenfenq = orig_pere_petitenfenq + 200 if (orig_pere_petitenfenq!=. &
orig_mere_petitenfenq==. & orig_pere_petitenfenq>=620 & orig_pere_petitenfenq<=712) /*si seule
l'origine du père du petit-enfant de l'enquêté est renseignée et qu'il est petit-
enfant d'immigrés (y compris originaire des DOM) (G3), on catégorise le petit-enfant de
l'enquêté comme arrière-petit-enfant d'immigrés (G4) de même origine que son père*/
replace orig_petitenfenq = orig_mere_petitenfenq + 200 if (orig_pere_petitenfenq==. &
orig_mere_petitenfenq!=. & orig_mere_petitenfenq>=120 & orig_mere_petitenfenq<=212) /*si seule
l'origine de la mère du petit-enfant de l'enquêté est renseignée et qu'elle est
immigrée (y compris originaire des DOM) (G1), on catégorise le petit-enfant de l'enquêté
comme enfant d'immigrés (G2) de même origine que sa mère*/
replace orig_petitenfenq = orig_mere_petitenfenq + 300 if (orig_pere_petitenfenq==. &
orig_mere_petitenfenq!=. & orig_mere_petitenfenq>=320 & orig_mere_petitenfenq<=412) /*si seule
l'origine de la mère du petit-enfant de l'enquêté est renseignée et qu'elle est
enfant d'immigrés (y compris originaire des DOM) (G2), on catégorise le petit-enfant de
l'enquêté comme petit-enfant d'immigrés (G3) de même origine que sa mère*/
replace orig_petitenfenq = 504 if (orig_pere_petitenfenq==. &
orig_mere_petitenfenq!=. & orig_mere_petitenfenq==503) /*si seule
l'origine de la mère du petit-enfant de l'enquêté est renseignée et qu'elle est petit-
enfant du groupe "majoritaire" , on catégorise le petit-enfant de
l'enquêté comme arrière-petit-enfant (G4) de même origine que sa mère*/
replace orig_petitenfenq = orig_mere_petitenfenq + 200 if (orig_pere_petitenfenq==. &
orig_mere_petitenfenq!=. & orig_mere_petitenfenq>=620 & orig_mere_petitenfenq<=712) /*si seule
l'origine de la mère du petit-enfant de l'enquêté est renseignée et qu'elle est petit-
enfant d'immigrés (y compris originaire des DOM) (G3), on catégorise le petit-enfant de
l'enquêté comme arrière-petit-enfant d'immigrés (G4) de même origine que sa mère*/
replace orig_petitenfenq = orig_pere_petitenfenq + 200 if (orig_pere_petitenfenq!=. &
orig_mere_petitenfenq!=. & orig_pere_petitenfenq>=120 & orig_pere_petitenfenq<=212) /*si
l'origine des 2 parents du petit-enfant de l'enquêté est renseignée et que son père est
immigré (y compris originaire des DOM) (G1), on catégorise le petit-enfant de l'enquêté
comme enfant d'immigrés (G2) de même origine que son père*/
replace orig_petitenfenq = orig_pere_petitenfenq + 300 if (orig_pere_petitenfenq!=. &
orig_mere_petitenfenq!=. & orig_pere_petitenfenq>=320 & orig_pere_petitenfenq<=412) /*si
l'origine des 2 parents du petit-enfant de l'enquêté est renseignée et que son père est
enfant d'immigrés (y compris originaire des DOM) (G2), on catégorise le petit-enfant de
l'enquêté comme petit-enfant d'immigrés (G3) de même origine que son père*/
replace orig_petitenfenq = 504 if (orig_pere_petitenfenq!=. &
orig_mere_petitenfenq!=. & orig_pere_petitenfenq==503) /*si
l'origine des 2 parents du petit-enfant de l'enquêté est renseignée et que son père est
petit-enfant du groupe "majoritaire" , on catégorise le petit-enfant de
l'enquêté comme arrière-petit-enfant (G4) de même origine que son père*/
replace orig_petitenfenq = orig_pere_petitenfenq + 200 if (orig_pere_petitenfenq!=. &
orig_mere_petitenfenq!=. & orig_pere_petitenfenq>=620 & orig_pere_petitenfenq<=712) /*si
l'origine des 2 parents du petit-enfant de l'enquêté est renseignée et que son père est
petit-enfant d'immigrés (y compris originaire des DOM) (G3), on catégorise le petit-enfant de
l'enquêté comme arrière-petit-enfant d'immigrés (G4) de même origine que son père*/
/*tab orig_petitenfenq, missing*/ /*Cette variable est renseignée pour n individus et elle
va de n à n*/
*Je calcule maintenant les REGIONNAISPERE_IND et REGIONNAISMERE_IND des petits-enfants de
l'enquêté
generate regionnaispere_petitenfenq = regionnais_parent1_petitenf if
(sexe_parent1_petitenf==1)
replace regionnaispere_petitenfenq = regionnais_parent2_petitenf if
(sexe_parent1_petitenf==2)
/*tab regionnaispere_petitenfenq, missing*/
generate regionnaismere_petitenfenq = regionnais_parent2_petitenf if
(sexe_parent1_petitenf==1)
replace regionnaismere_petitenfenq = regionnais_parent1_petitenf if
(sexe_parent1_petitenf==2)
tab regionnaismere_petitenfenq, missing
save petitsenf, replace
*Je réimporte ces variables dans REPRO
use repro
merge 1:1 ident noi using petitsenf, keepusing(orig_petitenfenq regionnaispere_petitenfenq
regionnaismere_petitenfenq) /*keep(match master)*/

```

```

/*il y a bien 105 appariements, c'est bon*/
drop _merge

*Je complète ORIG_IND avec ORIG_PETITENFENQ
replace orig_ind = orig_petitenfenq if (lienego_ind=="21") /*il y a
bien 105 changements, c'est bon*/
/*tab orig_ind lienego_ind, missing*/ /*restent
toujours 22 petits-enfants des enquêtés sans ORIG_IND attribué*/
*Je termine de compléter ORIG_IND (22 petits-enfants de l'enquêté vivant sans aucun de leurs
parents) avec les ORIG_IND de leurs oncles et tantes, probablement similaires aux ORIG_IND
d'un au moins de leurs parents; ou si nécessaire avec l'ORIG_IND de leur grand-parent enquêté
egen orig_oncle_tante_petitenfenq = mode(orig_enfenq), by(ident)
replace orig_ind = orig_oncle_tante_petitenfenq if (lienego_ind=="21" & orig_ind==.) /*il y a
21 changements*/
/*tab orig_ind, missing*/ /*reste
encore 1 petit-enfant des enquêtés sans ORIG_IND attribué*/
generate check_petitenf = 1 if (lienego_ind=="21" & orig_ind==.) /*ce
dernier petit-enfant sans ORIG_IND attribué est le petit-enfant d'un enquêté G1 autre Europe,
et comme c'est la seule information dont on dispose sur lui, je le catégorise comme G3 autre
Europe*/
egen orig_grandparent_petitenfenq = mode(orig_enq), by(ident)
replace orig_ind = orig_grandparent_petitenfenq + 500 if (lienego_ind=="21" & orig_ind==. &
orig_grandparent_petitenfenq>=120 & orig_grandparent_petitenfenq<=212) /*il y a bien 1
changement, c'est bon*/
*Je récapitule cette variable ORIG_IND
/*tab orig_ind, missing*/ /*Cette variable est renseignée pour 81 780 individus
et va de 120 à 849*/
/*tab lienego_ind orig_ind, missing*/ /*Cette variable est renseignée pour des personnes de
tous liens avec l'enquêté SAUF Autres liens familiaux, Liens familiaux indéterminés, Amis et
Autres liens non familiaux, ce qui est normal*/
/*ORIG_IND est (re)codée comme suit:
120=immigré (G1) originaire du Maghreb
129=immigré (G1) originaire d'autre Afrique
130=immigré (G1) originaire d'Asie (sauf Turquie)
134=immigré (G1) originaire de Turquie
140=immigré (G1) originaire d'Europe du Sud
149=immigré (G1) originaire d'autre Europe
150=immigré (G1) originaire d'ailleurs dans le monde
212=natif d'un DOM
320=enfant d'1 ou 2 immigrés (G2) originaire(s) du Maghreb
329=enfant d'1 ou 2 immigrés (G2) originaire(s) d'autre Afrique
330=enfant d'1 ou 2 immigrés (G2) originaire(s) d'Asie (sauf Turquie)
334=enfant d'1 ou 2 immigrés (G2) originaire(s) de Turquie
340=enfant d'1 ou 2 immigrés (G2) originaire(s) d'Europe du Sud
349=enfant d'1 ou 2 immigrés (G2) originaire(s) d'autre Europe
350=enfant d'1 ou 2 immigrés (G2) originaire(s) d'ailleurs dans le monde
412=enfant d'1 ou 2 natifs d'un DOM
500=groupe "majoritaire", sans ascendant immigré (G0)
501=groupe "majoritaire", sans ascendant immigré (G1)
502=groupe "majoritaire", sans ascendant immigré (G2)
503=groupe "majoritaire", sans ascendant immigré (G3)
504=groupe "majoritaire", sans ascendant immigré (G4)
620=petit-enfant d'1 à 4 immigrés (G3) originaire(s) du Maghreb
629=petit-enfant d'1 à 4 immigrés (G3) originaire(s) d'autre Afrique
630=petit-enfant d'1 à 4 immigrés (G3) originaire(s) d'Asie (sauf Turquie)
634=petit-enfant d'1 à 4 immigrés (G3) originaire(s) de Turquie
640=petit-enfant d'1 à 4 immigrés (G3) originaire(s) d'Europe du Sud
649=petit-enfant d'1 à 4 immigrés (G3) originaire(s) d'autre Europe
650=petit-enfant d'1 à 4 immigrés (G3) originaire(s) d'ailleurs dans le monde
712=petit-enfant d'1 à 4 natifs d'un DOM
820=arrière-petit-enfant d'1 à 8 immigrés (G4) originaire(s) du Maghreb
840=arrière-petit-enfant d'1 à 8 immigrés (G4) originaire(s) d'Europe du Sud
849=arrière-petit-enfant d'1 à 8 immigrés (G4) originaire(s) d'autre Europe
. =Non réponse*/

*J'applique des labels à la variable ORIG_IND
label variable orig_ind "origine"
label define orig 120 "G1 Maghreb" 129 "G1 autre Afrique" 130 "G1 Asie" 134 "G1 Turquie" 140
"G1 Europe Sud" 149 "G1 autre Europe" 150 "G1 ailleurs" 212 "G1 DOM" ///
320 "G2 Maghreb" 329 "G2 autre Afrique" 330 "G2 Asie" 334 "G2 Turquie" 340
"G2 Europe Sud" 349 "G2 autre Europe" 350 "G2 ailleurs" 390 "G2 ne sait pas d'où" 412 "G2 DOM"
///
500 "G0 Groupe majoritaire" 501 "G1 Groupe majoritaire" 502 "G2 Groupe
majoritaire" 503 "G3 Groupe majoritaire" 504 "G4 Groupe majoritaire" ///
620 "G3 Maghreb" 629 "G3 autre Afrique" 630 "G3 Asie" 634 "G3 Turquie" 640
"G3 Europe Sud" 649 "G3 autre Europe" 650 "G3 ailleurs" 712 "G3 DOM"
///

```

"G4 Europe Sud" 849 "G4 autre Europe"

```
label values orig_ind orig
/*tab orig_ind, missing*/
```

```
*Je complète REGIONNAISPERE_IND avec REGIONNAISPERE_PETITENFENQ
replace regionnaispere_ind = regionnaispere_petitenfenq if (lienego_ind=="21" &
regionnaispere_ind==.)
tab regionnaispere_ind, missing
*Je complète REGIONNAISMERE_IND avec REGIONNAISMERE_PETITENFENQ
replace regionnaismere_ind = regionnaismere_petitenfenq if (lienego_ind=="21" &
regionnaismere_ind==.)
tab regionnaismere_ind, missing
*Je récapitule ces variables REGIONNAISPERE_IND et REGIONNAISMERE_IND
tab regionnaispere_ind, missing /*Cette variable est renseignée pour 66 183
individus et va de 11 à 90*/
/*tab lienego_ind regionnaispere_ind, missing*/ /*Cette variable est renseignée pour des
personnes de tous liens avec l'enquête SAUF les parents de l'enquête et ses grands-parents et
beaux-parents, ainsi que les Autres liens familiaux, Liens familiaux indéterminés, Amis et
Autres liens non familiaux, ce qui est normal*/
tab regionnaismere_ind, missing /*Cette variable est renseignée pour 67
403 individus et va de 11 à 90*/
/*tab lienego_ind regionnaismere_ind, missing*/ /*Cette variable est renseignée pour des
personnes de tous liens avec l'enquête SAUF les parents de l'enquête et ses grands-parents et
beaux-parents, ainsi que les Autres liens familiaux, Liens familiaux indéterminés, Amis et
Autres liens non familiaux, ce qui est normal*/
/*REGIONNAISPERE_IND et REGIONNAISMERE_IND sont (re)codées comme suit:
11=France métropolitaine
12=DOM
20=Maghreb
29=autre Afrique
30=Asie
34=Turquie
40=Europe du Sud
49=autre Europe
50=Ailleurs dans le monde
90=NSP
.=Non réponse*/
```

```
*J'applique des labels à la variable LIENEGO_IND (maintenant seulement, car Stata ne permet
pas d'appliquer de labels à des variables string)
destring lienego_ind, replace
label variable lienego_ind "lien avec l'enquête"
label define lien 0 "Enquête" 1 "Conjoint" 2 "Enfant cohabitant" 3 "Parent" 10 "Frère,
sœur" 200 "Enfant décédé non cohabitant" 201 "Enfant en vie non cohabitant" 21 "Petit-enfant"
22 "Grand-parent" 31 "Beau-fils, belle-fille" 32 "Beau-parent" 40 "Autre lien familial" 50
"Lien familial indéterminé" 60 "Ami" 90 "Autre lien non familial"
label values lienego_ind lien
tab lienego_ind, missing
```

```
*Je supprime les bases de données séparées
erase petitsenf.dta
erase famille_petitsenf.dta
```

```
*Je réordonne dans la table REPRO ces 24 variables (+ 2 identifiants) dans le bon ordre, celui
du Tableau 5.1
order ident noi poids_ind sexe_ind anais_ind prenom_ind lienego_ind regionnaiss_ind
residence_ind group2_ind regionnaispere_ind regionnaismere_ind orig_ind autrepereparent natpniv2
natmniv2 n_indper n_indmer lnp1_gr lnp2_gr lnm1_gr lnm2_gr relig_ind r_impvie orig_relig_ind
x_vufri
```

```
*Je range dans la table REPRO les 83 045 individus dans le bon ordre
sort ident noi
```

/\*Chapitre 5: L'échantillon retenu\*/

```
*Effectif renseigné et nb de prénoms différents du champ "maximal"
tab lienego_ind, missing /*N = 83 045*/
codebook prenom_ind /*N = 18 161*/
```

```
*Champ qui s'approche le plus du champ représentatif des habitants de 0-50 / 0-60 ans en
France métropolitaine en 2008
generate bon_champ = 1 if (lienego_ind==0)
/*enquête,
N = 21 761*/
replace bon_champ = 1 if (lienego_ind==2 | lienego_ind==201) & (anais_ind<=1990) &
```

```

(residence_ind==1) & (orig_enq==320 | orig_enq==329 | orig_enq==330 | orig_enq==334 |
orig_enq==340 | orig_enq==349 | orig_enq==350) /*enfant d'enquête qui est vivant, majeur et
habite en métropole en 2008, l'enquête étant lui-même enfant d'immigrés de
18-50 ans, N = 1 043*/
replace bon_champ = 1 if (lienego_ind==2 | lienego_ind==201) & (anais_ind<=1990) &
(residence_ind==1) & (orig_enq==412)
/*enfant d'enquête qui est vivant, majeur et habite en métropole en 2008, l'enquête étant lui-
même enfant de personnes nées dans un DOM de 18-50 ans, N = 38*/
replace bon_champ = 1 if (lienego_ind==2 | lienego_ind==201) & (anais_ind >1990) &
(residence_ind==1) & (orig_enq!=502)
/*enfant d'enquête qui est vivant, mineur et habite en métropole en 2008, l'enquête étant lui-
même "minoritaire", N = 16 710*/
replace bon_champ = 1 if (lienego_ind==2 | lienego_ind==201) & (anais_ind >1990) &
(residence_ind==1) & (orig_enq==502) & (orig_ind==503)
/*enfant d'enquête qui est vivant, mineur et habite en métropole en 2008, l'enquête étant lui-
même "majoritaire" et l'autre parent aussi, N = 2 656*/
tab lienego_ind if (bon_champ==1), missing /*le bon champ compte 42 208 individus*/
codebook prenom_ind if (bon_champ==1) /*le bon champ compte 42 208 individus qui
portent 11 151 prénoms différents*/

```

/\*Chapitre 5: La pondération des individus\*/

\*Pour chacune des variables ci-dessous, la pondération ne change pas la somme des individus du champ maximal (83 045) mais seulement la distribution de leurs modalités de réponse.

```

tab sexe_ind [aw=poids_ind], missing /*Sur données pondérées, on compte
49,77% d'hommes, 50,22% de femmes et 0,01% de non réponse.*/
/*tab anais_ind [aw=poids_ind], missing*/ /*Sur données pondérées, on compte
0,04 % de non réponse*/

```

```

ssc install groups
groups prenom_ind [aw=poids_ind], order(h) select(5) /*Sur données pondérées, les 5
prénoms les plus fréquents sont Sébastien, Michel, Nathalie, Catherine et Frédéric, portés par
4,05% des individus du champ maximal.*/

```

```

tab lienego_ind [aw=poids_ind], missing
tab regionnais_ind [aw=poids_ind], missing
tab group2_ind [aw=poids_ind], missing
tab regionnaispere_ind [aw=poids_ind], missing
tab regionnaismere_ind [aw=poids_ind], missing
/*tab orig_ind [aw=poids_ind], missing*/
tab autrepere [aw=poids_ind], missing
tab natpniv2 [aw=poids_ind], missing
/*tab natmniv2 [aw=poids_ind], missing*/
tab n_indper [aw=poids_ind], missing
tab n_indmer [aw=poids_ind], missing
tab lnp1_gr [aw=poids_ind], missing
tab lnp2_gr [aw=poids_ind], missing
tab lnm1_gr [aw=poids_ind], missing
tab lnm2_gr [aw=poids_ind], missing
tab relig_ind [aw=poids_ind], missing
tab r_impvie [aw=poids_ind], missing
tab orig_relig_ind [aw=poids_ind], missing
tab x_vufri [aw=poids_ind], missing

```

/\*Chapitre 5: Vérification de la validité "interne" de la table REPRO (par rapport à l'enquête TeO)\*/

\*Je vérifie que REPRO contient bien 83 045 individus et les 24 variables (+ 2 identifiants) en colonnes

```

codebook ident noi poids_ind sexe_ind anais_ind prenom_ind lienego_ind regionnais_ind
residence_ind group2_ind regionnaispere_ind regionnaismere_ind orig_ind autrepere natpniv2
natmniv2 n_indper n_indmer lnp1_gr lnp2_gr lnm1_gr lnm2_gr relig_ind r_impvie orig_relig_ind
x_vufri /*c'est bon*/

```

\*J'ai déjà vérifié si les variables de vérification sont cohérentes entre elles:

```

/* - les valeurs des variables de pondération sont les mêmes (pour tout individu) que
ces variables proviennent des tables THL2_PRENOM (POIDSI) , INDIV2_PRENOM (POIDSI) ou
ENFANTS2_PRENOM (POIDSE) ; cf. ci-dessus variable incoherencepoids
- les valeurs des variables de sexe sont les mêmes (pour tout individu) que
ces variables proviennent des tables THL2_PRENOM (SEXE) , INDIV2_PRENOM (SEXEE) ou
ENFANTS2_PRENOM (E_SEXE) ; cf. ci-dessus variable incoherencesexe
- les valeurs des variables d'année de naissance sont les mêmes (pour tout individu) que
ces variables proviennent des tables THL2_PRENOM (ANAIIS) , INDIV2_PRENOM (ANAISE) ou
ENFANTS2_PRENOM (E_ANAIS) ; cf. ci-dessus variable incoherenceanais
- les valeurs des variables de prénom sont les mêmes (pour tout individu) que
ces variables proviennent des tables THL2_PRENOM (PRENOM) , INDIV2_PRENOM (PRENOME) ou
ENFANTS2_PRENOM (E_PRENOM) ; cf. ci-dessus variable incoherenceprenom

```

```

- les valeurs des variables de lieu de naissance sont les mêmes (pour tout individu) que
ces variables proviennent des tables THL2_PRENOM (REGIONNAIS2) ou
ENFANTS2_PRENOM (REGIONNAISENF2); cf. ci-dessus variable incoherenceregionnais*/

*J'ai déjà vérifié si les variables renseignent le bon nombre d'individus du bon type:
/* - les variables calculables pour tous les individus sont bien renseignées pour jusqu'à 83
045 individus (POIDS_IND, SEXE_IND, ANAIS_IND, PRENOM_IND, LIENEGO_IND, REGIONNAIS_IND,
ORIG_IND)
- les variables calculables pour les seuls enquêtés sont bien renseignées pour jusqu'à 21
761 enquêtés (NATPNIV2, NATMNIV2, N_INDPER, N_INDMER, LNP1_GR, LNP2_GR, LNM1_GR, LNM2_GR,
R_IMPVIE, X_VUFRI)
- les variables calculables pour les enquêtés et seulement certains de leurs apparentés
sont bien renseignées pour entre 21 761 et 83 045 individus, de 35 004 pour celle renseignée
pour les enquêtés et leurs conjoints actuels (GROUP2_IND) jusqu'à 67 403 pour celles
renseignées pour la plupart des apparentés des enquêtés (REGIONNAISPERE_IND,
REGIONNAISMERE_IND), en passant par 42 651 pour celle renseignée pour les enquêtés et leurs
conjoints actuels ainsi que les parents des enquêtés (RELIG_IND), et 59 082 pour celle
renseignée pour les enquêtés et leurs enfants ainsi que les frères et soeurs des enquêtés
(ORIG_RELIG_IND)*/

*Je vérifie si REPRO permet de reproduire à l'identique (à l'unité d'effectif non pondéré
près, et à la décimale de pourcentage d'effectif pondéré près) les tris à plat présentés dans
le dictionnaire des codes (Insee 2011):
*la répartition par sexe des individus de la table THL2_PRENOM (Insee 2011, p. 444)
  tab sexe_ind if (lienego_ind!=200 & lienego_ind!=201)
, missing /*c'est bon*/
  tab sexe_ind if (lienego_ind!=200 & lienego_ind!=201) [aw=poids_i]
, missing /*c'est bon*/
*la répartition par sexe des enquêtés de la table INDIV2_PRENOM (Insee 2011, p. 445)
  tab sexe_ind if (lienego_ind==0)
, missing /*c'est bon*/
  tab sexe_ind if (lienego_ind==0)
[aw=poids_ind], missing /*c'est bon*/
*la répartition par sexe des individus de la table ENFANTS2_PRENOM (Insee 2011, p. 209)
  tab sexe_ind if (lienego_ind==2 | lienego_ind==200 | lienego_ind==201)
, missing /*c'est bon*/
  tab sexe_ind if (lienego_ind==2 | lienego_ind==200 | lienego_ind==201)
[aw=poids_ind], missing /*c'est bon*/
*la répartition par année de naissance des individus de la table THL2_PRENOM (Insee 2011, p.
93)
  summarize anais_ind if (lienego_ind!=200 & lienego_ind!=201)
, detail /*c'est bon*/
  summarize anais_ind if (lienego_ind!=200 & lienego_ind!=201) [aw=poids_i]
, detail /*c'est bon*/
*la répartition par année de naissance des enquêtés de la table INDIV2_PRENOM (Insee 2011, p.
93)
  summarize anais_ind if (lienego_ind==0)
, detail /*c'est bon*/
  summarize anais_ind if (lienego_ind==0)
[aw=poids_ind], detail /*c'est PAS bon: la "moyenne pondérée" présentée en bas de la p. 93
n'est en fait pas pondérée*/
*la répartition par année de naissance des individus de la table ENFANTS2_PRENOM (Insee 2011,
p. 205)
  summarize anais_ind if (lienego_ind==2 | lienego_ind==200 | lienego_ind==201)
, detail /*c'est bon*/
  summarize anais_ind if (lienego_ind==2 | lienego_ind==200 | lienego_ind==201)
[aw=poids_ind], detail /*on ne peut pas savoir si c'est bon: p. 205, les infos pondérées ne
sont pas présentées*/
*la répartition par lien à l'enquêté des individus de la table THL2_PRENOM (Insee 2011, p.
291)
  tab lienego_ind if (lienego_ind!=200 & lienego_ind!=201)
, missing /*c'est bon*/
  tab lienego_ind if (lienego_ind!=200 & lienego_ind!=201) [aw=poids_i]
, missing /*c'est bon*/
*la répartition par lien à la migration des enquêtés de la table INDIV2_PRENOM (Insee 2011, p.
241-243)
  tab group2 if (lienego_ind==0)
, missing /*c'est bon*/
  tab group2 if (lienego_ind==0)
[aw=poids_ind], missing /*c'est bon*/
  tab group2ca if (lienego_ind==0)
, missing /*c'est bon*/
  tab group2ca if (lienego_ind==0)
[aw=poids_ind], missing /*c'est bon*/
  tab group2pc if (lienego_ind==0)
, missing /*c'est bon*/
  tab group2pc if (lienego_ind==0)

```

```

[aw=poids_ind], missing /*c'est bon*/
*la répartition par choix de nationalité du parent si né dans une ancienne colonie française
de la table INDIV2_PRENOM (Insee 2011, p. 316)
  tab n_indper if (lienego_ind==0)
, missing /*c'est bon*/
  tab n_indper if (lienego_ind==0)
[aw=poids_ind], missing /*c'est bon*/
  tab n_indmer if (lienego_ind==0)
, missing /*c'est bon*/
/*tab n_indmer if (lienego_ind==0)
[aw=poids_ind], missing /*c'est bon*/
*la répartition par religiosité des enquêtés de la table INDIV2_PRENOM (Insee 2011, p. 419)
  tab r_impvie if (lienego_ind==0)
, missing /*c'est bon*/
  tab r_impvie if (lienego_ind==0)
[aw=poids_ind], missing /*c'est bon*/
*la répartition par opinion sur "On me voit comme un Français" des enquêtés de la table
INDIV2_PRENOM (Insee 2011, p. 491)
  tab x_vufri if (lienego_ind==0)
, missing /*c'est bon*/
  tab x_vufri if (lienego_ind==0)
[aw=poids_ind], missing /*c'est bon*/

*Les seules modalités de réponse imprévues ou bizarres concernent certains prénoms
  sort prenom_ind
*Ces modalités bizarres sont, pour environ 150 d'entre elles, placées en début et en fin de
fichier, mais aussi au début et à la fin de chaque le lettre du prénom; je les traite plus
loin

*Je vérifie si ORIG_IND, absente du dictionnaire des codes, permet de reproduire à l'identique
(à l'unité d'effectif non pondéré près, et à la décimale de pourcentage d'effectif pondéré
près) les tris à plat présentés dans le dictionnaire des codes sur un champ partiel:
*la répartition par GROUP1 des enquêtés (Insee 2011, p. 239)
/*tab orig_ind if (lienego_ind==0), missing*/ /*c'est bon, à ceci près que
orig_ind catégorise comme "Non réponse" 1 individu que GROUP1 catégorise comme enfant d'1 ou 2
immigrés (parce qu'on ignore son origine géographique)*/
/*tab orig_ind if (lienego_ind==0) [aw=poids_ind], missing*/ /*c'est bon*/
*Je vérifie si ORIG_IND est liée à LIENEGO_IND de façon logique
/*tab orig_ind lienego_ind, missing*/ /*c'est bon*/

*Tableau 5.3. Distribution de la variable d'origine des individus (ORIG_IND) (enquête TeO), N
= 42 208
  tab orig_ind if (bon_champ==1) , missing /*données non pondérées [seul tableau
d'effectifs publié dont au moins 1 modalité ne comprend qu'1 individu]*/
  tab orig_ind if (bon_champ==1) [aw=poids_ind], missing /*données pondérées [seul tableau
d'effectifs publié dont au moins 1 modalité ne comprend qu'1 individu]*/

/*Chapitre 5: Analyses préliminaires sur la table REPRO*/

*LIENEGO_IND
tab lienego_ind if (bon_champ==1) [aw=poids_ind], missing /*les 42 208 habitants
comprennent 74 % d'enquêtés, 24 % d'enfants cohabitants, et 2 % d'enfants non cohabitants*/
tab lienego_ind sexe_ind if (bon_champ==1) [aw=poids_ind], missing /*idem par sexe: pas de
surprise*/

*SEXE_IND
tab sexe_ind if (bon_champ==1) [aw=poids_ind], missing /*les 42 208 habitants
comprennent 49,54 % d'hommes et 50,46 % de femmes*/

*ANAIIS_IND
summarize anais_ind if (bon_champ==1) [aw=poids_ind], detail /*les 42 208 habitants sont
nés de 1948 à 2009, en moyenne vers 1976*/
generate quinquennais_ind = anais_ind
  recode quinquennais_ind (1900/1904=1900) (1905/1909=1905) (1910/1914=1910) (1915/1919=1915)
(1920/1924=1920) (1925/1929=1925) (1930/1934=1930) (1935/1939=1935) (1940/1944=1940)
(1945/1949=1945) (1950/1954=1950) (1955/1959=1955) (1960/1964=1960) (1965/1969=1965)
(1970/1974=1970) (1975/1979=1975) (1980/1984=1980) (1985/1989=1985) (1990/1994=1990)
(1995/1999=1995) (2000/2004=2000) (2005/2009=2005)
  tab quinquennais_ind, missing /*comme anais_ind, cette
variable est renseignée pour 82 961 individus (on compte 84 Non réponse); elle va de 1900-1904
à 2005-2009*/

*Figure 5.4. Pyramide des âges des personnes des 18-50 / 18-60 ans en ménage ordinaire et de
leurs enfants de 0-17 ans habitant en France métropolitaine en 2008, N = 42 208
  tab quinquennais_ind sexe_ind if (bon_champ==1) [aw=poids_ind], missing /*je mets ce
résultat sous Excel pour construire la pyramide des âges*/

```





```

1 seul espace*/
replace prenom_ind = strtrim(prenom_ind) /*on supprime les espaces initiaux et
finaux*/
sort prenom_ind

generate nombredansprenom = 0 if strmatch(prenom_ind,"*0*") /*on remplace ci-dessous certains
de ces chiffres 0 par la lettre O*/
replace nombredansprenom = 1 if strmatch(prenom_ind,"*1*") /*on remplace ci-dessous certains
de ces chiffres 1 par la lettre I*/
replace nombredansprenom = 2 if strmatch(prenom_ind,"*2*")
replace nombredansprenom = 3 if strmatch(prenom_ind,"*3*")
replace nombredansprenom = 4 if strmatch(prenom_ind,"*4*")
replace nombredansprenom = 5 if strmatch(prenom_ind,"*5*")
replace nombredansprenom = 6 if strmatch(prenom_ind,"*6*")
replace nombredansprenom = 7 if strmatch(prenom_ind,"*7*")
replace nombredansprenom = 8 if strmatch(prenom_ind,"*8*")
replace nombredansprenom = 9 if strmatch(prenom_ind,"*9*")
tab nombredansprenom, missing
sort nombredansprenom /*j'inclus les corrections ci-
dessous, parmi d'autres corrections réalisées après examen manuel des prénoms non catégorisés
de la table REPRO*/

replace prenom_ind = subinstr(prenom_ind,"1BDOURAMANE","ABDOURAMANE",.)
replace prenom_ind = subinstr(prenom_ind,"ADAM1","ADAM",.)
replace prenom_ind = subinstr(prenom_ind,"ADA0","ADAO",.)
replace prenom_ind = subinstr(prenom_ind,"A◇CHA","AICHA",.) /*ne marche pas*/
replace prenom_ind = subinstr(prenom_ind,"A◇DA","AIDA",.) /*ne marche pas*/
replace prenom_ind = subinstr(prenom_ind,"A◇DAN","AIDAN",.) /*ne marche pas*/
replace prenom_ind = subinstr(prenom_ind,"A◇MEN","AIMEN",.) /*ne marche pas*/
replace prenom_ind = subinstr(prenom_ind,"A◇SSA","AISSA",.) /*ne marche pas*/
replace prenom_ind = subinstr(prenom_ind,"AI SSATA","AISSATA",.)
replace prenom_ind = subinstr(prenom_ind,"AN DREAS","ANDREAS",.)
replace prenom_ind = subinstr(prenom_ind,"1NNISSA","ANNISSA",.)
replace prenom_ind = subinstr(prenom_ind,"ASSE◇TOU","ASSETOU",.) /*ne marche pas*/
replace prenom_ind = subinstr(prenom_ind,"B APTISTE","BAPTISTE",.)
replace prenom_ind = subinstr(prenom_ind,"B ELINDA","BELINDA",.)
replace prenom_ind = subinstr(prenom_ind,"B IRAHIMA","IBRAHIMA",.)
replace prenom_ind = subinstr(prenom_ind,"BUSRA|","BUSRA",.)
replace prenom_ind = subinstr(prenom_ind,"CAMILLE 2","CAMILLE",.)
replace prenom_ind = subinstr(prenom_ind,"CARIN1A","CARINIA",.)
replace prenom_ind = subinstr(prenom_ind,"2CATARINA","CATARINA",.)
replace prenom_ind = subinstr(prenom_ind,"C◇ELINE","CELINE",.) /*ne marche pas*/
replace prenom_ind = subinstr(prenom_ind,"CHLO◇","CHLOE",.) /*ne marche pas*/
replace prenom_ind = subinstr(prenom_ind,"CHR ISTOPHE","CHRISTOPHE",.)
replace prenom_ind = subinstr(prenom_ind,"CHRISTO|HE","CHRISTOPHE",.)
replace prenom_ind = subinstr(prenom_ind,"CLAUDINE","CLAUDINE",.)
replace prenom_ind = subinstr(prenom_ind,"C◇ME","COME",.) /*ne marche pas*/
replace prenom_ind = subinstr(prenom_ind,"CORINNNE","CORINNE",.)
replace prenom_ind = subinstr(prenom_ind,"DAN IELE","DANIELE",.)
replace prenom_ind = subinstr(prenom_ind,"DAV ID","DAVID",.)
replace prenom_ind = subinstr(prenom_ind,"DEMBO","DEMBO",.)
replace prenom_ind = subinstr(prenom_ind,"DIDIER4","DIDIER",.)
replace prenom_ind = subinstr(prenom_ind,"FABRICE2","FABRICE",.)
replace prenom_ind = subinstr(prenom_ind,"FATI1A","FATIHA",.)
replace prenom_ind = subinstr(prenom_ind,"F3ERNANDO","FERNANDO",.)
replace prenom_ind = subinstr(prenom_ind,"GAB RIEL","GABRIEL",.)
replace prenom_ind = subinstr(prenom_ind,"GUILLAUME 1","GUILLAUME",.)
replace prenom_ind = subinstr(prenom_ind,"GUILLAUME","GUILLAUME",.)
replace prenom_ind = subinstr(prenom_ind,"GGWENAELE","GWENAELE",.)
replace prenom_ind = subinstr(prenom_ind,"GWENN AEL","GWENNAEL",.)
replace prenom_ind = subinstr(prenom_ind,"GWENA LLE","GWENAELE",.)
replace prenom_ind = subinstr(prenom_ind,"HAN1","HANI",.)
replace prenom_ind = subinstr(prenom_ind,"ISMA◇L","ISMAEL",.) /*ne marche pas*/
replace prenom_ind = subinstr(prenom_ind,"ISS1A","ISSA",.)
replace prenom_ind = subinstr(prenom_ind,"J BAPTISTE","JEAN BAPTISTE",.)
replace prenom_ind = subinstr(prenom_ind,"J CLAUDE","JEAN CLAUDE",.)
replace prenom_ind = subinstr(prenom_ind,"J FRANCOIS","JEAN FRANCOIS",.)
replace prenom_ind = subinstr(prenom_ind,"J JACQUES","JEAN JACQUES",.)
replace prenom_ind = subinstr(prenom_ind,"J MICHEL","JEAN MICHEL",.)
replace prenom_ind = subinstr(prenom_ind,"J NOEL","JEAN NOEL",.)
replace prenom_ind = subinstr(prenom_ind,"J PAUL","JEAN PAUL",.)
replace prenom_ind = subinstr(prenom_ind,"J PHI","JEAN PHILIPPE",.)
replace prenom_ind = subinstr(prenom_ind,"J PIERRE","JEAN PIERRE",.)
replace prenom_ind = subinstr(prenom_ind,"JEA N PIERRE","JEAN PIERRE",.)
replace prenom_ind = subinstr(prenom_ind,"JEAN MARC 1","JEAN MARC",.)

```

```

replace prenom_ind = subinstr(prenom_ind,"JOSEPH1","JOSEPH",.)
replace prenom_ind = subinstr(prenom_ind,"JUL IE","JULIE",.)
replace prenom_ind = subinstr(prenom_ind,"JOEL","JOEL",.) /*ne marche pas*/
replace prenom_ind = subinstr(prenom_ind,"KARIM1","KARIM",.)
replace prenom_ind = subinstr(prenom_ind,"KADIATOU2","KADIATOU",.)
replace prenom_ind = subinstr(prenom_ind,"KIMLY2","KIMLYZ",.)
replace prenom_ind = subinstr(prenom_ind,"LAURIN E","LAURINE",.)
replace prenom_ind = subinstr(prenom_ind,"MAREV A","MAREVA",.)
replace prenom_ind = subinstr(prenom_ind,"MARIA 1","MARIA",.)
replace prenom_ind = subinstr(prenom_ind,"LMARINA","MARINA",.)
replace prenom_ind = subinstr(prenom_ind,"1MATHIAS","MATHIAS",.)
replace prenom_ind = subinstr(prenom_ind,"MEL1","MELI",.)
replace prenom_ind = subinstr(prenom_ind,"MES*OUDA","MESSOUDA",.)
replace prenom_ind = subinstr(prenom_ind,"MICKAEL","MICKAEL",.) /*ne marche pas*/
replace prenom_ind = subinstr(prenom_ind,"MON IQUE","MONIQUE",.)
replace prenom_ind = subinstr(prenom_ind,"NAB ILA","NABILA",.)
replace prenom_ind = subinstr(prenom_ind,"NATHALLIE","NATHALIE",.)
replace prenom_ind = subinstr(prenom_ind,"NOEMIE","NOEMIE",.) /*ne marche pas*/
replace prenom_ind = subinstr(prenom_ind,"OGUZHAN","OGUZHAN",.)
replace prenom_ind = subinstr(prenom_ind,"PATRICK","PATRICK",.) /*ne marche pas*/
replace prenom_ind = subinstr(prenom_ind,"PHILIPP","PHILIPPE",.) /*ne marche pas*/
replace prenom_ind = subinstr(prenom_ind,"QUUANG6DRI","QUUANG DRI",.)
replace prenom_ind = subinstr(prenom_ind,"RACHELLE","RACHELLE",.)
replace prenom_ind = subinstr(prenom_ind,"PATRICE","PATRICE",.)
replace prenom_ind = subinstr(prenom_ind,"RICARDO","RICARDO",.)
replace prenom_ind = subinstr(prenom_ind,"SAID","SAID",.) /*ne marche pas*/
replace prenom_ind = subinstr(prenom_ind,"SANDRINE 090199","SANDRINE",.)
replace prenom_ind = subinstr(prenom_ind,"YOUTHYE1","YOUTHYEI",.)
replace prenom_ind = subinstr(prenom_ind,"SANDY2","SANDY",.)
replace prenom_ind = subinstr(prenom_ind,"SERIMAN1","SERIMANI",.)
replace prenom_ind = subinstr(prenom_ind,"THOMAS1SQ","THOMAS",.)
replace prenom_ind = subinstr(prenom_ind,"VERONIQUE","VERONIQUE",.)
replace prenom_ind = subinstr(prenom_ind,"VIRGINIE","VIRGINIE",.)
replace prenom_ind = subinstr(prenom_ind,"VIRGINIE 201019","VIRGINIE",.)
replace prenom_ind = subinstr(prenom_ind,"WILLIAM","WILLIAM",.)
replace prenom_ind = subinstr(prenom_ind,"YAELE","YAELE",.) /*ne marche pas*/
replace prenom_ind = subinstr(prenom_ind,"YOUTHYE1","YOUTHYE",.)
replace prenom_ind = subinstr(prenom_ind,"ZOHRA","ZOHRA",.)
save repro, replace

*Je nettoie de façon harmonisée les prénoms de la table INSEE
use insee, clear
rename sexe sexe_ind
tab sexe_ind, missing /*le Fichier INSEE des prénoms (1900-
2019) comprend 297 223 prénoms-année de garçons et 354 833 prénoms-année de filles*/
rename preusuel prenom_ind
codebook prenom_ind /*le Fichier INSEE des prénoms (1900-
2019) comprend 34 264 prénoms-sexe différents, sur 652 056 prénoms-sexe-année différents*/
rename annais annais_ind
replace annais_ind = "" if (annais_ind=="XXXX")
destring annais_ind, replace
tab annais_ind, missing /*le Fichier INSEE des prénoms (1900-
2019) comprend de 1 708 prénoms-sexe en 1900 à 13 140 prénoms-sexe en 2019, + 35 864 prénoms-
sexe d'année de naissance inconnue, soit au total 652 056 prénoms-sexe-année*/
generate decennais_ind = annais_ind
recode decennais_ind (1900/1909=1900) (1910/1919=1910) (1920/1929=1920) (1930/1939=1930)
(1940/1949=1940) (1950/1959=1950) (1960/1969=1960) (1970/1979=1970) (1980/1989=1980)
(1990/1999=1990) (2000/2009=2000) (2010/2019=2010) (.=)
tab decennais_ind, missing /*Comme annais_ind, cette variable est renseignée pour 652 056
prénoms-sexe-année; elle va de la décennie 1900 à la décennie 2010*/
/*tab nombre, missing*/

replace prenom_ind = subinstr(prenom_ind,"Ã","A",.) /*on remplace les signes qui ont
remplacé le A majuscule avec accent circonflexe par un A majuscule tout simple*/
replace prenom_ind = subinstr(prenom_ind,"Ä","A",.) /*on remplace les signes qui ont
remplacé le A majuscule avec tréma par un A majuscule tout simple*/
replace prenom_ind = subinstr(prenom_ind,"Æ","A",.) /*on remplace les signes qui ont
remplacé le A majuscule avec accent grave par un A majuscule tout simple*/
replace prenom_ind = subinstr(prenom_ind,"À","C",.) /*on remplace les signes qui ont
remplacé le C majuscule avec cédille par un C majuscule tout simple*/
replace prenom_ind = subinstr(prenom_ind,"Ä","E",.) /*on remplace les signes qui ont
remplacé le E majuscule avec accent aigu par un E majuscule tout simple*/
replace prenom_ind = subinstr(prenom_ind,"Å","E",.) /*on remplace les signes qui ont
remplacé le E majuscule avec accent grave par un E majuscule tout simple*/
replace prenom_ind = subinstr(prenom_ind,"Š","E",.) /*on remplace les signes qui ont
remplacé le E majuscule avec accent circonflexe par un E majuscule tout simple*/

```

```

    replace prenom_ind = subinstr(prenom_ind,"Ä","E",.) /*on remplace les signes qui ont
remplacé le E majuscule avec tréma par un E majuscule tout simple*/
    replace prenom_ind = subinstr(prenom_ind,"Å","O",.) /*on remplace les signes qui ont
remplacé le O majuscule avec accent circonflexe par un O majuscule tout simple*/
    replace prenom_ind = subinstr(prenom_ind,"Ä-","O",.) /*on remplace les signes qui ont
remplacé le O majuscule avec tréma par un O majuscule tout simple*/
    replace prenom_ind = subinstr(prenom_ind,"Ä","U",.) /*on remplace les signes qui ont
remplacé le U majuscule avec accent aigu par un U majuscule tout simple*/
    replace prenom_ind = subinstr(prenom_ind,"Ä","U",.) /*on remplace les signes qui ont
remplacé le U majuscule avec accent circonflexe par un U majuscule tout simple*/
    replace prenom_ind = subinstr(prenom_ind,"Äæ","U",.) /*on remplace les signes qui ont
remplacé le U majuscule avec tréma par un U majuscule tout simple*/
    replace prenom_ind = subinstr(prenom_ind,"Ž","I",.) /*on remplace les signes qui ont
remplacé le I majuscule avec accent circonflexe par un I majuscule tout simple*/
    replace prenom_ind = subinstr(prenom_ind,"Ä","I",.) /*on remplace les signes qui ont
remplacé le I majuscule avec tréma par un I majuscule tout simple; on le fait en
dernier pour éviter les erreurs*/
    replace prenom_ind = subinstr(prenom_ind,"I","I",.) /*on remplace les signes qui ont
remplacé le I majuscule avec tréma par un I majuscule tout simple*/
    replace prenom_ind = subinstr(prenom_ind,"'", "'",.) /*on remplace les apostrophes
par des espaces, comme dans TeO*/
    replace prenom_ind = subinstr(prenom_ind,"-"," ",.) /*on remplace les traits d'union
par des espaces, comme dans TeO*/

save insee, replace

```

\*Tableau 5.5. Top 3 des prénoms par sexe et par décennie de naissance, selon l'enquête TeO (2008), N = 32 022, et selon le fichier INSEE des prénoms (1900-2019), N = 85 047 407  
use repro

```

generate decennais_ind = anais_ind
    recode decennais_ind (1900/1909=1900) (1910/1919=1910) (1920/1929=1920) (1930/1939=1930)
(1940/1949=1940) (1950/1959=1950) (1960/1969=1960) (1970/1979=1970) (1980/1989=1980)
(1990/1999=1990) (2000/2009=2000) (.=.)
    tab decennais_ind, missing /*Comme anais_ind, cette variable est renseignée pour 82 961
individus (on compte 84 Non réponse); elle va de la décennie 1900 à la décennie 2000*/
    tab sexe_ind if bon_champ==1 & (sexe_ind==1 | sexe_ind==2) & (regionnais_ind==11 |
regionnais_ind==12) & (decennais_ind>=1950 & decennais_ind<=2000) [aw=poids_ind], missing
/*les habitants qui sont nés en France entière des années 1950 aux années 2000 sont N = 32
022*/
    groups prenom_ind if bon_champ==1 & decennais_ind==1950 & sexe_ind==1 & (regionnais_ind==11
| regionnais_ind==12) [aw=poids_ind], order(h) select(3) /*les hommes qui sont nés en France
entière de 1950 à 1959 ont comme 3 prénoms les plus fréquents Patrick, Michel, Christian*/
    groups prenom_ind if bon_champ==1 & decennais_ind==1950 & sexe_ind==2 & (regionnais_ind==11
| regionnais_ind==12) [aw=poids_ind], order(h) select(3) /*les femmes qui sont nées en France
entière de 1950 à 1959 ont comme 3 prénoms les plus fréquents Martine, Françoise, Catherine*/
    groups prenom_ind if bon_champ==1 & decennais_ind==1960 & sexe_ind==1 & (regionnais_ind==11
| regionnais_ind==12) [aw=poids_ind], order(h) select(3) /*les hommes qui sont nés en France
entière de 1960 à 1969 ont comme 3 prénoms les plus fréquents Philippe, Eric, Thierry*/
    groups prenom_ind if bon_champ==1 & decennais_ind==1960 & sexe_ind==2 & (regionnais_ind==11
| regionnais_ind==12) [aw=poids_ind], order(h) select(3) /*les femmes qui sont nées en France
entière de 1960 à 1969 ont comme 3 prénoms les plus fréquents Catherine, Isabelle, Sylvie*/
    groups prenom_ind if bon_champ==1 & decennais_ind==1970 & sexe_ind==1 & (regionnais_ind==11
| regionnais_ind==12) [aw=poids_ind], order(h) select(3) /*les hommes qui sont nés en France
entière de 1970 à 1979 ont comme 3 prénoms les plus fréquents Stéphane, Frédéric, Sébastien*/
    groups prenom_ind if bon_champ==1 & decennais_ind==1970 & sexe_ind==2 & (regionnais_ind==11
| regionnais_ind==12) [aw=poids_ind], order(h) select(3) /*les femmes qui sont nées en France
entière de 1970 à 1979 ont comme 3 prénoms les plus fréquents Sandrine, Nathalie, Stéphanie*/
    groups prenom_ind if bon_champ==1 & decennais_ind==1980 & sexe_ind==1 & (regionnais_ind==11
| regionnais_ind==12) [aw=poids_ind], order(h) select(3) /*les hommes qui sont nés en France
entière de 1980 à 1989 ont comme 3 prénoms les plus fréquents Sébastien, Julien, Nicolas*/
    groups prenom_ind if bon_champ==1 & decennais_ind==1980 & sexe_ind==2 & (regionnais_ind==11
| regionnais_ind==12) [aw=poids_ind], order(h) select(3) /*les femmes qui sont nées en France
entière de 1980 à 1989 ont comme 3 prénoms les plus fréquents Elodie, Emilie, Aurélie*/
    groups prenom_ind if bon_champ==1 & decennais_ind==1990 & sexe_ind==1 & (regionnais_ind==11
| regionnais_ind==12) [aw=poids_ind], order(h) select(3) /*les hommes qui sont nés en France
entière de 1990 à 1999 ont comme 3 prénoms les plus fréquents Nicolas, Julien, Thomas*/
    groups prenom_ind if bon_champ==1 & decennais_ind==1990 & sexe_ind==2 & (regionnais_ind==11
| regionnais_ind==12) [aw=poids_ind], order(h) select(3) /*les femmes qui sont nées en France
entière de 1990 à 1999 ont comme 3 prénoms les plus fréquents Elodie, Manon, Laura*/
    groups prenom_ind if bon_champ==1 & decennais_ind==2000 & sexe_ind==1 & (regionnais_ind==11
| regionnais_ind==12) [aw=poids_ind], order(h) select(3) /*les hommes qui sont nés en France
entière de 2000 à 2009 ont comme 3 prénoms les plus fréquents Lucas, Enzo, Tom*/
    groups prenom_ind if bon_champ==1 & decennais_ind==2000 & sexe_ind==2 & (regionnais_ind==11
| regionnais_ind==12) [aw=poids_ind], order(h) select(3) /*les femmes qui sont nées en France
entière de 2000 à 2009 ont comme 3 prénoms les plus fréquents Manon, Léa, Clara*/
save repro, replace
/*Le Top 3 des prénoms par sexe et par décennie de naissance selon le fichier INSEE des

```

prénoms (1900-2019) est issu des analyses ci-dessous destinées à produire la Figure 5.7a\*/

\*Figure 5.6a. Nombre de prénoms différents portés en 2008, par sexe et par décennie de naissance, N = 32 022

```
use repro
codebook prenom_ind if bon_champ==1 & decennais_ind==1950 & sexe_ind==1 &
(regionnais_ind==11 | regionnais_ind==12) /*les hommes qui sont nés en France entière de
1950 à 1959 sont 463 et portent 147 prénoms différents*/
codebook prenom_ind if bon_champ==1 & decennais_ind==1950 & sexe_ind==2 &
(regionnais_ind==11 | regionnais_ind==12) /*les femmes qui sont nées en France entière de
1950 à 1959 sont 581 et portent 238 prénoms différents*/
codebook prenom_ind if bon_champ==1 & decennais_ind==1960 & sexe_ind==1 &
(regionnais_ind==11 | regionnais_ind==12) /*les hommes qui sont nés en France entière de
1960 à 1969 sont 1 273 et portent 351 prénoms différents*/
codebook prenom_ind if bon_champ==1 & decennais_ind==1960 & sexe_ind==2 &
(regionnais_ind==11 | regionnais_ind==12) /*les femmes qui sont nées en France entière de
1960 à 1969 sont 1 510 et portent 461 prénoms différents*/
codebook prenom_ind if bon_champ==1 & decennais_ind==1970 & sexe_ind==1 &
(regionnais_ind==11 | regionnais_ind==12) /*les hommes qui sont nés en France entière de
1970 à 1979 sont 1 758 et portent 571 prénoms différents*/
codebook prenom_ind if bon_champ==1 & decennais_ind==1970 & sexe_ind==2 &
(regionnais_ind==11 | regionnais_ind==12) /*les femmes qui sont nées en France entière de
1970 à 1979 sont 1 996 et portent 695 prénoms différents*/
codebook prenom_ind if bon_champ==1 & decennais_ind==1980 & sexe_ind==1 &
(regionnais_ind==11 | regionnais_ind==12) /*les hommes qui sont nés en France entière de
1980 à 1989 sont 2 745 et portent 982 prénoms différents*/
codebook prenom_ind if bon_champ==1 & decennais_ind==1980 & sexe_ind==2 &
(regionnais_ind==11 | regionnais_ind==12) /*les femmes qui sont nées en France entière de
1980 à 1989 sont 2 786 et portent 1 059 prénoms différents*/
codebook prenom_ind if bon_champ==1 & decennais_ind==1990 & sexe_ind==1 &
(regionnais_ind==11 | regionnais_ind==12) /*les hommes qui sont nés en France entière de
1990 à 1999 sont 4 422 et portent 1 540 prénoms différents*/
codebook prenom_ind if bon_champ==1 & decennais_ind==1990 & sexe_ind==2 &
(regionnais_ind==11 | regionnais_ind==12) /*les femmes qui sont nées en France entière de
1990 à 1999 sont 4 245 et portent 1 553 prénoms différents*/
codebook prenom_ind if bon_champ==1 & decennais_ind==2000 & sexe_ind==1 &
(regionnais_ind==11 | regionnais_ind==12) /*les hommes qui sont nés en France entière de
2000 à 2009 sont 5 204 et portent 1 970 prénoms différents*/
codebook prenom_ind if bon_champ==1 & decennais_ind==2000 & sexe_ind==2 &
(regionnais_ind==11 | regionnais_ind==12) /*les femmes qui sont nées en France entière de
2000 à 2009 sont 5 039 et portent 2 135 prénoms différents*/
/*codebook prenom_ind if bon_champ==1 & decennais_ind==. & sexe_ind==1 &
(regionnais_ind==11 | regionnais_ind==12)*/ /*les hommes qui sont nés en France entière en .
sont 9 et portent 9 prénoms différents*/
/*codebook prenom_ind if bon_champ==1 & decennais_ind==. & sexe_ind==2 &
(regionnais_ind==11 | regionnais_ind==12)*/ /*les femmes qui sont nées en France entière en .
sont 10 et portent 10 prénoms différents*/
save repro, replace
```

\*Figure 5.6b. Nombre de prénoms différents attribués depuis 1900, par sexe et par année de naissance, N = 85 047 407

```
use insee
tabstat nombre if (sexe_ind==1), by(anais_ind) statistics(n sum) missing /*indicateurs pour
les nouveaux-nés garçons, contraint de comptabiliser tous les prénoms rares comme s'ils
constituaient ensemble 1 seul prénom*/
tabstat nombre if (sexe_ind==2), by(anais_ind) statistics(n sum) missing /*indicateurs pour
les nouveaux-nés filles, contraint de comptabiliser tous les prénoms rares comme s'ils
constituaient ensemble 1 seul prénom*/
save insee, replace
/*Au total, le Fichier INSEE des prénoms 2019 comprend:
- les prénoms attribués à 85 868 884 nouveau-nés de 1900 à 2019, y compris les prénoms
"rares" et y compris les prénoms dont l'année d'attribution n'est pas renseignée;
- les prénoms attribués à 85 047 407 nouveau-nés de 1900 à 2019, y compris les prénoms
"rares" mais pas les prénoms dont l'année d'attribution n'est pas renseignée;
- les prénoms attribués à 84 264 960 nouveau-nés de 1900 à 2019, à l'exclusion des prénoms
"rares" et des prénoms dont l'année d'attribution n'est pas renseignée*/
```

\*Figure 5.7a. Part des habitants en 2008 qui portent l'un des 10 prénoms les plus attribués de leur décennie de naissance, par sexe et par décennie de naissance, N = 32 022

```
use repro
groups prenom_ind if bon_champ==1 & decennais_ind==1950 & sexe_ind==1 & (regionnais_ind==11
| regionnais_ind==12) [aw=poids_ind], order(h) select(10) /*Top 10 des prénoms des hommes nés
en France entière dans les années 1950, et leur fréquence*/
groups prenom_ind if bon_champ==1 & decennais_ind==1950 & sexe_ind==2 & (regionnais_ind==11
| regionnais_ind==12) [aw=poids_ind], order(h) select(10) /*Top 10 des prénoms des femmes nées
en France entière dans les années 1950, et leur fréquence*/
groups prenom_ind if bon_champ==1 & decennais_ind==1960 & sexe_ind==1 & (regionnais_ind==11
```



```

/*Top 10 des prénoms des hommes nés en France entière dans les années 1970, et leur
fréquence*/
groups prenom_ind if (decennais_ind==1970 & sexe_ind==2) [aw=nombre], order(h) select(10)
/*Top 10 des prénoms des femmes nées en France entière dans les années 1970, et leur
fréquence*/
groups prenom_ind if (decennais_ind==1980 & sexe_ind==1) [aw=nombre], order(h) select(11)
/*Top 10 des prénoms des hommes nés en France entière dans les années 1980, et leur
fréquence; j'examine les 11 prénoms les plus attribués, sans quoi la présence de la catégorie
"prénoms rares" empêche d'observer le 10e prénom le plus attribué et sa fréquence*/
groups prenom_ind if (decennais_ind==1980 & sexe_ind==2) [aw=nombre], order(h) select(11)
/*Top 10 des prénoms des femmes nées en France entière dans les années 1980, et leur
fréquence; j'examine les 11 prénoms les plus attribués, sans quoi la présence de la catégorie
"prénoms rares" empêche d'observer le 10e prénom le plus attribué et sa fréquence*/
groups prenom_ind if (decennais_ind==1990 & sexe_ind==1) [aw=nombre], order(h) select(11)
/*Top 10 des prénoms des hommes nés en France entière dans les années 1990, et leur
fréquence; j'examine les 11 prénoms les plus attribués, sans quoi la présence de la catégorie
"prénoms rares" empêche d'observer le 10e prénom le plus attribué et sa fréquence*/
groups prenom_ind if (decennais_ind==1990 & sexe_ind==2) [aw=nombre], order(h) select(11)
/*Top 10 des prénoms des hommes nés en France entière dans les années 2000, et leur
fréquence; j'examine les 11 prénoms les plus attribués, sans quoi la présence de la catégorie
"prénoms rares" empêche d'observer le 10e prénom le plus attribué et sa fréquence*/
groups prenom_ind if (decennais_ind==2000 & sexe_ind==1) [aw=nombre], order(h) select(11)
/*Top 10 des prénoms des femmes nées en France entière dans les années 2000, et leur
fréquence; j'examine les 11 prénoms les plus attribués, sans quoi la présence de la catégorie
"prénoms rares" empêche d'observer le 10e prénom le plus attribué et sa fréquence*/
groups prenom_ind if (decennais_ind==2010 & sexe_ind==1) [aw=nombre], order(h) select(11)
/*Top 10 des prénoms des hommes nés en France entière dans les années 2010, et leur
fréquence; j'examine les 11 prénoms les plus attribués, sans quoi la présence de la catégorie
"prénoms rares" empêche d'observer le 10e prénom le plus attribué et sa fréquence*/
groups prenom_ind if (decennais_ind==2010 & sexe_ind==2) [aw=nombre], order(h) select(11)
/*Top 10 des prénoms des femmes nées en France entière dans les années 2010, et leur
fréquence; j'examine les 11 prénoms les plus attribués, sans quoi la présence de la catégorie
"prénoms rares" empêche d'observer le 10e prénom le plus attribué et sa fréquence*/
save insee, replace

```

\*Figure 5.7c. Part des nouveau-nés en France qui se voient attribuer l'un des 10 prénoms les plus attribués de la décennie, par sexe et par décennie de naissance  
use insee

```

generate prenom_ind_dupaquier = prenom_ind
replace prenom_ind_dupaquier = "JEAN" if strmatch(prenom_ind,"JEAN *") & (sexe_ind==1)
replace prenom_ind_dupaquier = "MARIE" if strmatch(prenom_ind,"MARIE *") & (sexe_ind==2)
groups prenom_ind_dupaquier if (decennais_ind==1900 & sexe_ind==1) [aw=nombre], order(h)
select(10) /*idem que supra*/
groups prenom_ind_dupaquier if (decennais_ind==1900 & sexe_ind==2) [aw=nombre], order(h)
select(10) /*idem que supra*/
groups prenom_ind_dupaquier if (decennais_ind==1910 & sexe_ind==1) [aw=nombre], order(h)
select(10) /*idem que supra*/
groups prenom_ind_dupaquier if (decennais_ind==1910 & sexe_ind==2) [aw=nombre], order(h)
select(10) /*idem que supra*/
groups prenom_ind_dupaquier if (decennais_ind==1920 & sexe_ind==1) [aw=nombre], order(h)
select(10) /*idem que supra*/
groups prenom_ind_dupaquier if (decennais_ind==1920 & sexe_ind==2) [aw=nombre], order(h)
select(10) /*idem que supra*/
groups prenom_ind_dupaquier if (decennais_ind==1930 & sexe_ind==1) [aw=nombre], order(h)
select(10) /*idem que supra*/
groups prenom_ind_dupaquier if (decennais_ind==1930 & sexe_ind==2) [aw=nombre], order(h)
select(10) /*idem que supra*/
groups prenom_ind_dupaquier if (decennais_ind==1940 & sexe_ind==1) [aw=nombre], order(h)
select(10) /*idem que supra*/
groups prenom_ind_dupaquier if (decennais_ind==1940 & sexe_ind==2) [aw=nombre], order(h)
select(10) /*idem que supra*/
groups prenom_ind_dupaquier if (decennais_ind==1950 & sexe_ind==1) [aw=nombre], order(h)
select(10) /*idem que supra*/
groups prenom_ind_dupaquier if (decennais_ind==1950 & sexe_ind==2) [aw=nombre], order(h)
select(10) /*idem que supra*/
groups prenom_ind_dupaquier if (decennais_ind==1960 & sexe_ind==1) [aw=nombre], order(h)
select(10) /*idem que supra*/
groups prenom_ind_dupaquier if (decennais_ind==1960 & sexe_ind==2) [aw=nombre], order(h)
select(10) /*idem que supra*/
groups prenom_ind_dupaquier if (decennais_ind==1970 & sexe_ind==1) [aw=nombre], order(h)
select(10) /*idem que supra*/
groups prenom_ind_dupaquier if (decennais_ind==1970 & sexe_ind==2) [aw=nombre], order(h)
select(10) /*idem que supra*/

```

```

groups prenom_ind_dupaquier if (decennais_ind==1980 & sexe_ind==1) [aw=nombre], order(h)
select(11) /*idem que supra*/
groups prenom_ind_dupaquier if (decennais_ind==1980 & sexe_ind==2) [aw=nombre], order(h)
select(11) /*idem que supra*/
groups prenom_ind_dupaquier if (decennais_ind==1990 & sexe_ind==1) [aw=nombre], order(h)
select(11) /*idem que supra*/
groups prenom_ind_dupaquier if (decennais_ind==1990 & sexe_ind==2) [aw=nombre], order(h)
select(11) /*idem que supra*/
groups prenom_ind_dupaquier if (decennais_ind==2000 & sexe_ind==1) [aw=nombre], order(h)
select(11) /*idem que supra*/
groups prenom_ind_dupaquier if (decennais_ind==2000 & sexe_ind==2) [aw=nombre], order(h)
select(11) /*idem que supra*/
groups prenom_ind_dupaquier if (decennais_ind==2010 & sexe_ind==1) [aw=nombre], order(h)
select(11) /*idem que supra*/
groups prenom_ind_dupaquier if (decennais_ind==2010 & sexe_ind==2) [aw=nombre], order(h)
select(11) /*idem que supra*/
save insee, replace

```

\*Figure 5.8a. Nombre de prénoms du Top 10 de la décennie de naissance qui figuraient dans le Top 10 de la décennie précédente, par sexe et par décennie de naissance (TeO)  
/\*faite à la main sous Excel\*/

\*Figure 5.8b. Nombre de prénoms du Top 10 de la décennie de naissance qui figuraient dans le Top 10 de la décennie précédente, par sexe et par décennie de naissance (fichier INSEE)  
/\*faite à la main sous Excel\*/

\*Figure 5.8c. Nombre de prénoms du Top 10 de la décennie en France qui figuraient dans le Top 10 de la décennie précédente, par sexe et par décennie de naissance (fichier INSEE + Dupâquier et al. 1987)  
/\*faite à la main sous Excel\*/

\*Figure 5.8d-f. Nombre de prénoms du Top 10 de la décennie qui figuraient dans le Top 10 de la décennie précédente, par sexe et par décennie de naissance, en Angleterre-Galles, Etats-Unis et Brésil  
/\*faites à la main sous Excel\*/

\*Figure 5.9a. Part de prénoms portés en 2008 qui sont des prénoms composés, par sexe et par décennie de naissance, N = 32 022

```

use repro
generate prenom_compose = 0
replace prenom_compose = 1 if strmatch(prenom_ind,"?* *?")
    tab decennais_ind prenom_compose if bon_champ==1 & sexe_ind==1 & (regionnais_ind==11 |
regionnais_ind==12) [aw=poids_ind], row /*nofreq*/ missing
    tab decennais_ind prenom_compose if bon_champ==1 & sexe_ind==2 & (regionnais_ind==11 |
regionnais_ind==12) [aw=poids_ind], row /*nofreq*/ missing
save repro, replace

```

\*Figure 5.9b. Part de prénoms attribués depuis 1900 qui sont des prénoms composés, par sexe et par décennie de naissance, N = 84 264 960

```

use insee
generate prenom_compose = 0 if (prenom_ind!="_PRENOMS_RARES")
replace prenom_compose = 1 if strmatch(prenom_ind,"?* *?")
    tab decennais_ind prenom_compose if (sexe_ind==1) [aw=nombre], row nofreq
    tab decennais_ind prenom_compose if (sexe_ind==2) [aw=nombre], row nofreq
save insee, replace

```

\*Figure 5.9c. Nombre moyen de lettres des prénoms portés en 2008, par sexe et par décennie de naissance, N = 32 022

```

use repro
generate taille_prenom = strlen(prenom_ind)
    tabstat taille_prenom if bon_champ==1 & sexe_ind==1 & (regionnais_ind==11 |
regionnais_ind==12) [aw=poids_ind], by(decennais_ind) statistics(min max mean)
    tabstat taille_prenom if bon_champ==1 & sexe_ind==2 & (regionnais_ind==11 |
regionnais_ind==12) [aw=poids_ind], by(decennais_ind) statistics(min max mean)
save repro, replace

```

\*Figure 5.9d. Nombre moyen de lettres des prénoms attribués depuis 1900, par sexe et par décennie de naissance, N = 84 264 960

```

use insee
generate taille_prenom = strlen(prenom_ind) if (prenom_ind!="_PRENOMS_RARES")
    tabstat taille_prenom if (sexe_ind==1) [aw=nombre], by(decennais_ind) statistics(n min max mean)
    tabstat taille_prenom if (sexe_ind==2) [aw=nombre], by(decennais_ind) statistics(n min max mean)
save insee, replace

```

\*Figure 5.10a. Part de prénoms portés en 2008 qui débutent par une voyelle, par sexe et par





```

ôôôôô
ùùùùù
ý
*/

```

```

*Je nettoie de façon harmonisée la NOMENCLATURE1
use nomenclature1
  replace prenom_ind = subinstr(prenom_ind,"'", " ",.) /*on remplace les apostrophes par des
espaces, comme dans les prénoms TeO*/
  replace prenom_ind = subinstr(prenom_ind,"`", " ",.)
  replace prenom_ind = subinstr(prenom_ind,"'", " ",.)
  replace prenom_ind = subinstr(prenom_ind,"-", " ",.)
  replace prenom_ind = subinstr(prenom_ind,"_", " ",.)
  replace prenom_ind = subinstr(prenom_ind,".", " ",.)
  replace prenom_ind = subinstr(prenom_ind,"'", " ",.)
  replace prenom_ind = subinstr(prenom_ind,";", " ",.)
  replace prenom_ind = subinstr(prenom_ind,":", " ",.)
  replace prenom_ind = subinstr(prenom_ind,"?", " ",.)
  replace prenom_ind = subinstr(prenom_ind,"!", " ",.)
  replace prenom_ind = subinstr(prenom_ind,"/", " ",.)
  replace prenom_ind = subinstr(prenom_ind,"\\", " ",.)
  replace prenom_ind = subinstr(prenom_ind,"\\", " ",.)
  replace prenom_ind = subinstr(prenom_ind,"|", " ",.)
  replace prenom_ind = subinstr(prenom_ind,"'", " ",.)
  replace prenom_ind = subinstr(prenom_ind,"^", " ",.)
  replace prenom_ind = subinstr(prenom_ind,"À", "A",.)
  replace prenom_ind = subinstr(prenom_ind,"Á", "A",.)
  replace prenom_ind = subinstr(prenom_ind,"Â", "A",.)
  replace prenom_ind = subinstr(prenom_ind,"Ã", "A",.)
  replace prenom_ind = subinstr(prenom_ind,"Ä", "A",.)
  replace prenom_ind = subinstr(prenom_ind,"Å", "A",.)
  replace prenom_ind = subinstr(prenom_ind,"Ç", "C",.)
  replace prenom_ind = subinstr(prenom_ind,"Ć", "C",.)
  replace prenom_ind = subinstr(prenom_ind,"Ĉ", "C",.)
  replace prenom_ind = subinstr(prenom_ind,"Ė", "E",.)
  replace prenom_ind = subinstr(prenom_ind,"Ē", "E",.)
  replace prenom_ind = subinstr(prenom_ind,"Ê", "E",.)
  replace prenom_ind = subinstr(prenom_ind,"Ë", "E",.)
  replace prenom_ind = subinstr(prenom_ind,"İ", "I",.)
  replace prenom_ind = subinstr(prenom_ind,"Í", "I",.)
  replace prenom_ind = subinstr(prenom_ind,"Î", "I",.)
  replace prenom_ind = subinstr(prenom_ind,"Ī", "I",.)
  replace prenom_ind = subinstr(prenom_ind,"Ñ", "N",.)
  replace prenom_ind = subinstr(prenom_ind,"Ò", "O",.)
  replace prenom_ind = subinstr(prenom_ind,"Ó", "O",.)
  replace prenom_ind = subinstr(prenom_ind,"Ô", "O",.)
  replace prenom_ind = subinstr(prenom_ind,"Õ", "O",.)
  replace prenom_ind = subinstr(prenom_ind,"Ö", "O",.)
  replace prenom_ind = subinstr(prenom_ind,"Û", "U",.)
  replace prenom_ind = subinstr(prenom_ind,"Ü", "U",.)
  replace prenom_ind = subinstr(prenom_ind,"Ů", "U",.)
  replace prenom_ind = subinstr(prenom_ind,"Ů", "U",.)
  replace prenom_ind = subinstr(prenom_ind,"Ů", "U",.)
  replace prenom_ind = subinstr(prenom_ind,"Ý", "Y",.)
  replace prenom_ind = stritrim(prenom_ind) /*on remplace les espaces consécutifs par
1 seul espace*/
  replace prenom_ind = strtrim(prenom_ind) /*on supprime les espaces initiaux et
finaux*/
sort prenom_ind

duplicates report prenom_ind nomenclature1 male_nomenclature1 /*il n'y a pas de doublons dans
les 4 045 prénoms*/

rename male_nomenclature1 sexe_ind
destring sexe_ind, replace
recode sexe_ind (1=1) (0=2)

tab nomenclature1 sexe_ind, missing /*la nomenclature 1 contient 3 396 prénoms "arabes", 577
prénoms "africains" et 72 prénoms "arabes et africains", soit au total 4 045 prénoms
différents*/

*Je recode la nomenclature 1 en modalités numériques
  replace nomenclature1="05" if (nomenclature1=="arabe")
  replace nomenclature1="06" if (nomenclature1=="africain")
  replace nomenclature1="05" if (nomenclature1=="arabe et africain") /*ces 72 prénoms Fatima,
Khadija, Amina, etc., catégorisés comme "arabes et africains", sont recatégorisés comme
"arabo-musulmans" (05)*/
  tab nomenclature1, missing

```

```

save nomenclature1, replace

*Je nettoie de façon harmonisée la NOMENCLATURE2
use nomenclature2
replace prenom_ind = subinstr(prenom_ind,"'", " ",.) /*on remplace les apostrophes par des
espaces, comme dans les prénoms TeO*/
replace prenom_ind = subinstr(prenom_ind,"`", " ",.)
replace prenom_ind = subinstr(prenom_ind,"'", " ",.)
replace prenom_ind = subinstr(prenom_ind,"-", " ",.)
replace prenom_ind = subinstr(prenom_ind," ", " ",.)
replace prenom_ind = subinstr(prenom_ind,".", " ",.)
replace prenom_ind = subinstr(prenom_ind,";", " ",.)
replace prenom_ind = subinstr(prenom_ind,":", " ",.)
replace prenom_ind = subinstr(prenom_ind,"?", " ",.)
replace prenom_ind = subinstr(prenom_ind,"!", " ",.)
replace prenom_ind = subinstr(prenom_ind,"/", " ",.)
replace prenom_ind = subinstr(prenom_ind,"\\", " ",.)
replace prenom_ind = subinstr(prenom_ind,"\\", " ",.)
replace prenom_ind = subinstr(prenom_ind,"|", " ",.)
replace prenom_ind = subinstr(prenom_ind,"'", " ",.)
replace prenom_ind = subinstr(prenom_ind,"^", " ",.)
replace prenom_ind = subinstr(prenom_ind,"Ä", "A",.)
replace prenom_ind = subinstr(prenom_ind,"Å", "A",.)
replace prenom_ind = subinstr(prenom_ind,"Ä", "A",.)
replace prenom_ind = subinstr(prenom_ind,"Å", "A",.)
replace prenom_ind = subinstr(prenom_ind,"Ä", "A",.)
replace prenom_ind = subinstr(prenom_ind,"Å", "A",.)
replace prenom_ind = subinstr(prenom_ind,"Ç", "C",.)
replace prenom_ind = subinstr(prenom_ind,"Ç", "C",.)
replace prenom_ind = subinstr(prenom_ind,"Ç", "C",.)
replace prenom_ind = subinstr(prenom_ind,"È", "E",.)
replace prenom_ind = subinstr(prenom_ind,"È", "E",.)
replace prenom_ind = subinstr(prenom_ind,"È", "E",.)
replace prenom_ind = subinstr(prenom_ind,"È", "E",.)
replace prenom_ind = subinstr(prenom_ind,"Ì", "I",.)
replace prenom_ind = subinstr(prenom_ind,"Ì", "I",.)
replace prenom_ind = subinstr(prenom_ind,"Ì", "I",.)
replace prenom_ind = subinstr(prenom_ind,"Ñ", "N",.)
replace prenom_ind = subinstr(prenom_ind,"Ò", "O",.)
replace prenom_ind = subinstr(prenom_ind,"Ó", "O",.)
replace prenom_ind = subinstr(prenom_ind,"Ô", "O",.)
replace prenom_ind = subinstr(prenom_ind,"Ô", "O",.)
replace prenom_ind = subinstr(prenom_ind,"Ô", "O",.)
replace prenom_ind = subinstr(prenom_ind,"Û", "U",.)
replace prenom_ind = subinstr(prenom_ind,"Û", "U",.)
replace prenom_ind = subinstr(prenom_ind,"Û", "U",.)
replace prenom_ind = subinstr(prenom_ind,"Ü", "U",.)
replace prenom_ind = subinstr(prenom_ind,"Ý", "Y",.)
replace prenom_ind = stritrim(prenom_ind) /*on remplace les espaces consécutifs par
1 seul espace*/
replace prenom_ind = strtrim(prenom_ind) /*on supprime les espaces initiaux et
finaux*/
sort prenom_ind

duplicates report prenom_ind /*il y a 1 384 + 114 = 1 498 doublons, dont 692 + 76 = 768
en surplus*/
duplicates drop prenom_ind, force /*ces 768 doublons en surplus sont éliminés*/
tab nomenclature2, missing /*la nomenclature 2 contient 403 prénoms "africains", 2
095 prénoms "asiatiques", 781 prénoms "britanniques", 987 prénoms "européens", 6 725 prénoms
"français", 472 prénoms "germaniques", 127 prénoms "juifs", 7 581 prénoms "musulmans" et 3 449
prénoms "méditerranéens", soit au total 22 620 prénoms différents*/

*Je recode la nomenclature 2 en modalités numériques
replace nomenclature2="01" if (nomenclature2=="français")
replace nomenclature2="02" if (nomenclature2=="méditerranéen")
replace nomenclature2="03" if (nomenclature2=="britannique")
replace nomenclature2="04" if (nomenclature2=="germanique")
replace nomenclature2="04" if (nomenclature2=="européen")
replace nomenclature2="05" if (nomenclature2=="musulman")
replace nomenclature2="06" if (nomenclature2=="africain")
replace nomenclature2="07" if (nomenclature2=="asiatique")
replace nomenclature2="08" if (nomenclature2=="juif")
tab nomenclature2, missing

```

```

save nomenclature2, replace

*Je nettoie de façon harmonisée la NOMENCLATURE3
use nomenclature3
  replace prenom_ind = subinstr(prenom_ind,"'", " ",.) /*on remplace les apostrophes par des
espaces, comme dans les prénoms TeO*/
  replace prenom_ind = subinstr(prenom_ind,"`", " ",.)
  replace prenom_ind = subinstr(prenom_ind,"'", " ",.)
  replace prenom_ind = subinstr(prenom_ind,"-", " ",.)
  replace prenom_ind = subinstr(prenom_ind,"_", " ",.)
  replace prenom_ind = subinstr(prenom_ind,".", " ",.)
  replace prenom_ind = subinstr(prenom_ind,",", " ",.)
  replace prenom_ind = subinstr(prenom_ind,":", " ",.)
  replace prenom_ind = subinstr(prenom_ind,"?", " ",.)
  replace prenom_ind = subinstr(prenom_ind,"!", " ",.)
  replace prenom_ind = subinstr(prenom_ind,"/", " ",.)
  replace prenom_ind = subinstr(prenom_ind,"\\", " ",.)
  replace prenom_ind = subinstr(prenom_ind,"\\", " ",.)
  replace prenom_ind = subinstr(prenom_ind,"|", " ",.)
  replace prenom_ind = subinstr(prenom_ind,"'", " ",.)
  replace prenom_ind = subinstr(prenom_ind,"^", " ",.)
  replace prenom_ind = subinstr(prenom_ind,"À", "A",.)
  replace prenom_ind = subinstr(prenom_ind,"Á", "A",.)
  replace prenom_ind = subinstr(prenom_ind,"Â", "A",.)
  replace prenom_ind = subinstr(prenom_ind,"Ã", "A",.)
  replace prenom_ind = subinstr(prenom_ind,"Ä", "A",.)
  replace prenom_ind = subinstr(prenom_ind,"Å", "A",.)
  replace prenom_ind = subinstr(prenom_ind,"Ç", "C",.)
  replace prenom_ind = subinstr(prenom_ind,"Ç", "C",.)
  replace prenom_ind = subinstr(prenom_ind,"Ĉ", "C",.)
  replace prenom_ind = subinstr(prenom_ind,"Ĉ", "C",.)
  replace prenom_ind = subinstr(prenom_ind,"È", "E",.)
  replace prenom_ind = subinstr(prenom_ind,"É", "E",.)
  replace prenom_ind = subinstr(prenom_ind,"Ê", "E",.)
  replace prenom_ind = subinstr(prenom_ind,"Ë", "E",.)
  replace prenom_ind = subinstr(prenom_ind,"Ī", "I",.)
  replace prenom_ind = subinstr(prenom_ind,"Ī", "I",.)
  replace prenom_ind = subinstr(prenom_ind,"Ī", "I",.)
  replace prenom_ind = subinstr(prenom_ind,"Ĭ", "I",.)
  replace prenom_ind = subinstr(prenom_ind,"Ñ", "N",.)
  replace prenom_ind = subinstr(prenom_ind,"Ō", "O",.)
  replace prenom_ind = subinstr(prenom_ind,"Ō", "O",.)
  replace prenom_ind = subinstr(prenom_ind,"Ō", "O",.)
  replace prenom_ind = subinstr(prenom_ind,"Ō", "O",.)
  replace prenom_ind = subinstr(prenom_ind,"Ū", "U",.)
  replace prenom_ind = subinstr(prenom_ind,"Ū", "U",.)
  replace prenom_ind = subinstr(prenom_ind,"Ū", "U",.)
  replace prenom_ind = subinstr(prenom_ind,"Ū", "U",.)
  replace prenom_ind = subinstr(prenom_ind,"Ÿ", "Y",.)
  replace prenom_ind = stritrim(prenom_ind) /*on remplace les espaces consécutifs par
1 seul espace*/
  replace prenom_ind = strtrim(prenom_ind) /*on supprime les espaces initiaux et
finaux*/
sort prenom_ind

duplicates report prenom_ind nomenclature3 male_nomenclature3 /*il y a 1 308 + 162 + 32
+ 10 = 1 512 doublons, dont 654 + 108 + 24 + 8 = 794 en surplus*/
duplicates drop prenom_ind nomenclature3 male_nomenclature3, force /*ces 794 doublons en
surplus sont éliminés*/

duplicates tag prenom_ind, generate(ametautre)
tab ametautre, missing /*il reste 44 prénoms
classifiés à la fois comme arabo-musulmans et comme autres, dont 22 en surplus*/
replace nomenclature3 = "arabo-musulman et autre" if (ametautre==1)
duplicates drop prenom_ind nomenclature3 male_nomenclature3, force /*ces 22 doublons en
surplus sont éliminés*/

rename male_nomenclature3 sexe_ind
destring sexe_ind, replace
tab nomenclature3 sexe_ind, missing /*la nomenclature 3
contient 4 375 prénoms "arabo-musulmans", 10 267 prénoms "autres" et 22 prénoms "à la fois
arabo-musulmans et autres", soit au total 14 664 prénoms différents*/

*Je recode la nomenclature 3 en modalités numériques
  replace nomenclature3="05" if (nomenclature3=="arabo-musulman")
  replace nomenclature3="09" if (nomenclature3=="autre")

```

```

    replace nomenclature3="05" if (nomenclature3=="arabo-musulman et autre") /*ces 22 prénoms
Yanis, Elias, Omer, etc., catégorisés comme "arabo-musulmans et autres", sont recatégorisés
comme "arabo-musulmans" (05)*/
    tab nomenclature3, missing

save nomenclature3, replace

*J'apparie les nomenclatures 1, 2 et 3 à REPRO
use repro
merge m:1 prenom_ind sexe_ind using nomenclature1.dta, keepusing(nomenclature1) keep(master
match)
    drop _merge
merge m:1 prenom_ind          using nomenclature2.dta, keepusing(nomenclature2) keep(master
match)
    drop _merge
merge m:1 prenom_ind sexe_ind using nomenclature3.dta, keepusing(nomenclature3) keep(master
match)
    drop _merge
    sort ident noi
order ident noi poids_ind sexe_ind anais_ind prenom_ind nomenclature1 nomenclature2
nomenclature3 lienego_ind regionnais_ind residence_ind group2_ind regionnaispere_ind
regionnaismere_ind orig_ind autrepereparent natpniv2 natmniv2 n_indper n_indmer lnp1_gr lnp2_gr
lnm1_gr lnm2_gr relig_ind r_impvie orig_relig_ind x_vufri
    tab nomenclature1 [aw=poids_ind], missing                                /*1a
nomenclature 1 s'apparie à 7 % des individus Te0*/
    tab nomenclature2 [aw=poids_ind], missing                                /*1a
nomenclature 2 s'apparie à 91 % des individus Te0*/
    tab nomenclature3 [aw=poids_ind], missing                                /*1a
nomenclature 3 s'apparie à 48 % des individus Te0*/
sort ident noi

*Je recode les nomenclatures 1 à 3
    tab nomenclature1 [aw=poids_ind], missing                                /**/
    replace nomenclature1="09" if (nomenclature1=="")                        /**/
    tab nomenclature1 [aw=poids_ind], missing                                /**/

    tab nomenclature2 [aw=poids_ind], missing                                /**/
    replace nomenclature2="09" if (nomenclature2=="")                        /**/
    tab nomenclature2 [aw=poids_ind], missing                                /**/

    tab nomenclature3 [aw=poids_ind], missing                                /**/
    replace nomenclature3="09" if (nomenclature3=="")                        /**/
    tab nomenclature3 [aw=poids_ind], missing                                /**/

*J'examine quels sont les prénoms non catégorisés les plus fréquents
generate nomenclatures = 0 if (nomenclature1=="09" & nomenclature2=="09" &
nomenclature3=="09")
    tab nomenclatures [aw=poids_ind], missing                                /*9 % des
individus du champ maximal ont un prénom d'origine non catégorisée*/
    groups prenom_ind [aw=poids_ind] if (nomenclatures==0), order(h) select(50) /*les prénoms
non catégorisés les plus fréquents comprennent "Jean-François", "Marie-Thérèse", "Jean-
Philippe" et autres prénoms composés.*/

*Je vais donc chercher à catégoriser les prénoms composés qui ne sont catégorisés par aucune
des 3 premières nomenclatures.
*Pour cela je commence par découper les prénoms composés en plusieurs prénoms composants.
split prenom_ind /*les prénoms composés des individus Te0 sont composés, au plus, de 4 prénoms
composants*/
order ident noi poids_ind sexe_ind anais_ind prenom_ind prenom_ind1 prenom_ind2 prenom_ind3
prenom_ind4 regionnais_ind nomenclature1 nomenclature2 nomenclature3 lienego_ind orig_ind
relig_ind r_impvie x_vufri
codebook prenom_ind1 /*          83 044 individus sur 83 045 ont un 1e prénom composant, i.e.
1 seul individu n'a pas de prénom*/
codebook prenom_ind2 /*seulement 5 281 individus sur 83 045 ont un 2e prénom composant*/
codebook prenom_ind3 /*seulement 203 individus sur 83 045 ont un 3e prénom composant*/
codebook prenom_ind4 /*seulement 2 individus sur 83 045 ont un 4e prénom composant*/
save repro, replace

*J'apparie les nomenclatures 1, 2 et 3 à chacun des 4 prénoms composants
*J'apparie la nomenclature 1 au 1er prénom composant
    use nomenclature1
rename prenom_ind prenom_ind1
rename nomenclature1 nomenclature1_ind1
    save nomenclature1, replace
    use repro
merge m:1 prenom_ind1 sexe_ind using nomenclature1.dta, keepusing(nomenclature1_ind1)
keep(master match)

```

```

drop _merge
save repro, replace
*J'apparie la nomenclature 2 au 1er prénom composant
    use nomenclature2
rename prenom_ind prenom_ind1
rename nomenclature2 nomenclature2_ind1
    save nomenclature2, replace
    use repro
merge m:1 prenom_ind1          using nomenclature2.dta, keepusing(nomenclature2_ind1)
keep(master match)
drop _merge
save repro, replace
*J'apparie la nomenclature 3 au 1er prénom composant
    use nomenclature3
rename prenom_ind prenom_ind1
rename nomenclature3 nomenclature3_ind1
    save nomenclature3, replace
    use repro
merge m:1 prenom_ind1 sexe_ind using nomenclature3.dta, keepusing(nomenclature3_ind1)
keep(master match)
drop _merge
save repro, replace
*J'apparie la nomenclature 1 au 2e prénom composant
    use nomenclature1
rename prenom_ind1 prenom_ind2
rename nomenclature1_ind1 nomenclature1_ind2
    save nomenclature1, replace
    use repro
merge m:1 prenom_ind2 sexe_ind using nomenclature1.dta, keepusing(nomenclature1_ind2)
keep(master match)
drop _merge
save repro, replace
*J'apparie la nomenclature 2 au 2e prénom composant
    use nomenclature2
rename prenom_ind1 prenom_ind2
rename nomenclature2_ind1 nomenclature2_ind2
    save nomenclature2, replace
    use repro
merge m:1 prenom_ind2          using nomenclature2.dta, keepusing(nomenclature2_ind2)
keep(master match)
drop _merge
save repro, replace
*J'apparie la nomenclature 3 au 2e prénom composant
    use nomenclature3
rename prenom_ind1 prenom_ind2
rename nomenclature3_ind1 nomenclature3_ind2
    save nomenclature3, replace
    use repro
merge m:1 prenom_ind2 sexe_ind using nomenclature3.dta, keepusing(nomenclature3_ind2)
keep(master match)
drop _merge
save repro, replace
*J'apparie la nomenclature 1 au 3e prénom composant
    use nomenclature1
rename prenom_ind2 prenom_ind3
rename nomenclature1_ind2 nomenclature1_ind3
    save nomenclature1, replace
    use repro
merge m:1 prenom_ind3 sexe_ind using nomenclature1.dta, keepusing(nomenclature1_ind3)
keep(master match)
drop _merge
save repro, replace
*J'apparie la nomenclature 2 au 3e prénom composant
    use nomenclature2
rename prenom_ind2 prenom_ind3
rename nomenclature2_ind2 nomenclature2_ind3
    save nomenclature2, replace
    use repro
merge m:1 prenom_ind3          using nomenclature2.dta, keepusing(nomenclature2_ind3)
keep(master match)
drop _merge
save repro, replace
*J'apparie la nomenclature 3 au 3e prénom composant
    use nomenclature3
rename prenom_ind2 prenom_ind3
rename nomenclature3_ind2 nomenclature3_ind3
    save nomenclature3, replace

```

```

    use repro
merge m:1 prenom_ind3 sexe_ind using nomenclature3.dta, keepusing(nomenclature3_ind3)
keep(master match)
    drop _merge
save repro, replace
*J'apparie la nomenclature 1 au 4e prénom composant
    use nomenclature1
rename prenom_ind3 prenom_ind4
rename nomenclature1_ind3 nomenclature1_ind4
    save nomenclature1, replace
    use repro
merge m:1 prenom_ind4 sexe_ind using nomenclature1.dta, keepusing(nomenclature1_ind4)
keep(master match)
    drop _merge
save repro, replace
*J'apparie la nomenclature 2 au 4e prénom composant
    use nomenclature2
rename prenom_ind3 prenom_ind4
rename nomenclature2_ind3 nomenclature2_ind4
    save nomenclature2, replace
    use repro
merge m:1 prenom_ind4 sexe_ind using nomenclature2.dta, keepusing(nomenclature2_ind4)
keep(master match)
    drop _merge
save repro, replace
*J'apparie la nomenclature 3 au 4e prénom composant
    use nomenclature3
rename prenom_ind3 prenom_ind4
rename nomenclature3_ind3 nomenclature3_ind4
    save nomenclature3, replace
    use repro
merge m:1 prenom_ind4 sexe_ind using nomenclature3.dta, keepusing(nomenclature3_ind4)
keep(master match)
    drop _merge
save repro, replace

*Je fais les tris à plat de ces variables
replace nomenclature1_ind1="09" if (nomenclature1_ind1=="")
    tab nomenclature1_ind1, missing /*ok*/
replace nomenclature2_ind1="09" if (nomenclature2_ind1=="")
    tab nomenclature2_ind1, missing /*ok*/
replace nomenclature3_ind1="09" if (nomenclature3_ind1=="")
    tab nomenclature3_ind1, missing /*ok*/
replace nomenclature1_ind2="09" if (nomenclature1_ind2=="")
    tab nomenclature1_ind2, missing /*ok*/
replace nomenclature2_ind2="09" if (nomenclature2_ind2=="")
    tab nomenclature2_ind2, missing /*ok*/
replace nomenclature3_ind2="09" if (nomenclature3_ind2=="")
    tab nomenclature3_ind2, missing /*ok*/
replace nomenclature1_ind3="09" if (nomenclature1_ind3=="")
    tab nomenclature1_ind3, missing /*ok*/
replace nomenclature2_ind3="09" if (nomenclature2_ind3=="")
/*tab nomenclature2_ind3, missing*/ /*ok*/
replace nomenclature3_ind3="09" if (nomenclature3_ind3=="")
    tab nomenclature3_ind3, missing /*ok*/
replace nomenclature1_ind4="09" if (nomenclature1_ind4=="")
    tab nomenclature1_ind4, missing /*ok*/
replace nomenclature2_ind4="09" if (nomenclature2_ind4=="")
/*tab nomenclature2_ind4, missing*/ /*ok*/
replace nomenclature3_ind4="09" if (nomenclature3_ind4=="")
    tab nomenclature3_ind4, missing /*ok*/

*Enfin je crée orig_prenom_ind dans REPRO, en 7 étapes.
generate orig_prenom_simple = nomenclature1_ind1 + nomenclature2_ind1 + nomenclature3_ind1 if
((prenom_ind=="") | (prenom_ind1!="" & prenom_ind2==""))
/*tab orig_prenom_simple, missing*/
generate orig_prenom_compose_2 = nomenclature1_ind1 + nomenclature2_ind1 + nomenclature3_ind1
+ nomenclature1_ind2 + nomenclature2_ind2 + nomenclature3_ind2 if (prenom_ind1!="" &
prenom_ind2!="" & prenom_ind3=="")
/*tab orig_prenom_compose_2, missing*/
generate orig_prenom_compose_3 = nomenclature1_ind1 + nomenclature2_ind1 + nomenclature3_ind1
+ nomenclature1_ind2 + nomenclature2_ind2 + nomenclature3_ind2 + nomenclature1_ind3 +
nomenclature2_ind3 + nomenclature3_ind3 if (prenom_ind1!="" & prenom_ind2!="" &
prenom_ind3!="")
/*tab orig_prenom_compose_3, missing*/
generate orig_prenom_compose_4 = nomenclature1_ind1 + nomenclature2_ind1 + nomenclature3_ind1
+ nomenclature1_ind2 + nomenclature2_ind2 + nomenclature3_ind2 + nomenclature1_ind3 +

```

```

nomenclature2_ind3 + nomenclature3_ind3 + nomenclature1_ind4 + nomenclature2_ind4 +
nomenclature3_ind4 if (prenom_ind1!="" & prenom_ind2!="" & prenom_ind3!="" & prenom_ind4!="")
/*tab orig_prenom_compose_4, missing*/

* (1) Si un prénom simple est non renseigné ou ne voit son origine catégorisée par aucune des
3 nomenclatures, ou si un prénom composé constitué de 2 à 4 "composants" (termes séparés par
un trait d'union, un espace, une apostrophe, etc.) ne voit son origine catégorisée par aucune
des 3 nomenclatures et qu'aucun de ses composants ne voit son origine catégorisée par l'une
des nomenclatures, alors on le catégorise comme "inclassable".
generate orig_prenom_ind = "09" if (orig_prenom_simple=="090909")
replace orig_prenom_ind = "09" if (orig_prenom_compose_2=="090909090909")
replace orig_prenom_ind = "09" if (orig_prenom_compose_3=="0909090909090909")
/*tab orig_prenom_ind [aw=poids_ind], missing*/
/* 5 045 individus ( 6 % des individus) portent un prénom simple dont l'origine
n'est catégorisée par aucune des 3 nomenclatures*/

* (2) Si un prénom simple ne voit son origine catégorisée que par une seule des 3
nomenclatures, alors on le catégorise de ladite origine.
**prénom simple catégorisé par 1 seule nomenclature : la nomenclature 1
replace orig_prenom_ind = substr(orig_prenom_simple,1,2) if
(strmatch(orig_prenom_simple,"??0909") & orig_prenom_ind=="")
**prénom simple catégorisé par 1 seule nomenclature : la nomenclature 2
replace orig_prenom_ind = substr(orig_prenom_simple,3,2) if
(strmatch(orig_prenom_simple,"09??09") & orig_prenom_ind=="")
**prénom simple catégorisé par 1 seule nomenclature : la nomenclature 3
replace orig_prenom_ind = substr(orig_prenom_simple,5,2) if
(strmatch(orig_prenom_simple,"0909??") & orig_prenom_ind=="")
tab orig_prenom_ind [aw=poids_ind], missing
/*66 653 individus (80 % des individus) portent un prénom simple dont
l'origine n'est catégorisée que par la nomenclature 1, ou la 2, ou la 3*/

* (3) Si un prénom simple voit son origine catégorisée par au moins 2 des 3 nomenclatures et
que ces origines sont les mêmes, alors on le catégorise de ladite origine.
**prénom simple catégorisé de la même façon par 2 nomenclatures : les nomenclatures 1 et 2
replace orig_prenom_ind = "05" if (orig_prenom_simple=="050509")
replace orig_prenom_ind = "06" if (orig_prenom_simple=="060609")
**prénom simple catégorisé de la même façon par 2 nomenclatures : les nomenclatures 1 et 3
replace orig_prenom_ind = "05" if (orig_prenom_simple=="050905")
**prénom simple catégorisé de la même façon par 2 nomenclatures : les nomenclatures 2 et 3
replace orig_prenom_ind = "05" if (orig_prenom_simple=="090505")
**prénom simple catégorisé de la même façon par 3 nomenclatures : les nomenclatures 1, 2 et 3
replace orig_prenom_ind = "05" if (orig_prenom_simple=="050505")
/*tab orig_prenom_ind [aw=poids_ind], missing*/
/* 5 641 individus ( 7 % des individus) portent un prénom simple dont l'origine
est catégorisée de façon identique par les nomenclatures 1, 2 et/ou 3*/

* (4) Si un prénom simple voit son origine catégorisée par au moins 2 des 3 nomenclatures mais
que ces origines ne sont pas les mêmes, alors on le catégorise d'une origine mixte ou ambiguë,
constituée des 2 ou 3 origines différentes attribuées par les nomenclatures.
**prénom simple catégorisé de façons différentes par 2 nomenclatures : les nomenclatures 1 et
2
replace orig_prenom_ind = "15" if (orig_prenom_simple=="050109")
replace orig_prenom_ind = "25" if (orig_prenom_simple=="050209")
replace orig_prenom_ind = "35" if (orig_prenom_simple=="050309")
replace orig_prenom_ind = "45" if (orig_prenom_simple=="050409")
replace orig_prenom_ind = "56" if (orig_prenom_simple=="050609")
replace orig_prenom_ind = "57" if (orig_prenom_simple=="050709")
replace orig_prenom_ind = "58" if (orig_prenom_simple=="050809")
replace orig_prenom_ind = "16" if (orig_prenom_simple=="060109")
replace orig_prenom_ind = "26" if (orig_prenom_simple=="060209")
replace orig_prenom_ind = "36" if (orig_prenom_simple=="060309")
replace orig_prenom_ind = "46" if (orig_prenom_simple=="060409")
replace orig_prenom_ind = "56" if (orig_prenom_simple=="060509")
replace orig_prenom_ind = "67" if (orig_prenom_simple=="060709")
replace orig_prenom_ind = "68" if (orig_prenom_simple=="060809")
**prénom simple catégorisé de façons différentes par 2 nomenclatures : les nomenclatures 1 et
3
replace orig_prenom_ind = "56" if (orig_prenom_simple=="060905")
**prénom simple catégorisé de façons différentes par 2 nomenclatures : les nomenclatures 2 et
3
replace orig_prenom_ind = "15" if (orig_prenom_simple=="090105")
replace orig_prenom_ind = "25" if (orig_prenom_simple=="090205")
replace orig_prenom_ind = "35" if (orig_prenom_simple=="090305")
replace orig_prenom_ind = "45" if (orig_prenom_simple=="090405")
replace orig_prenom_ind = "56" if (orig_prenom_simple=="090605")
replace orig_prenom_ind = "57" if (orig_prenom_simple=="090705")
replace orig_prenom_ind = "58" if (orig_prenom_simple=="090805")

```

```

**prénom simple catégorisé de façons différentes par 3 nomenclatures : les nomenclatures 1, 2
et 3
replace orig_prenom_ind = "15" if (orig_prenom_simple=="050105")
replace orig_prenom_ind = "25" if (orig_prenom_simple=="050205")
replace orig_prenom_ind = "35" if (orig_prenom_simple=="050305")
replace orig_prenom_ind = "45" if (orig_prenom_simple=="050405")
replace orig_prenom_ind = "56" if (orig_prenom_simple=="050605")
replace orig_prenom_ind = "57" if (orig_prenom_simple=="050705")
replace orig_prenom_ind = "58" if (orig_prenom_simple=="050805")
replace orig_prenom_ind = "156" if (orig_prenom_simple=="060105")
replace orig_prenom_ind = "256" if (orig_prenom_simple=="060205")
replace orig_prenom_ind = "356" if (orig_prenom_simple=="060305")
replace orig_prenom_ind = "456" if (orig_prenom_simple=="060405")
replace orig_prenom_ind = "56" if (orig_prenom_simple=="060505")
replace orig_prenom_ind = "56" if (orig_prenom_simple=="060605")
replace orig_prenom_ind = "567" if (orig_prenom_simple=="060705")
replace orig_prenom_ind = "568" if (orig_prenom_simple=="060805")
/*tab orig_prenom_ind [aw=poids_ind], missing*/
/* 431 individus ( 1 % des individus) portent un prénom simple dont l'origine
est catégorisée de façons différentes par les nomenclatures 1, 2 et/ou 3*/

* (5) Si un prénom composé constitué de 2 "composants" ne voit son origine catégorisée par
aucune des 3 nomenclatures, mais qu'au moins l'un de ses composants voit son origine
catégorisée par au moins l'une des nomenclatures, alors on le catégorise selon l'origine du
premier des composants dont l'origine est catégorisée.
**prénom composé catégorisé par 1 seule nomenclature
replace orig_prenom_ind = substr(orig_prenom_compose_2,1,2) if
(strmatch(orig_prenom_compose_2,"050909?????") & orig_prenom_ind=="")
replace orig_prenom_ind = substr(orig_prenom_compose_2,1,2) if
(strmatch(orig_prenom_compose_2,"060909?????") & orig_prenom_ind=="")
replace orig_prenom_ind = substr(orig_prenom_compose_2,7,2) if
(strmatch(orig_prenom_compose_2,"090909??0909") & orig_prenom_ind=="")
replace orig_prenom_ind = substr(orig_prenom_compose_2,3,2) if
(strmatch(orig_prenom_compose_2,"09??09?????") & orig_prenom_ind=="")
replace orig_prenom_ind = substr(orig_prenom_compose_2,9,2) if
(strmatch(orig_prenom_compose_2,"09090909??09") & orig_prenom_ind=="")
replace orig_prenom_ind = substr(orig_prenom_compose_2,5,2) if
(strmatch(orig_prenom_compose_2,"0909???????") & orig_prenom_ind=="")
replace orig_prenom_ind = substr(orig_prenom_compose_2,11,2) if
(strmatch(orig_prenom_compose_2,"0909090909??") & orig_prenom_ind=="")
**prénom composé catégorisé de la même façon par 2 nomenclatures
replace orig_prenom_ind = substr(orig_prenom_compose_2,1,2) if
(strmatch(orig_prenom_compose_2,"050509?????") & orig_prenom_ind=="")
replace orig_prenom_ind = substr(orig_prenom_compose_2,1,2) if
(strmatch(orig_prenom_compose_2,"060609?????") & orig_prenom_ind=="")
replace orig_prenom_ind = substr(orig_prenom_compose_2,7,2) if
(strmatch(orig_prenom_compose_2,"090909050509") & orig_prenom_ind=="")
replace orig_prenom_ind = substr(orig_prenom_compose_2,7,2) if
(strmatch(orig_prenom_compose_2,"090909060609") & orig_prenom_ind=="")
replace orig_prenom_ind = substr(orig_prenom_compose_2,7,2) if
(strmatch(orig_prenom_compose_3,"0909090505090909") & orig_prenom_ind=="")
replace orig_prenom_ind = substr(orig_prenom_compose_2,7,2) if
(strmatch(orig_prenom_compose_3,"0909090606090909") & orig_prenom_ind=="")
replace orig_prenom_ind = substr(orig_prenom_compose_2,1,2) if
(strmatch(orig_prenom_compose_2,"050905?????") & orig_prenom_ind=="")
replace orig_prenom_ind = substr(orig_prenom_compose_2,7,2) if
(strmatch(orig_prenom_compose_2,"090909050905") & orig_prenom_ind=="")
replace orig_prenom_ind = substr(orig_prenom_compose_2,3,2) if
(strmatch(orig_prenom_compose_2,"090505?????") & orig_prenom_ind=="")
replace orig_prenom_ind = substr(orig_prenom_compose_2,9,2) if
(strmatch(orig_prenom_compose_2,"090909090505") & orig_prenom_ind=="")
**prénom composé catégorisé de la même façon par 3 nomenclatures : les nomenclatures 1, 2 et 3
replace orig_prenom_ind = substr(orig_prenom_compose_2,1,2) if
(strmatch(orig_prenom_compose_2,"050505?????") & orig_prenom_ind=="")
replace orig_prenom_ind = substr(orig_prenom_compose_2,7,2) if
(strmatch(orig_prenom_compose_2,"090909050505") & orig_prenom_ind=="")
**prénom composé catégorisé de façons différentes par 2 nomenclatures
replace orig_prenom_ind = "25" if (strmatch(orig_prenom_compose_2,"050209?????") &
orig_prenom_ind=="")
replace orig_prenom_ind = "56" if (strmatch(orig_prenom_compose_2,"050609?????") &
orig_prenom_ind=="")
replace orig_prenom_ind = "57" if (strmatch(orig_prenom_compose_2,"050709?????") &
orig_prenom_ind=="")
replace orig_prenom_ind = "26" if (strmatch(orig_prenom_compose_2,"060209?????") &
orig_prenom_ind=="")
replace orig_prenom_ind = "56" if (strmatch(orig_prenom_compose_2,"060509?????") &
orig_prenom_ind=="")

```



```

    replace orig_prenom_ind = "67" if (strmatch(orig_prenom_compose_2,"060709?????") &
orig_prenom_ind=="")
    replace orig_prenom_ind = "25" if (strmatch(orig_prenom_compose_2,"090909050209") &
orig_prenom_ind=="")
    replace orig_prenom_ind = "56" if (strmatch(orig_prenom_compose_2,"090909050609") &
orig_prenom_ind=="")
    replace orig_prenom_ind = "56" if (strmatch(orig_prenom_compose_2,"090909060509") &
orig_prenom_ind=="")
    replace orig_prenom_ind = "56" if (strmatch(orig_prenom_compose_2,"060905?????") &
orig_prenom_ind=="")
    replace orig_prenom_ind = "56" if (strmatch(orig_prenom_compose_2,"090909060905") &
orig_prenom_ind=="")
    replace orig_prenom_ind = "35" if (strmatch(orig_prenom_compose_2,"090305?????") &
orig_prenom_ind=="")
    replace orig_prenom_ind = "56" if (strmatch(orig_prenom_compose_2,"090605?????") &
orig_prenom_ind=="")
    replace orig_prenom_ind = "56" if (strmatch(orig_prenom_compose_2,"090909090605") &
orig_prenom_ind=="")
**prénom composé catégorisé de façons différentes par 3 nomenclatures
    replace orig_prenom_ind = "15" if (strmatch(orig_prenom_compose_2,"050105?????") &
orig_prenom_ind=="")
    replace orig_prenom_ind = "56" if (strmatch(orig_prenom_compose_2,"060505?????") &
orig_prenom_ind=="")
    replace orig_prenom_ind = "56" if (strmatch(orig_prenom_compose_2,"060605?????") &
orig_prenom_ind=="")
    replace orig_prenom_ind = "15" if (strmatch(orig_prenom_compose_2,"090909050105") &
orig_prenom_ind=="")
    replace orig_prenom_ind = "56" if (strmatch(orig_prenom_compose_2,"090909060505") &
orig_prenom_ind=="")
    replace orig_prenom_ind = "56" if (strmatch(orig_prenom_compose_2,"090909060605") &
orig_prenom_ind=="")
/*tab orig_prenom_ind [aw=poids_ind], missing*/
/* 5 229 individus ( 6 % des nouveaux-nés) portent un prénom double dont
l'origine d'un composant est catégorisée par les nomenclatures 1, 2 et 3*/

```

\* (6) Si un prénom composé constitué de 3 "composants" ne voit son origine catégorisée par aucune des 3 nomenclatures, mais qu'au moins l'un de ses composants voit son origine catégorisée par au moins l'une des nomenclatures, alors on le catégorise de l'origine du premier des composants dont l'origine est catégorisée.

```

    replace orig_prenom_ind = substr(orig_prenom_compose_3,13,2) if
(strmatch(orig_prenom_compose_3,"090909090909??0909") & orig_prenom_ind=="")
    replace orig_prenom_ind = substr(orig_prenom_compose_3,1,2) if
(strmatch(orig_prenom_compose_3,"050505????????????") & orig_prenom_ind=="")
    replace orig_prenom_ind = substr(orig_prenom_compose_3,1,2) if
(strmatch(orig_prenom_compose_3,"050509????????????") & orig_prenom_ind=="")
    replace orig_prenom_ind = substr(orig_prenom_compose_3,1,2) if
(strmatch(orig_prenom_compose_3,"060609????????????") & orig_prenom_ind=="")
    replace orig_prenom_ind = substr(orig_prenom_compose_3,3,2) if
(strmatch(orig_prenom_compose_3,"09??09????????????") & orig_prenom_ind=="") /*ici figure le
(piégeux) prénom composé "Maria de Fatima", catégorisé comme originaire d'Europe du Sud
(Portugal), comme il se doit */
    replace orig_prenom_ind = "56" if
(strmatch(orig_prenom_compose_3,"060505????????????") & orig_prenom_ind=="")
    replace orig_prenom_ind = "56" if
(strmatch(orig_prenom_compose_3,"060509????????????") & orig_prenom_ind=="")
    replace orig_prenom_ind = "56" if
(strmatch(orig_prenom_compose_3,"060605????????????") & orig_prenom_ind=="")
    replace orig_prenom_ind = substr(orig_prenom_compose_3,7,2) if
(strmatch(orig_prenom_compose_3,"090909050505????") & orig_prenom_ind=="")
    replace orig_prenom_ind = "56" if
(strmatch(orig_prenom_compose_3,"090909060905????") & orig_prenom_ind=="")
    replace orig_prenom_ind = substr(orig_prenom_compose_3,9,2) if
(strmatch(orig_prenom_compose_3,"09090909??09????") & orig_prenom_ind=="")
    replace orig_prenom_ind = substr(orig_prenom_compose_3,13,2) if
(strmatch(orig_prenom_compose_3,"0909090909050509") & orig_prenom_ind=="")
    replace orig_prenom_ind = substr(orig_prenom_compose_3,13,2) if
(strmatch(orig_prenom_compose_3,"0909090909060609") & orig_prenom_ind=="")
    replace orig_prenom_ind = substr(orig_prenom_compose_3,15,2) if
(strmatch(orig_prenom_compose_3,"09090909090909??09") & orig_prenom_ind=="")
    replace orig_prenom_ind = "56" if
(strmatch(orig_prenom_compose_3,"0909090909060509") & orig_prenom_ind=="")
    replace orig_prenom_ind = "56" if
(strmatch(orig_prenom_compose_3,"0909090909050609") & orig_prenom_ind=="")
    replace orig_prenom_ind = "56" if
(strmatch(orig_prenom_compose_3,"0909090909060905") & orig_prenom_ind=="")
/*tab orig_prenom_ind [aw=poids_ind], missing*/
/* 46 individus ( 0 % des nouveaux-nés) portent un prénom triple dont

```

```

l'origine d'un composant est catégorisée par les nomenclatures 1, 2 et 3*/

* (7) Si un prénom composé constitué de 4 "composants" ne voit son origine catégorisée par
aucune des 3 nomenclatures, mais qu'au moins l'un de ses composants voit son origine
catégorisée par au moins l'une des nomenclatures, alors on le catégorise selon l'origine du
premier des composants dont l'origine est catégorisée.
  replace orig_prenom_ind = substr(orig_prenom_compose_4,3,2) if
(orig_prenom_compose_4=="090709????????????????")
  replace orig_prenom_ind = substr(orig_prenom_compose_4,9,2) if
(orig_prenom_compose_4=="090909090709????????????")
  /*tab orig_prenom_ind [aw=poids_ind], missing*/
  /*      0 individus ( 0 % des nouveaux-nés) portent un prénom quadruple dont
l'origine d'un composant est catégorisée par les nomenclatures 1, 2 et 3*/

*J'examine manuellement dans le fichier si je ne peux pas améliorer le codage
sort orig_prenom_ind prenom_ind
order ident noi poids_ind sexe_ind anais_ind regionnais_ind prenom_ind orig_prenom_ind
lienego_ind orig_ind relig_ind r_impvie x_vufri
sort ident noi
*Parmi les prénoms non catégorisés, certains contiennent probablement des erreurs
d'orthographe, mais dans le doute je n'y touche pas:
/*"ALAEANDRE", "ALEXANDA", "ALESXI", "ALLEXANDRE", "ALEXXANDRE",
"BAPISTE", "BEJAMIN",
"CHISTELLE", "CHISTINNE", "CHRTINE", "CHRSTINE", "CHISTOPHE", "CHTYSTELLE", "CLHOE",
"CHRITOPHE",
"DEPHINE", "DIEGO",
"ELSABETH", "EISABETH",
"GRABRIELLA", "GREGORY", "GREPOR",
"ISABELLE",
"LAETITA", "LEATITIA", "LAETITTIA", "LUCDOVIC",
"MARI CHRISTINE",
"OCSAR",
"STEPHANMIE",
etc.*/

*J'examine le "top 20" des prénoms de chaque modalité de cette variable ORIG_PRENOM_IND, pour
vérifier si elles sont globalement conformes à l'intuition et éventuellement les recoder
codebook prenom_ind if (orig_prenom_ind=="01") /*Prénoms
français":
incluent 1 958 prénoms différents
portés par 34 157 individus, comme "Sébastien", "Michel", "Nathalie", etc. */
  groups prenom_ind if (orig_prenom_ind=="01") [aw=poids_ind], order(h) select(20)
codebook prenom_ind if (orig_prenom_ind=="02") /*Prénoms
latin":
incluent 1 043 prénoms différents
portés par 6 725 individus, comme "Patricia", "Laura", "Sabrina", etc. */
  groups prenom_ind if (orig_prenom_ind=="02") [aw=poids_ind], order(h) select(20)
codebook prenom_ind if (orig_prenom_ind=="03") /*Prénoms
international ou anglophone":
incluent 429 prénoms différents
portés par 4 519 individus, comme "Audrey", "Franck", "Kevin", etc. */
  groups prenom_ind if (orig_prenom_ind=="03") [aw=poids_ind], order(h) select(20)
codebook prenom_ind if (orig_prenom_ind=="04") /*Prénoms
autre Europe":
incluent 354 prénoms différents
portés par 1 466 individus, comme "Karine", "Ingrid", "Dimitri", etc. */
  groups prenom_ind if (orig_prenom_ind=="04") [aw=poids_ind], order(h) select(20)
codebook prenom_ind if (orig_prenom_ind=="05") /*Prénoms
arabo-musulman":
incluent 3 542 prénoms différents
portés par 19 850 individus, comme "Mohamed", "Fatima", "Nadia", etc. */
  groups prenom_ind if (orig_prenom_ind=="05") [aw=poids_ind], order(h) select(20)
codebook prenom_ind if (orig_prenom_ind=="06") /*Prénoms
africain":
incluent 184 prénoms différents
portés par 597 individus, comme "Awa", "Aissatou", "Aissata", etc. */
  groups prenom_ind if (orig_prenom_ind=="06") [aw=poids_ind], order(h) select(20)
codebook prenom_ind if (orig_prenom_ind=="07") /*Prénoms
asiatique":
incluent 862 prénoms différents
portés par 1 303 individus, comme "Maeva", "Noa", "Yan", etc. */
  groups prenom_ind if (orig_prenom_ind=="07") [aw=poids_ind], order(h) select(20)
codebook prenom_ind if (orig_prenom_ind=="08") /*Prénoms
juif":
incluent 50 prénoms différents
portés par 793 individus, comme "Sarah", "Myriam", "Samuel", etc. */
  groups prenom_ind if (orig_prenom_ind=="08") [aw=poids_ind], order(h) select(20)
codebook prenom_ind if (orig_prenom_ind=="09") /*Prénoms
inclassable":
incluent 9 275 prénoms différents
portés par 11 666 individus, comme "Fiona", "Killian", "Teo", etc. */
  groups prenom_ind if (orig_prenom_ind=="09") [aw=poids_ind], order(h) select(20)
codebook prenom_ind if (orig_prenom_ind=="15") /*Prénoms
français et arabo-musulman":
incluent 14 prénoms différents
portés par 147 individus, comme "Sofiane", "Sophiane", "Omer", etc. */
/*groups prenom_ind if (orig_prenom_ind=="15") [aw=poids_ind], order(h) select(20)*/

```

```

codebook prenom_ind if (orig_prenom_ind=="16")                                /*Prénoms
"français et africain":                                                    incluent      4 prénoms différents
portés par      26 individus, comme "Candice", "Marième", "Dédé" et "Dior" */
/*groups prenom_ind if (orig_prenom_ind=="16") [aw=poids_ind], order(h) select(20)*/
codebook prenom_ind if (orig_prenom_ind=="25")                                /*Prénoms
"latin et arabo-musulman":                                                incluent     16 prénoms différents
portés par     178 individus, comme "Aminata", "Manel", "Orane", etc. */
/*groups prenom_ind if (orig_prenom_ind=="25") [aw=poids_ind], order(h) select(20)*/
codebook prenom_ind if (orig_prenom_ind=="256")                             /*Prénoms
"latin, arabo-musulman et africain":                                       incluent      1 prénom
porté par      34 individus,      "Adama" */
groups prenom_ind if (orig_prenom_ind=="256") [aw=poids_ind], order(h) select(20)
codebook prenom_ind if (orig_prenom_ind=="26")                             /*Prénoms
"latin et africain":                                                       incluent      3 prénoms différents
portés par     47 individus,      "Binta", "Fanta" et "Diogo" */
/*groups prenom_ind if (orig_prenom_ind=="26") [aw=poids_ind], order(h) select(20)*/
codebook prenom_ind if (orig_prenom_ind=="35")                             /*Prénoms
"international ou anglophone, et arabo-musulman":                       incluent      5 prénoms différents
portés par     18 individus, comme "Mody", "Sammy", "Duran", etc. */
groups prenom_ind if (orig_prenom_ind=="35") [aw=poids_ind], order(h) select(20)
codebook prenom_ind if (orig_prenom_ind=="36")                             /*Prénoms
"international ou anglophone, et africain":                               incluent      2 prénoms différents
portés par      5 individus,      "Candy" et "Tiana" */
/*groups prenom_ind if (orig_prenom_ind=="36") [aw=poids_ind], order(h) select(20)*/
codebook prenom_ind if (orig_prenom_ind=="45")                             /*Prénoms
"autre Europe et arabo-musulman":                                          incluent      7 prénoms différents
portés par     31 individus, comme "Wassila", "Nada", "Liès", etc. */
/*groups prenom_ind if (orig_prenom_ind=="45") [aw=poids_ind], order(h) select(20)*/
codebook prenom_ind if (orig_prenom_ind=="46")                             /*Prénoms
"autre Europe et africain":                                                incluent      1 prénom
porté par      2 individus,      "Lala" */
groups prenom_ind if (orig_prenom_ind=="46") [aw=poids_ind], order(h) select(20)
codebook prenom_ind if (orig_prenom_ind=="56")                             /*Prénoms
"arabo-musulman et africain":                                              incluent     275 prénoms différents
portés par 1 416 individus, comme "Mamadou", "Hadda", "Tidiane", etc. */
groups prenom_ind if (orig_prenom_ind=="56") [aw=poids_ind], order(h) select(20)
codebook prenom_ind if (orig_prenom_ind=="567")                           /*Prénoms
"arabo-musulman, africain et asiatique":                                  incluent      1 prénom
porté par      1 individu,      "s" */
/*groups prenom_ind if (orig_prenom_ind=="567") [aw=poids_ind], order(h) select(20)*/
codebook prenom_ind if (orig_prenom_ind=="57")                             /*Prénoms
"arabo-musulman et asiatique":                                            incluent     16 prénoms différents
portés par     47 individus, comme "May", "Keltoum", "Hana", etc. */
/*groups prenom_ind if (orig_prenom_ind=="57") [aw=poids_ind], order(h) select(20)*/
codebook prenom_ind if (orig_prenom_ind=="58")                             /*Prénoms
"arabo-musulman et juif":                                                 incluent      6 prénoms différents
portés par     15 individus, comme "Aaron", "Nissim", "Rivka", etc. */
/*groups prenom_ind if (orig_prenom_ind=="58") [aw=poids_ind], order(h) select(20)*/
codebook prenom_ind if (orig_prenom_ind=="67")                             /*Prénoms
"africain et asiatique":                                                  incluent      2 prénoms différents
portés par      2 individus,      "Sokha" et "Sokha Ry" */
/*groups prenom_ind if (orig_prenom_ind=="67") [aw=poids_ind], order(h) select(20)*/

```

```

*Je regroupe les catégories trop pauvres en prénoms, selon la façon dont behindthename et
Wikipedia catégorisent le prénom le plus fréquent de chacune d'entre elles
replace orig_prenom_ind = "05" if (orig_prenom_ind=="15") /*"français et arabo-musulman"
devient "arabo-musulman"*/
replace orig_prenom_ind = "06" if (orig_prenom_ind=="16") /*"français et africain"
devient "africain"*/
replace orig_prenom_ind = "05" if (orig_prenom_ind=="25") /*"latin et arabo-musulman"
devient "arabo-musulman"*/
replace orig_prenom_ind = "06" if (orig_prenom_ind=="256") /*"latin, arabo-musulman et
africain" devient "africain"*/
replace orig_prenom_ind = "06" if (orig_prenom_ind=="26") /*"latin et africain"
devient "africain"*/
replace orig_prenom_ind = "03" if (orig_prenom_ind=="35") /*"international ou anglophone, et
arabo-musulman" devient "international ou anglophone"*/
replace orig_prenom_ind = "03" if (orig_prenom_ind=="36") /*"international ou anglophone, et
africain" devient "international ou anglophone"*/
replace orig_prenom_ind = "05" if (orig_prenom_ind=="45") /*"autre Europe, et arabo-musulman"
devient "arabo-musulman"*/
replace orig_prenom_ind = "04" if (orig_prenom_ind=="46") /*"autre Europe et africain"
devient "autre Europe"*/
replace orig_prenom_ind = "06" if (orig_prenom_ind=="56") /*"arabo-musulman et africain"
devient "africain"*/
replace orig_prenom_ind = "09" if (orig_prenom_ind=="567") /*"arabo-musulman, africain et
asiatique" devient "inclassable"*/

```

```

replace orig_prenom_ind = "03" if (orig_prenom_ind=="57") /**arabo-musulman et asiatique"
devient "arabo-musulman"*/
replace orig_prenom_ind = "08" if (orig_prenom_ind=="58") /**arabo-musulman et juif"
devient "juif"*/
replace orig_prenom_ind = "07" if (orig_prenom_ind=="67") /**africain et asiatique"
devient "asiatique"*/
replace orig_prenom_ind = "08" if (orig_prenom_ind=="68") /**africain et juif"
devient "juif"*/ /*ce recodage-là ne sera utile que pour catégoriser les prénoms du Fichier
INSEE des prénoms, ci-dessous*/
    tab orig_prenom_ind [aw=poids_ind], missing

*J'applique des labels à la variable ORIG_PRENOM_IND
destring orig_prenom_ind, replace
    label variable orig_prenom_ind "origine du prénom"
    label define prenomind 1 "Français" 2 "Latin" 3 "International ou anglophone" 4 "Autre
Europe" 5 "Arabo-musulman" 6 "Africain" 7 "Asiatique" 8 "Juif" 9 "Inclassable"
    label values orig_prenom_ind prenomind
    tab orig_prenom_ind [aw=poids_ind], missing /**/

save repro, replace

*ORIG_PRENOM_IND est finalement codée comme suit:
/*1=Français
2=Latin
3=International ou anglophone
4=Autre Europe
5=Arabo-musulman
6=Africain
7=Asiatique
8=Juif
9=Inclassable*/

/*Chapitre 6: La nomenclature synthétique des prénoms par origine: appariement au Fichier
INSEE des prénoms*/

*J'apparie les nomenclatures 1, 2 et 3 (nomenclature synthétique) au Fichier INSEE des prénoms
    use nomenclature1
    rename prenom_ind4 prenom_ind
    rename nomenclature1_ind4 nomenclature1
    save nomenclature1, replace
    use nomenclature2
    rename prenom_ind4 prenom_ind
    rename nomenclature2_ind4 nomenclature2
    save nomenclature2, replace
    use nomenclature3
    rename prenom_ind4 prenom_ind
    rename nomenclature3_ind4 nomenclature3
    save nomenclature3, replace

use insee
merge m:1 prenom_ind sexe_ind using nomenclature1.dta, keepusing(nomenclature1) keep(master
match)
    drop _merge
merge m:1 prenom_ind sexe_ind using nomenclature2.dta, keepusing(nomenclature2) keep(master
match)
    drop _merge
merge m:1 prenom_ind sexe_ind using nomenclature3.dta, keepusing(nomenclature3) keep(master
match)
    drop _merge
    tab nomenclature1 [aw=nombre], missing           /*la nomenclature 1 s'apparie à 4 % des
prénoms du Fichier INSEE*/
    tab nomenclature2 [aw=nombre], missing           /*la nomenclature 2 s'apparie à 91 % des
prénoms du Fichier INSEE*/
    tab nomenclature3 [aw=nombre], missing           /*la nomenclature 3 s'apparie à 49 % des
prénoms du Fichier INSEE*/

*Je recode les nomenclatures 1 à 3
    tab nomenclature1 [aw=nombre], missing           /**/
    replace nomenclature1="09" if (nomenclature1=="")
    tab nomenclature1 [aw=nombre], missing           /**/

    tab nomenclature2 [aw=nombre], missing           /**/
    replace nomenclature2="09" if (nomenclature2=="")
    tab nomenclature2 [aw=nombre], missing           /**/

    tab nomenclature3 [aw=nombre], missing           /**/

```

```

replace nomenclature3="09" if (nomenclature3=="")
      tab nomenclature3 [aw=nombre], missing      /**/

*J'examine quels sont les prénoms non catégorisés les plus fréquents
generate nomenclatures = 0 if (nomenclature1=="09" & nomenclature2=="09" &
nomenclature3=="09")
      tab nomenclatures [aw=nombre], missing      /*7 % des nouveau-
nés ont un prénom d'origine non catégorisée*/
      groups prenom_ind [aw=nombre] if (nomenclatures==0), order(h) select(100) /*les prénoms non
catégorisés les plus fréquents comprennent "Jean-François", "Marie-Thérèse", "Marie-Christine"
et autres prénoms composés.*/

*Je vais donc chercher à catégoriser les prénoms composés qui ne sont catégorisés par aucune
des 3 premières nomenclatures.
*Pour cela je commence par découper les prénoms composés en plusieurs prénoms composants.
split prenom_ind /*les prénoms composés des individus INSEE sont composés, au plus, de 3
prénoms composants*/
codebook prenom_ind1 /*          652 056 prénoms-année sur 652 056 ont un 1e prénom
composant*/
codebook prenom_ind2 /*seulement  42 270 prénoms-année sur 652 056 ont un 2e prénom
composant*/
codebook prenom_ind3 /*seulement    56 prénoms-année sur 652 056 ont un 3e prénom
composant*/
save insee, replace

*J'apparie les nomenclatures 1, 2 et 3 à chacun des 3 prénoms composants
*J'apparie la nomenclature 1 au 1er prénom composant
      use nomenclature1
rename prenom_ind prenom_ind1
rename nomenclature1 nomenclature1_ind1
      save nomenclature1, replace
      use insee
merge m:1 prenom_ind1 sexe_ind using nomenclature1.dta, keepusing(nomenclature1_ind1)
keep(master match)
drop _merge
save insee, replace
*J'apparie la nomenclature 2 au 1er prénom composant
      use nomenclature2
rename prenom_ind prenom_ind1
rename nomenclature2 nomenclature2_ind1
      save nomenclature2, replace
      use insee
merge m:1 prenom_ind1          using nomenclature2.dta, keepusing(nomenclature2_ind1)
keep(master match)
drop _merge
save insee, replace
*J'apparie la nomenclature 3 au 1er prénom composant
      use nomenclature3
rename prenom_ind prenom_ind1
rename nomenclature3 nomenclature3_ind1
      save nomenclature3, replace
      use insee
merge m:1 prenom_ind1 sexe_ind using nomenclature3.dta, keepusing(nomenclature3_ind1)
keep(master match)
drop _merge
save insee, replace
*J'apparie la nomenclature 1 au 2e prénom composant
      use nomenclature1
rename prenom_ind1 prenom_ind2
rename nomenclature1_ind1 nomenclature1_ind2
      save nomenclature1, replace
      use insee
merge m:1 prenom_ind2 sexe_ind using nomenclature1.dta, keepusing(nomenclature1_ind2)
keep(master match)
drop _merge
save insee, replace
*J'apparie la nomenclature 2 au 2e prénom composant
      use nomenclature2
rename prenom_ind1 prenom_ind2
rename nomenclature2_ind1 nomenclature2_ind2
      save nomenclature2, replace
      use insee
merge m:1 prenom_ind2          using nomenclature2.dta, keepusing(nomenclature2_ind2)
keep(master match)
drop _merge
save insee, replace
*J'apparie la nomenclature 3 au 2e prénom composant

```

```

        use nomenclature3
rename prenom_ind1 prenom_ind2
rename nomenclature3_ind1 nomenclature3_ind2
    save nomenclature3, replace
    use insee
merge m:1 prenom_ind2 sexe_ind using nomenclature3.dta, keepusing(nomenclature3_ind2)
keep(master match)
drop _merge
save insee, replace
*J'apparie la nomenclature 1 au 3e prénom composant
    use nomenclature1
rename prenom_ind2 prenom_ind3
rename nomenclature1_ind2 nomenclature1_ind3
    save nomenclature1, replace
    use insee
merge m:1 prenom_ind3 sexe_ind using nomenclature1.dta, keepusing(nomenclature1_ind3)
keep(master match)
drop _merge
save insee, replace
*J'apparie la nomenclature 2 au 3e prénom composant
    use nomenclature2
rename prenom_ind2 prenom_ind3
rename nomenclature2_ind2 nomenclature2_ind3
    save nomenclature2, replace
    use insee
merge m:1 prenom_ind3 sexe_ind using nomenclature2.dta, keepusing(nomenclature2_ind3)
keep(master match)
drop _merge
save insee, replace
*J'apparie la nomenclature 3 au 3e prénom composant
    use nomenclature3
rename prenom_ind2 prenom_ind3
rename nomenclature3_ind2 nomenclature3_ind3
    save nomenclature3, replace
    use insee
merge m:1 prenom_ind3 sexe_ind using nomenclature3.dta, keepusing(nomenclature3_ind3)
keep(master match)
drop _merge
save insee, replace

use insee/*, clear*/
*Je fais les tris à plat de ces variables
replace nomenclature1_ind1="09" if (nomenclature1_ind1=="")
    tab nomenclature1_ind1, missing /*ok*/
replace nomenclature2_ind1="09" if (nomenclature2_ind1=="")
    tab nomenclature2_ind1, missing /*ok*/
replace nomenclature3_ind1="09" if (nomenclature3_ind1=="")
    tab nomenclature3_ind1, missing /*ok*/
replace nomenclature1_ind2="09" if (nomenclature1_ind2=="")
    tab nomenclature1_ind2, missing /*ok*/
replace nomenclature2_ind2="09" if (nomenclature2_ind2=="")
    tab nomenclature2_ind2, missing /*ok*/
replace nomenclature3_ind2="09" if (nomenclature3_ind2=="")
    tab nomenclature3_ind2, missing /*ok*/
replace nomenclature1_ind3="09" if (nomenclature1_ind3=="")
    tab nomenclature1_ind3, missing /*ok*/
replace nomenclature2_ind3="09" if (nomenclature2_ind3=="")
    tab nomenclature2_ind3, missing /*ok*/
replace nomenclature3_ind3="09" if (nomenclature3_ind3=="")
    tab nomenclature3_ind3, missing /*ok*/

*Enfin je crée orig_prenom_ind, en 6 étapes.
generate orig_prenom_simple = nomenclature1_ind1 + nomenclature2_ind1 + nomenclature3_ind1 if
((prenom_ind=="") | (prenom_ind1!="" & prenom_ind2==""))
    /*tab orig_prenom_simple, missing*/
generate orig_prenom_compose_2 = nomenclature1_ind1 + nomenclature2_ind1 + nomenclature3_ind1
+ nomenclature1_ind2 + nomenclature2_ind2 + nomenclature3_ind2 if (prenom_ind1!="" &
prenom_ind2!="" & prenom_ind3=="")
    /*tab orig_prenom_compose_2, missing*/
generate orig_prenom_compose_3 = nomenclature1_ind1 + nomenclature2_ind1 + nomenclature3_ind1
+ nomenclature1_ind2 + nomenclature2_ind2 + nomenclature3_ind2 + nomenclature1_ind3 +
nomenclature2_ind3 + nomenclature3_ind3 if (prenom_ind1!="" & prenom_ind2!="" &
prenom_ind3!="")
    /*tab orig_prenom_compose_3, missing*/

* (1) Si un prénom simple est non renseigné ou ne voit son origine catégorisée par aucune des
3 nomenclatures, ou si un prénom composé constitué de 2 à 3 "composants" (termes séparés par

```

un trait d'union, un espace, une apostrophe, etc.) ne voit son origine catégorisée par aucune des 3 nomenclatures et qu'aucun de ses composants ne voit son origine catégorisée par l'une des nomenclatures, alors on le catégorise comme "inclassable".

```
generate orig_prenom_ind = "09" if (orig_prenom_simple=="090909")
  replace orig_prenom_ind = "09" if (orig_prenom_compose_2=="090909090909")
  replace orig_prenom_ind = "09" if (orig_prenom_compose_3=="0909090909090909")
    tab orig_prenom_ind [aw=nombre], missing
/* 6 % des nouveau-nés se sont vus attribuer un prénom simple dont
l'origine n'est catégorisée par aucune des 3 nomenclatures*/
```

\* (2) Si un prénom simple ne voit son origine catégorisée que par une seule des 3 nomenclatures, alors on le catégorise de ladite origine.

```
**prénom simple catégorisé par 1 seule nomenclature : la nomenclature 1
  replace orig_prenom_ind = substr(orig_prenom_simple,1,2) if
(strmatch(orig_prenom_simple,"??0909") & orig_prenom_ind=="")
**prénom simple catégorisé par 1 seule nomenclature : la nomenclature 2
  replace orig_prenom_ind = substr(orig_prenom_simple,3,2) if
(strmatch(orig_prenom_simple,"09??09") & orig_prenom_ind=="")
**prénom simple catégorisé par 1 seule nomenclature : la nomenclature 3
  replace orig_prenom_ind = substr(orig_prenom_simple,5,2) if
(strmatch(orig_prenom_simple,"0909??") & orig_prenom_ind=="")
    tab orig_prenom_ind [aw=nombre], missing
/*87 % des nouveau-nés se sont vus attribuer un prénom simple dont
l'origine n'est catégorisée que par la nomenclature 1, ou la 2, ou la 3*/
```

\* (3) Si un prénom simple voit son origine catégorisée par au moins 2 des 3 nomenclatures et que ces origines sont les mêmes, alors on le catégorise de ladite origine.

```
**prénom simple catégorisé de la même façon par 2 nomenclatures : les nomenclatures 1 et 2
  replace orig_prenom_ind = "05" if (orig_prenom_simple=="050509")
  replace orig_prenom_ind = "06" if (orig_prenom_simple=="060609")
**prénom simple catégorisé de la même façon par 2 nomenclatures : les nomenclatures 1 et 3
  replace orig_prenom_ind = "05" if (orig_prenom_simple=="050905")
**prénom simple catégorisé de la même façon par 2 nomenclatures : les nomenclatures 2 et 3
  replace orig_prenom_ind = "05" if (orig_prenom_simple=="090505")
**prénom simple catégorisé de la même façon par 3 nomenclatures : les nomenclatures 1, 2 et 3
  replace orig_prenom_ind = "05" if (orig_prenom_simple=="050505")
    tab orig_prenom_ind [aw=nombre], missing
/* 3 % des nouveau-nés se sont vus attribuer un prénom simple dont l'origine
est catégorisée de façon identique par les nomenclatures 1, 2 et/ou 3*/
```

\* (4) Si un prénom simple voit son origine catégorisée par au moins 2 des 3 nomenclatures mais que ces origines ne sont pas les mêmes, alors on le catégorise d'une origine mixte ou ambiguë, constituée des 2 ou 3 origines différentes attribuées par les nomenclatures.

```
**prénom simple catégorisé de façons différentes par 2 nomenclatures : les nomenclatures 1 et 2
  replace orig_prenom_ind = "15" if (orig_prenom_simple=="050109")
  replace orig_prenom_ind = "25" if (orig_prenom_simple=="050209")
  replace orig_prenom_ind = "35" if (orig_prenom_simple=="050309")
  replace orig_prenom_ind = "45" if (orig_prenom_simple=="050409")
  replace orig_prenom_ind = "56" if (orig_prenom_simple=="050609")
  replace orig_prenom_ind = "57" if (orig_prenom_simple=="050709")
  replace orig_prenom_ind = "58" if (orig_prenom_simple=="050809")
  replace orig_prenom_ind = "16" if (orig_prenom_simple=="060109")
  replace orig_prenom_ind = "26" if (orig_prenom_simple=="060209")
  replace orig_prenom_ind = "36" if (orig_prenom_simple=="060309")
  replace orig_prenom_ind = "46" if (orig_prenom_simple=="060409")
  replace orig_prenom_ind = "56" if (orig_prenom_simple=="060509")
  replace orig_prenom_ind = "67" if (orig_prenom_simple=="060709")
  replace orig_prenom_ind = "68" if (orig_prenom_simple=="060809")
**prénom simple catégorisé de façons différentes par 2 nomenclatures : les nomenclatures 1 et 3
  replace orig_prenom_ind = "56" if (orig_prenom_simple=="060905")
**prénom simple catégorisé de façons différentes par 2 nomenclatures : les nomenclatures 2 et 3
  replace orig_prenom_ind = "15" if (orig_prenom_simple=="090105")
  replace orig_prenom_ind = "25" if (orig_prenom_simple=="090205")
  replace orig_prenom_ind = "35" if (orig_prenom_simple=="090305")
  replace orig_prenom_ind = "45" if (orig_prenom_simple=="090405")
  replace orig_prenom_ind = "56" if (orig_prenom_simple=="090605")
  replace orig_prenom_ind = "57" if (orig_prenom_simple=="090705")
  replace orig_prenom_ind = "58" if (orig_prenom_simple=="090805")
**prénom simple catégorisé de façons différentes par 3 nomenclatures : les nomenclatures 1, 2
et 3
  replace orig_prenom_ind = "15" if (orig_prenom_simple=="050105")
  replace orig_prenom_ind = "25" if (orig_prenom_simple=="050205")
  replace orig_prenom_ind = "35" if (orig_prenom_simple=="050305")
  replace orig_prenom_ind = "45" if (orig_prenom_simple=="050405")
```

```

replace orig_prenom_ind = "56" if (orig_prenom_simple=="050605")
replace orig_prenom_ind = "57" if (orig_prenom_simple=="050705")
replace orig_prenom_ind = "58" if (orig_prenom_simple=="050805")
replace orig_prenom_ind = "156" if (orig_prenom_simple=="060105")
replace orig_prenom_ind = "256" if (orig_prenom_simple=="060205")
replace orig_prenom_ind = "356" if (orig_prenom_simple=="060305")
replace orig_prenom_ind = "456" if (orig_prenom_simple=="060405")
replace orig_prenom_ind = "56" if (orig_prenom_simple=="060505")
replace orig_prenom_ind = "56" if (orig_prenom_simple=="060605")
replace orig_prenom_ind = "567" if (orig_prenom_simple=="060705")
replace orig_prenom_ind = "568" if (orig_prenom_simple=="060805")
/*tab orig_prenom_ind [aw=nombre], missing*/
/* 1 % des nouveau-nés se sont vus attribuer un prénom simple dont
l'origine est catégorisée de façons différentes par les nomenclatures 1, 2 et/ou 3*/

* (5) Si un prénom composé constitué de 2 "composants" ne voit son origine catégorisée par
aucune des 3 nomenclatures, mais qu'au moins l'un de ses composants voit son origine
catégorisée par au moins l'une des nomenclatures, alors on le catégorise selon l'origine du
premier des composants dont l'origine est catégorisée.
**prénom composé catégorisé par 1 seule nomenclature
replace orig_prenom_ind = substr(orig_prenom_compose_2,1,2) if
(strmatch(orig_prenom_compose_2,"050909?????") & orig_prenom_ind=="")
replace orig_prenom_ind = substr(orig_prenom_compose_2,1,2) if
(strmatch(orig_prenom_compose_2,"060909?????") & orig_prenom_ind=="")
replace orig_prenom_ind = substr(orig_prenom_compose_2,7,2) if
(strmatch(orig_prenom_compose_2,"090909??0909") & orig_prenom_ind=="")
replace orig_prenom_ind = substr(orig_prenom_compose_2,3,2) if
(strmatch(orig_prenom_compose_2,"09??09?????") & orig_prenom_ind=="")
replace orig_prenom_ind = substr(orig_prenom_compose_2,9,2) if
(strmatch(orig_prenom_compose_2,"09090909??09") & orig_prenom_ind=="")
replace orig_prenom_ind = substr(orig_prenom_compose_2,5,2) if
(strmatch(orig_prenom_compose_2,"0909???????") & orig_prenom_ind=="")
replace orig_prenom_ind = substr(orig_prenom_compose_2,11,2) if
(strmatch(orig_prenom_compose_2,"0909090909??") & orig_prenom_ind=="")
**prénom composé catégorisé de la même façon par 2 nomenclatures
replace orig_prenom_ind = substr(orig_prenom_compose_2,1,2) if
(strmatch(orig_prenom_compose_2,"050509?????") & orig_prenom_ind=="")
replace orig_prenom_ind = substr(orig_prenom_compose_2,1,2) if
(strmatch(orig_prenom_compose_2,"060609?????") & orig_prenom_ind=="")
replace orig_prenom_ind = substr(orig_prenom_compose_2,7,2) if
(strmatch(orig_prenom_compose_2,"090909050509") & orig_prenom_ind=="")
replace orig_prenom_ind = substr(orig_prenom_compose_2,7,2) if
(strmatch(orig_prenom_compose_2,"090909060609") & orig_prenom_ind=="")
replace orig_prenom_ind = substr(orig_prenom_compose_2,7,2) if
(strmatch(orig_prenom_compose_3,"0909090505090909") & orig_prenom_ind=="")
replace orig_prenom_ind = substr(orig_prenom_compose_2,7,2) if
(strmatch(orig_prenom_compose_3,"0909090606090909") & orig_prenom_ind=="")
replace orig_prenom_ind = substr(orig_prenom_compose_2,1,2) if
(strmatch(orig_prenom_compose_2,"050905?????") & orig_prenom_ind=="")
replace orig_prenom_ind = substr(orig_prenom_compose_2,7,2) if
(strmatch(orig_prenom_compose_2,"090909050905") & orig_prenom_ind=="")
replace orig_prenom_ind = substr(orig_prenom_compose_2,3,2) if
(strmatch(orig_prenom_compose_2,"090505?????") & orig_prenom_ind=="")
replace orig_prenom_ind = substr(orig_prenom_compose_2,9,2) if
(strmatch(orig_prenom_compose_2,"090909090505") & orig_prenom_ind=="")
**prénom composé catégorisé de la même façon par 3 nomenclatures : les nomenclatures 1, 2 et 3
replace orig_prenom_ind = substr(orig_prenom_compose_2,1,2) if
(strmatch(orig_prenom_compose_2,"050505?????") & orig_prenom_ind=="")
replace orig_prenom_ind = substr(orig_prenom_compose_2,7,2) if
(strmatch(orig_prenom_compose_2,"090909050505") & orig_prenom_ind=="")
**prénom composé catégorisé de façons différentes par 2 nomenclatures
replace orig_prenom_ind = "25" if (strmatch(orig_prenom_compose_2,"050209?????") &
orig_prenom_ind=="")
replace orig_prenom_ind = "56" if (strmatch(orig_prenom_compose_2,"050609?????") &
orig_prenom_ind=="")
replace orig_prenom_ind = "57" if (strmatch(orig_prenom_compose_2,"050709?????") &
orig_prenom_ind=="")
replace orig_prenom_ind = "26" if (strmatch(orig_prenom_compose_2,"060209?????") &
orig_prenom_ind=="")
replace orig_prenom_ind = "56" if (strmatch(orig_prenom_compose_2,"060509?????") &
orig_prenom_ind=="")
replace orig_prenom_ind = "67" if (strmatch(orig_prenom_compose_2,"060709?????") &
orig_prenom_ind=="")
replace orig_prenom_ind = "25" if (strmatch(orig_prenom_compose_2,"090909050209") &
orig_prenom_ind=="")
replace orig_prenom_ind = "56" if (strmatch(orig_prenom_compose_2,"090909050609") &
orig_prenom_ind=="")

```



```

    replace orig_prenom_ind = "56" if (strmatch(orig_prenom_compose_2,"090909060509") &
orig_prenom_ind=="")
    replace orig_prenom_ind = "56" if (strmatch(orig_prenom_compose_2,"060905??????") &
orig_prenom_ind=="")
    replace orig_prenom_ind = "56" if (strmatch(orig_prenom_compose_2,"090909060905") &
orig_prenom_ind=="")
    replace orig_prenom_ind = "35" if (strmatch(orig_prenom_compose_2,"090305??????") &
orig_prenom_ind=="")
    replace orig_prenom_ind = "56" if (strmatch(orig_prenom_compose_2,"090605??????") &
orig_prenom_ind=="")
    replace orig_prenom_ind = "56" if (strmatch(orig_prenom_compose_2,"090909090605") &
orig_prenom_ind=="")
**prénom composé catégorisé de façons différentes par 3 nomenclatures
    replace orig_prenom_ind = "15" if (strmatch(orig_prenom_compose_2,"050105??????") &
orig_prenom_ind=="")
    replace orig_prenom_ind = "56" if (strmatch(orig_prenom_compose_2,"060505??????") &
orig_prenom_ind=="")
    replace orig_prenom_ind = "56" if (strmatch(orig_prenom_compose_2,"060605??????") &
orig_prenom_ind=="")
    replace orig_prenom_ind = "15" if (strmatch(orig_prenom_compose_2,"090909050105") &
orig_prenom_ind=="")
    replace orig_prenom_ind = "56" if (strmatch(orig_prenom_compose_2,"090909060505") &
orig_prenom_ind=="")
    replace orig_prenom_ind = "56" if (strmatch(orig_prenom_compose_2,"090909060605") &
orig_prenom_ind=="")
/*tab orig_prenom_ind [aw=nombre], missing*/
/* 3 % des nouveau-nés se sont vus attribuer un prénom double dont
l'origine d'un composant est catégorisée par les nomenclatures 1, 2 et 3*/

* (6) Si un prénom composé constitué de 3 "composants" ne voit son origine catégorisée par
aucune des 3 nomenclatures, mais qu'au moins l'un de ses composants voit son origine
catégorisée par au moins l'une des nomenclatures, alors on le catégorise de l'origine du
premier des composants dont l'origine est catégorisée.
    replace orig_prenom_ind = substr(orig_prenom_compose_3,13,2) if
(strmatch(orig_prenom_compose_3,"090909090909??0909") & orig_prenom_ind=="")
    replace orig_prenom_ind = substr(orig_prenom_compose_3,1,2) if
(strmatch(orig_prenom_compose_3,"050505????????????") & orig_prenom_ind=="")
    replace orig_prenom_ind = substr(orig_prenom_compose_3,1,2) if
(strmatch(orig_prenom_compose_3,"050509????????????") & orig_prenom_ind=="")
    replace orig_prenom_ind = substr(orig_prenom_compose_3,1,2) if
(strmatch(orig_prenom_compose_3,"060609????????????") & orig_prenom_ind=="")
    replace orig_prenom_ind = substr(orig_prenom_compose_3,3,2) if
(strmatch(orig_prenom_compose_3,"09??09????????????") & orig_prenom_ind=="") /*ici figure le
(piégeux) prénom composé "Maria de Fatima", catégorisé comme originaire d'Europe du Sud
(Portugal), comme il se doit */
    replace orig_prenom_ind = "56" if
(strmatch(orig_prenom_compose_3,"060505????????????") & orig_prenom_ind=="")
    replace orig_prenom_ind = "56" if
(strmatch(orig_prenom_compose_3,"060509????????????") & orig_prenom_ind=="")
    replace orig_prenom_ind = "56" if
(strmatch(orig_prenom_compose_3,"060605????????????") & orig_prenom_ind=="")
    replace orig_prenom_ind = substr(orig_prenom_compose_3,7,2) if
(strmatch(orig_prenom_compose_3,"090909050505??????") & orig_prenom_ind=="")
    replace orig_prenom_ind = "56" if
(strmatch(orig_prenom_compose_3,"090909060905??????") & orig_prenom_ind=="")
    replace orig_prenom_ind = substr(orig_prenom_compose_3,9,2) if
(strmatch(orig_prenom_compose_3,"09090909??09??????") & orig_prenom_ind=="")
    replace orig_prenom_ind = substr(orig_prenom_compose_3,13,2) if
(strmatch(orig_prenom_compose_3,"0909090909050509") & orig_prenom_ind=="")
    replace orig_prenom_ind = substr(orig_prenom_compose_3,13,2) if
(strmatch(orig_prenom_compose_3,"0909090909060609") & orig_prenom_ind=="")
    replace orig_prenom_ind = substr(orig_prenom_compose_3,15,2) if
(strmatch(orig_prenom_compose_3,"09090909090909??09") & orig_prenom_ind=="")
    replace orig_prenom_ind = "56" if
(strmatch(orig_prenom_compose_3,"0909090909060509") & orig_prenom_ind=="")
    replace orig_prenom_ind = "56" if
(strmatch(orig_prenom_compose_3,"0909090909050609") & orig_prenom_ind=="")
    replace orig_prenom_ind = "56" if
(strmatch(orig_prenom_compose_3,"0909090909060905") & orig_prenom_ind=="")
/*tab orig_prenom_ind [aw=nombre], missing*/
/* 0 % des nouveau-nés se sont vus attribuer un prénom triple dont
l'origine d'un composant est catégorisée par les nomenclatures 1, 2 et 3*/

*J'examine le "top 20" des prénoms de chaque modalité de cette variable ORIG_PRENOM_IND, pour
vérifier si elles sont globalement conformes à l'intuition et éventuellement les recoder
codebook prenom_ind if (orig_prenom_ind=="01") /*Prénoms
"français": incluent 3 953 prénoms différents

```

```

attribués à 182 596 prénoms-sexe-année, comme "Marie", "Jean", "Pierre", etc. */
groups prenom_ind if (orig_prenom_ind=="01") [aw=nombre], order(h) select(10)
codebook prenom_ind if (orig_prenom_ind=="02") /*Prénoms
"latin": incluent 1 691 prénoms différents
attribués à 59 215 prénoms-sexe-année, comme "Patricia", "Laura", "Hugo", etc. */
groups prenom_ind if (orig_prenom_ind=="02") [aw=nombre], order(h) select(10)
codebook prenom_ind if (orig_prenom_ind=="03") /*Prénoms
"international ou anglophone": incluent 648 prénoms différents
attribués à 31 023 prénoms-sexe-année, comme "Franck", "Kevin", "Emma", etc. */
groups prenom_ind if (orig_prenom_ind=="03") [aw=nombre], order(h) select(10)
codebook prenom_ind if (orig_prenom_ind=="04") /*Prénoms
"autre Europe": incluent 658 prénoms différents
attribués à 20 852 prénoms-sexe-année, comme "Karine", "Lena", "Sacha", etc. */
groups prenom_ind if (orig_prenom_ind=="04") [aw=nombre], order(h) select(10)
codebook prenom_ind if (orig_prenom_ind=="05") /*Prénoms
"arabo-musulman": incluent 6 445 prénoms différents
attribués à 128 966 prénoms-sexe-année, comme "Inès", "Mohamed", "Nadia", etc. */
groups prenom_ind if (orig_prenom_ind=="05") [aw=nombre], order(h) select(10)
codebook prenom_ind if (orig_prenom_ind=="06") /*Prénoms
"africain": incluent 268 prénoms différents
attribués à 4 921 prénoms-sexe-année, comme "Shanna", "Awa", "Aissata", etc. */
groups prenom_ind if (orig_prenom_ind=="06") [aw=nombre], order(h) select(10)
codebook prenom_ind if (orig_prenom_ind=="07") /*Prénoms
"asiatique": incluent 283 prénoms différents
attribués à 4 558 prénoms-sexe-année, comme "Maeva", "Noa", "Loan", etc. */
groups prenom_ind if (orig_prenom_ind=="07") [aw=nombre], order(h) select(10)
codebook prenom_ind if (orig_prenom_ind=="08") /*Prénoms
"juif": incluent 84 prénoms différents
attribués à 3 815 prénoms-sexe-année, comme "Sarah", "Myriam", "Samuel", etc. */
groups prenom_ind if (orig_prenom_ind=="08") [aw=nombre], order(h) select(10)
codebook prenom_ind if (orig_prenom_ind=="09") /*Prénoms
"inclassable": incluent 17 072 prénoms différents
attribués à 203 571 prénoms-sexe-année, comme "Ethan", "Romane", "Noah", etc. */
groups prenom_ind if (orig_prenom_ind=="09") [aw=nombre], order(h) select(10)
codebook prenom_ind if (orig_prenom_ind=="15") /*Prénoms
"français et arabo-musulman": incluent 19 prénoms différents
attribués à 899 prénoms-sexe-année, comme "Sofiane", "Omer", "Elies", etc. */
groups prenom_ind if (orig_prenom_ind=="15") [aw=nombre], order(h) select(10)
codebook prenom_ind if (orig_prenom_ind=="16") /*Prénoms
"français et africain": incluent 7 prénoms différents
attribués à 229 prénoms-sexe-année, comme "Candice", "Marième", "Candie" etc. */
groups prenom_ind if (orig_prenom_ind=="16") [aw=nombre], order(h) select(10)
codebook prenom_ind if (orig_prenom_ind=="25") /*Prénoms
"latin et arabo-musulman": incluent 24 prénoms différents
attribués à 835 prénoms-sexe-année, comme "Manel", "Aminata", "Orane", etc. */
groups prenom_ind if (orig_prenom_ind=="25") [aw=nombre], order(h) select(10)
codebook prenom_ind if (orig_prenom_ind=="256") /*Prénoms
"latin, arabo-musulman et africain": incluent 1 prénom
attribué à 47 prénoms-sexe-année, "Adama" */
groups prenom_ind if (orig_prenom_ind=="256") [aw=nombre], order(h) select(10)
codebook prenom_ind if (orig_prenom_ind=="26") /*Prénoms
"latin et africain": incluent 6 prénoms différents
attribués à 151 prénoms-sexe-année, "Fanta", "Binta", "Diogo", etc. */
groups prenom_ind if (orig_prenom_ind=="26") [aw=nombre], order(h) select(10)
codebook prenom_ind if (orig_prenom_ind=="35") /*Prénoms
"international ou anglophone, et arabo-musulman": incluent 5 prénoms différents
attribués à 145 prénoms-sexe-année, comme "Sammy", "Arda", "Mody", etc. */
groups prenom_ind if (orig_prenom_ind=="35") [aw=nombre], order(h) select(10)
codebook prenom_ind if (orig_prenom_ind=="36") /*Prénoms
"international ou anglophone, et africain": incluent 2 prénoms différents
attribués à 107 prénoms-sexe-année, "Candy" et "Tiana" */
groups prenom_ind if (orig_prenom_ind=="36") [aw=nombre], order(h) select(10)
codebook prenom_ind if (orig_prenom_ind=="45") /*Prénoms
"autre Europe et arabo-musulman": incluent 13 prénoms différents
attribués à 365 prénoms-sexe-année, comme "Nada", "Wassila", "Aksel", etc. */
groups prenom_ind if (orig_prenom_ind=="45") [aw=nombre], order(h) select(10)
codebook prenom_ind if (orig_prenom_ind=="46") /*Prénoms
"autre Europe et africain": incluent 1 prénom
attribués à 46 prénoms-sexe-année, "Lala" */
groups prenom_ind if (orig_prenom_ind=="46") [aw=nombre], order(h) select(10)
codebook prenom_ind if (orig_prenom_ind=="56") /*Prénoms
"arabo-musulman et africain": incluent 381 prénoms différents
attribués à 8 818 prénoms-sexe-année, comme "Adem", "Mamadou", "Tamara", etc. */
groups prenom_ind if (orig_prenom_ind=="56") [aw=nombre], order(h) select(10)
codebook prenom_ind if (orig_prenom_ind=="567") /*Prénoms
"arabo-musulman, africain et asiatique": incluent 1 prénom
attribué à 7 prénoms-sexe-année, "Silamakan" */

```

```

    groups prenom_ind if (orig_prenom_ind=="567") [aw=nombre], order(h) select(10)
codebook prenom_ind if (orig_prenom_ind=="57") /*Prénoms
"arabo-musulman et asiatique": incluent 27 prénoms différents
attribués à 607 prénoms-sexe-année, comme "Ylan", "Hana", "May", etc. */
    groups prenom_ind if (orig_prenom_ind=="57") [aw=nombre], order(h) select(10)
codebook prenom_ind if (orig_prenom_ind=="58") /*Prénoms
"arabo-musulman et juif": incluent 9 prénoms différents
attribués à 250 prénoms-sexe-année, comme "Aaron", "Rivka", "Eliakim", etc. */
    groups prenom_ind if (orig_prenom_ind=="58") [aw=nombre], order(h) select(10)
codebook prenom_ind if (orig_prenom_ind=="67") /*Prénoms
"africain et asiatique": incluent 1 prénom
attribué à 3 prénoms-sexe-année, "Kouadio" */
    groups prenom_ind if (orig_prenom_ind=="67") [aw=nombre], order(h) select(10)
codebook prenom_ind if (orig_prenom_ind=="68") /*Prénoms
"africain et juif": incluent 2 prénoms différents
attribués à 30 prénoms-sexe-année, "Mardoche" et "Mardochee" */
    groups prenom_ind if (orig_prenom_ind=="68") [aw=nombre], order(h) select(10)

*Je regroupe les catégories trop pauvres en prénoms, selon la façon dont behindthename et
Wikipedia catégorisent le prénom le plus fréquent de chacune d'entre elles
replace orig_prenom_ind = "05" if (orig_prenom_ind=="15") /*"français et arabo-musulman"
devient "arabo-musulman"*/
replace orig_prenom_ind = "06" if (orig_prenom_ind=="16") /*"français et africain"
devient "africain"*/
replace orig_prenom_ind = "05" if (orig_prenom_ind=="25") /*"latin et arabo-musulman"
devient "arabo-musulman"*/
replace orig_prenom_ind = "06" if (orig_prenom_ind=="256") /*"latin, arabo-musulman et
africain" devient "africain"*/
replace orig_prenom_ind = "06" if (orig_prenom_ind=="26") /*"latin et africain"
devient "africain"*/
replace orig_prenom_ind = "03" if (orig_prenom_ind=="35") /*"international ou anglophone, et
arabo-musulman" devient "international ou anglophone"*/
replace orig_prenom_ind = "03" if (orig_prenom_ind=="36") /*"international ou anglophone, et
africain" devient "international ou anglophone"*/
replace orig_prenom_ind = "05" if (orig_prenom_ind=="45") /*"autre Europe, et arabo-musulman"
devient "arabo-musulman"*/
replace orig_prenom_ind = "04" if (orig_prenom_ind=="46") /*"autre Europe et africain"
devient "autre Europe"*/
replace orig_prenom_ind = "06" if (orig_prenom_ind=="56") /*"arabo-musulman et africain"
devient "africain"*/
replace orig_prenom_ind = "09" if (orig_prenom_ind=="567") /*"arabo-musulman, africain et
asiatique" devient "inclassable"*/
replace orig_prenom_ind = "03" if (orig_prenom_ind=="57") /*"arabo-musulman et asiatique"
devient "arabo-musulman"*/
replace orig_prenom_ind = "08" if (orig_prenom_ind=="58") /*"arabo-musulman et juif"
devient "juif"*/
replace orig_prenom_ind = "07" if (orig_prenom_ind=="67") /*"africain et asiatique"
devient "asiatique"*/
replace orig_prenom_ind = "08" if (orig_prenom_ind=="68") /*"africain et juif"
devient "juif"*/
    tab orig_prenom_ind [aw=nombre], missing

*J'applique des labels à la variable ORIG_PRENOM_IND
destring orig_prenom_ind, replace
    label variable orig_prenom_ind "origine du prénom"
    label define prenomind 1 "Français" 2 "Latin" 3 "International ou anglophone" 4 "Autre
Europe" 5 "Arabo-musulman" 6 "Africain" 7 "Asiatique" 8 "Juif" 9 "Inclassable"
    label values orig_prenom_ind prenomind
    tab orig_prenom_ind [aw=nombre], missing /**/

save insee, replace

*ORIG_PRENOM_IND est finalement codée comme suit:
/*1=Français
2=Latin
3=International ou anglophone
4=Autre Europe
5=Arabo-musulman
6=Africain
7=Asiatique
8=Juif
9=Inclassable*/

*Je supprime les 3 bases NOMENCLATURE1, NOMENCLATURE2 et NOMENCLATURE3, qui ont permis de
construire la nomenclature synthétique
erase nomenclature1.dta
erase nomenclature2.dta

```

erase nomenclature3.dta

/\*Chapitre 6: La nomenclature des prénoms par origine de l'INED: appariement à l'enquête TeO (base REPRO)\*/

```
*J'examine les 2 nomenclatures de l'INED issues de la table THL2_CAT1
use thl2_CAT1
*IDENT (identifiant de l'enquêté)
summarize ident /*Cette variable est bien renseignée pour les 75 121
habitants du logement et va de 1 à 21 761.*/
*NOI (numéro d'ordre de l'individu dans le logement)
/*tab noi, missing*/ /*Cette variable est bien renseignée pour les 75 121
habitants du logement et va de 1 à 19.*/
*CAT0
tab cat0, missing /*Cette variable va de 129 à 999 et n'est pas renseignée
pour 7 individus*/
recode cat0 (.=698)
rename cat0 cat0_thl
*CAT1
/*tab cat1, missing*/ /*Cette variable va de 0 à 698 et n'est pas renseignée
pour 7 individus*/
recode cat1 (.=698)
rename cat1 cat1_thl
save thl2_CAT1, replace
```

```
*J'examine les 4 nomenclatures de l'INED issues de la table INDIV2_CAT1
use indiv2_CAT1
*IDENT (identifiant de l'enquêté)
summarize ident /*Cette variable est bien renseignée pour les 21 761
enquêtés et va de 1 à 21 761.*/
*NOI (numéro d'ordre de l'enquêté dans le logement)
/*tab ne, missing*/ /*Cette variable est bien renseignée pour les 21 761
enquêtés et va de 1 à 9.*/
generate noi = ne
/*tab noi, missing*/
*CAT0
tab cat0, missing /*Cette variable va de 129 à 999 et est bien renseignée
pour les 21 761 individus*/
rename cat0 cat0_ind
*CAT1
/*tab cat1, missing*/ /*Cette variable va de 0 à 698 et est bien renseignée
pour les 21 761 individus*/
rename cat1 cat1_ind
*FNI2
summarize fni2 /*Cette variable va de 0 à 1 et est bien renseignée
pour les 21 761 individus*/
rename fni2 fni2_ind
*FNI_MOBILE
summarize fni_mobile /*Cette variable va de 0 à 1 et est bien renseignée
pour les 21 761 individus*/
rename fni_mobile fni_mobile_ind
save indiv2_CAT1, replace
```

```
*J'examine les 4 nomenclatures de l'INED issues de la table ENFANTS2_CAT1
use enfants2_CAT1
*IDENT (identifiant de l'enquêté, pour chaque enfant de l'enquêté)
summarize ident /*Cette variable est bien renseignée pour les 31 476 enfants et va de 3 à 21
760.*/
*NOI (numéro d'ordre de l'enfant de l'enquêté dans le logement, pour chaque enfant cohabitant
de l'enquêté)
sort ident e_noi_thl e_vie e_sexe e_anais e_nat e_lvmet e_situa e_diplo
tab e_noi_thl, missing /*Cette variable n'est pas renseignée pour les 7 924
enfants non cohabitants ou décédés, et elle va de 1 à 13*/
generate noi = e_noi_thl
tab noi, missing
generate noi_noncohab = 100 + _n /*Pour les 7 924 enfants non cohabitants ou décédés et qui
logiquement n'ont pas de NOI, on attribue un NOI arbitraire (égal à 100 + le numéro de leur
ligne dans la table ENFANTS2_PRENOM) afin de pouvoir les identifier uniquement*/
tostring noi_noncohab, replace
replace noi = noi_noncohab if (noi=="")
sort ident e_noi_thl e_anais e_sexe
*CAT0
tab cat0, missing /*Cette variable va de 129 à 999 et n'est pas renseignée
pour 34 individus*/
recode cat0 (.=698)
rename cat0 cat0_enf
```

```

*CAT1
/*tab cat1, missing*/
pour 34 individus*/
recode cat1 (.=698)
rename cat1 cat1_enf
*FNI2
summarize fni2
pour les 31 476 individus*/
rename fni2 fni2_enf
*FNI_MOBILE
summarize fni_mobile
pour 87 individus*/
rename fni_mobile fni_mobile_enf
save enfants2_CAT1, replace

*J'apparie les 2 nomenclatures de l'INED de la table thl2_CAT1 à la table REPRO
use repro
sort ident noi anais_ind sexe_ind prenom_ind
merge 1:1 ident noi using thl2_CAT1, keepusing(cat0_thl cat1_thl)
drop _merge

*J'apparie les 4 nomenclatures de l'INED de la table indiv2_CAT1 à la table REPRO
merge 1:1 ident noi using indiv2_CAT1, keepusing(cat0_ind cat1_ind fni2_ind fni_mobile_ind)
drop _merge

*J'apparie les 4 nomenclatures de l'INED de la table enfants2_CAT1 à la table REPRO
merge 1:1 ident noi using enfants2_CAT1, keepusing(cat0_enf cat1_enf fni2_enf fni_mobile_enf)
drop _merge

*Je vérifie que les nomenclatures de l'INED sont cohérentes entre elles pour chaque individu
(en lignes)
*CAT0
generate checkcat0_1 = 1 if (cat0_ind != cat0_thl & cat0_ind != .)
tab checkcat0_1, missing /*c'est bon*/
drop checkcat0_1
generate checkcat0_2 = 1 if (cat0_enf != cat0_thl & cat0_enf != . & cat0_thl != .)
tab checkcat0_2, missing /*c'est bon*/
drop checkcat0_2
*CAT1
generate checkcat1_1 = 1 if (cat1_ind != cat1_thl & cat1_ind != .)
tab checkcat1_1, missing /*c'est bon*/
drop checkcat1_1
generate checkcat1_2 = 1 if (cat1_enf != cat1_thl & cat1_enf != . & cat1_thl != .)
tab checkcat1_2, missing /*c'est bon*/
drop checkcat1_2
*FNI2
generate checkfni2 = 1 if (fni2_ind != fni2_enf & fni2_ind != . & fni2_enf != .)
tab checkfni2, missing /*c'est bon*/
drop checkfni2
*FNI_MOBILE
generate checkfni_mobile = 1 if (fni_mobile_ind != fni_mobile_enf & fni_mobile_ind != . &
fni_mobile_enf != .)
tab checkfni_mobile, missing /*c'est bon*/
drop checkfni_mobile

*Je crée les variables de nomenclature de l'INED

*Je crée la variable de nomenclature de l'INED CAT0_INED
generate cat0_ined = cat0_thl if (lienego_ind != 200 & lienego_ind != 201)
replace cat0_ined = cat0_enf if (lienego_ind == 200 | lienego_ind == 201)
tab cat0_ined, missing
*J'applique des labels à la variable bidon CAT0_INED
label variable cat0_ined "cat0 bidon"
label define cat0bidon 129 "France" 641 "Etats de langue arabe (hors Maghreb)" 695 "Langue
associée à un pays" 696 "Prénoms largement diffusés" 697 "Prénoms internationaux
(principalement anglo-américains)" 698 "Impossibles à classer" 999 "Erreur"
label values cat0_ined cat0bidon
tab cat0_ined, missing /*cette variable, communiquée par l'INED au réseau
Quetelet, au CASD et à la communauté scientifique, contient des modalités et des regroupements
bidon; c'est pourquoi je ne l'utiliserai pas*/

*Je crée la variable de nomenclature de l'INED CAT1_INED
generate cat1_ined = cat1_thl if (lienego_ind != 200 & lienego_ind != 201)
replace cat1_ined = cat1_enf if (lienego_ind == 200 | lienego_ind == 201)
/*tab cat1_ined, missing*/

*Je crée la variable de nomenclature de l'INED CAT1_INED_REC

```

```

*Je recode la nomenclature CAT1_INED comme indiqué dans le message de l'INED du 4/12/2020,
avec les corrections indiquées en vert ci-dessous:
generate catl_ined_rec = catl_ined
    recode catl_ined_rec (129 = 1/*1=Français*/) (137 153 161 323 327 661 = 2/*2=Latin*/) (135
168 368 696 697 = 3/*3=International ou anglophone*/) (100 121 122 125 126 128 132 134
/*prénoms grecs (134) regroupés avec prénoms d'autre Europe et non avec prénoms latins*/ 138
139 142 144 148 149 152 154 157 160 162 165 166 169 422 /*prénoms arméniens (422) regroupés
avec prénoms d'autre Europe et non avec prénoms arabo-musulmans*/ 430 441 611 651 = 4/*4=Autre
Europe, ou "Divers"*/) (163 /*turc*/ 221 252 425 439 450 622 641 643 662 = 5/*5=Arabo-
musulman*/) (224 /*prénoms érythréens et éthiopiens (amhariques) (224) regroupés avec prénoms
africains et non arabo-musulmans*/ 231/251 262/283 621 623 = 6/*6=Africain*/) (427 431/437 442
446 449 457/479 644 645 = 7/*7=Asiatique*/) (0 698 . = 9/*9=Inclassable*/)
*J'applique des labels à la variable CAT1_INED_REC
label variable catl_ined_rec "catl regroupée du prénom"
    label define catlregroupee 1 "Français" 2 "Latin" 3 "International ou anglophone" 4 "Autre
Europe" 5 "Arabo-musulman" 6 "Africain" 7 "Asiatique" 9 "Inclassable"
    label values catl_ined_rec catlregroupee
        tab catl_ined_rec [aw=poids_ind], missing

*Je vérifie que CAT1_INED_REC est cohérente pour tous les porteurs du même prénom (en
colonnes)
egen mode_catl_ined_rec = mode(catl_ined_rec), by(prenom_ind)
generate probleme_catl_ined_rec = 1 if (catl_ined_rec != mode_catl_ined_rec)
tab probleme_catl_ined_rec, missing /*25 individus voient associée à leur prénom une catégorie
d'origine qui n'est pas la catégorie d'origine modale de ce prénom, d'après CAT1*/
sort probleme_catl_ined_rec mode_catl_ined_rec
*I1 apparaît que, parmi ces 25 individus problématiques:
** 9 individus voient associée à leur prénom une catégorie d'origine qui n'est pas attendue
(vu sa catégorisation par CAT1), parce que apparemment l'INED a fait catégoriser ces prénoms
avant de les avoir nettoyés (espaces, underscores mal placés, etc.), d'où une probable erreur
dans leurs bases de données THL2_CAT1, INDIV2_CAT1 et ENFANTS2_CAT1;
**12 individus voient associée à leur prénom une catégorie d'origine qui n'est pas attendue
(vu sa catégorisation par CAT1), parce que l'appariement rigoureux est impossible entre les
enfants non cohabitants de la base ENFANTS2_CAT1 et ceux de la base REPRO (l'absence de
e_noi_thl ne peut pas toujours être compensé par le tri complet des tables), si bien que
certains enfants non cohabitants se sont vus attribuer la catégorie d'origine du prénom d'un
de leurs germains non cohabitants (et réciproquement);
** 4 individus voient associée à leur prénom la catégorie d'origine attendue, mais ce n'est
pas la catégorie modale parce qu'ils sont l'un des 2 seuls porteurs d'un prénom qui, chez
l'autre porteur, est mal catégorisé pour l'une des deux raisons ci-dessus.
**donc je corrige 17 de ces mauvaises catégorisations
replace catl_ined_rec = mode_catl_ined_rec if (probleme_catl_ined_rec==1 &
mode_catl_ined_rec>=1 & mode_catl_ined_rec<=9)
**et les 4 qui restent
egen min_catl_ined_rec = min(catl_ined_rec), by(prenom_ind)
replace catl_ined_rec = min_catl_ined_rec if (probleme_catl_ined_rec==1 &
mode_catl_ined_rec==.)
**les 4 dernières étant, comme on l'a vu, correctes

/*CAT1_INED_REC est (re)codée comme suit:
1=Français
2=Latin
3=International ou anglophone
4=autre Europe
5=Arabo-musulman
6=Africain
7=Asiatique
9=Inclassable*/

*Je crée la variable de nomenclature de l'INED FNI2
generate fni2_ined = fni2_ind if (lienego_ind == 0)
    replace fni2_ined = fni2_enf if (lienego_ind == 2 | lienego_ind == 200 | lienego_ind == 201)
        tab fni2_ined, missing /*29 808 individus sont non renseignés, et comme pour
CAT1_INED_REC cette variable contient 21 erreurs de catégorisation, que je ne corrige pas*/

*Je crée la variable de nomenclature de l'INED FNI_MOBILE
generate fni_mobile_ined = fni_mobile_ind if (lienego_ind == 0)
    replace fni_mobile_ined = fni_mobile_enf if (lienego_ind == 2 | lienego_ind == 200 |
lienego_ind == 201)
        /*tab fni_mobile_ined, missing*/ /*29 895 individus sont non renseignés, et comme pour
CAT1_INED_REC cette variable contient 21 erreurs de catégorisation, que je ne corrige pas*/

order ident noi poids_ind sexe_ind anais_ind prenom_ind orig_prenom_ind catl_ined_rec
lienego_ind regionnais_ind residence_ind group2_ind regionnaispere_ind regionnaismere_ind
orig_ind autrepere natpniv2 natmniv2 n_indper n_indmer lnp1_gr lnp2_gr lnm1_gr lnm2_gr
relig_ind r_impvie orig_relig_ind x_vufri
sort ident noi

```

```

save repro, replace

*Je supprime les 3 bases de données THL2_CAT1, INDIV2_CAT1 et ENFANTS2_CAT1, qui ont permis
d'importer les nomenclatures de l'INED
erase thl2_CAT1.dta
erase indiv2_CAT1.dta
erase enfants2_CAT1.dta

/*Chapitre 6: La nomenclature des prénoms par origine de l'INED: appariement au Fichier INSEE
des prénoms*/

use repro
duplicates report prenom_ind sexe_ind /*il y a 69 822 doublons, dont 64 486 sont en
surplus*/
duplicates drop prenom_ind sexe_ind, force /*ces 64 486 doublons en surplus sont éliminés*/
save repro_cat1.dta

use insee
merge m:1 prenom_ind sexe_ind using repro_cat1.dta, keepusing(cat1_ined_rec) keep(master
match)
drop _merge
groups prenom_ind [aw=nombre] if (cat1_ined_rec==.), order(h) select(20) /*les prénoms non
catégorisés les plus fréquents comprennent "Marceau", "Eveline", "Constant", etc.: la
nomenclature de l'INED s'est bien appariée*/
recode cat1_ined_rec (. = 9)
tab cat1_ined_rec [aw=nombre], missing /**/
save insee, replace

erase repro_cat1.dta

/*CAT1_INED_REC est (re)codée comme suit:
1=Français
2=Latin
3=International ou anglophone
4=autre Europe
5=Arabo-musulman
6=Africain
7=Asiatique
9=Inclassable*/

/*Chapitre 6: Comparaison entre la nomenclature des prénoms par origine de l'INED et la
nomenclature synthétique*/

*Tableau 6.7. Origine des prénoms en France, selon la nomenclature de l'INED et selon la
nomenclature synthétique
**1e colonne, N = 42 208
use repro
tab cat1_ined_rec if (bon_champ==1) [aw=poids_ind], missing
**2e colonne, N = 42 208
tab orig_prenom_ind if (bon_champ==1) [aw=poids_ind], missing
save repro, replace
**3e colonne, N = 84 264 960
use insee
tab cat1_ined_rec [aw=nombre] , missing
**4e colonne, N = 84 264 960
tab orig_prenom_ind [aw=nombre] , missing
save insee, replace

*Tableau 6.8. Origine des prénoms portés en France en 2008 (enquête TeO), N = 42 208
use repro
tab orig_prenom_ind cat1_ined_rec if (bon_champ==1) [aw=poids_ind], cell nofreq missing
save repro, replace
*Origine des prénoms attribués en France depuis 1900 (fichier INSEE des prénoms)
use insee
tab orig_prenom_ind cat1_ined_rec [aw=nombre] , cell nofreq missing
save insee, replace

*Tableau 6.9. Top 3 des prénoms "inclassables" selon l'origine de leur porteur en France en
2008 (enquête TeO)
use repro
**Nomenclature de l'INED, N = 989
tab cat1_ined_rec if (bon_champ==1) [aw=poids_ind], missing
/*groups prenom_ind if (bon_champ==1 & sexe_ind==1 & cat1_ined_rec==09) & (orig_ind>=500 &
orig_ind<=504) [aw=poids_ind], order(h) select(3)*/ /*les
habitants masculins du groupe "majoritaire" ont comme 3 prénoms "inclassables" les plus

```

[illegible]



```

ont comme 3 prénoms "inclassables" les plus fréquents s, s et s*/
/*groups prenom_ind if (bon_champ==1 & sexe_ind==2 & catl_ined_rec==09)
[aw=poids_ind], order(h) select(3)*/ /*les habitants féminins
ont comme 3 prénoms "inclassables" les plus fréquents s, s et s*/
**Nomenclature synthétique, N = 2 392
tab orig_prenom_ind if (bon_champ==1) [aw=poids_ind], missing
/*groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_prenom_ind==09) & (orig_ind>=500 &
orig_ind<=504) [aw=poids_ind], order(h) select(3)*/ /*les
habitants masculins du groupe "majoritaire" ont comme 3 prénoms "inclassables" les plus
fréquents Killian, s et Evan*/
/*groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_prenom_ind==09) & (orig_ind>=500 &
orig_ind<=504) [aw=poids_ind], order(h) select(3)*/ /*les
habitants féminins du groupe "majoritaire" ont comme 3 prénoms "inclassables" les plus
fréquents s, Lou et Mégane*/
/*groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_prenom_ind==09) & (orig_ind==212 |
orig_ind==412 | orig_ind==712) [aw=poids_ind], order(h) select(3)*/ /*les
habitants masculins originaires d'un DOM ont comme 3 prénoms "inclassables" les plus
fréquents s, s et s*/
/*groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_prenom_ind==09) & (orig_ind==212 |
orig_ind==412 | orig_ind==712) [aw=poids_ind], order(h) select(3)*/ /*les
habitants féminins originaires d'un DOM ont comme 3 prénoms "inclassables" les plus
fréquents s, s et s*/
/*groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_prenom_ind==09) & (orig_ind==140 |
orig_ind==340 | orig_ind==640 | orig_ind==840) [aw=poids_ind], order(h) select(3)*/ /*les
habitants masculins originaires d'Europe du Sud ont comme 3 prénoms "inclassables" les plus
fréquents s, Tiago et s*/
/*groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_prenom_ind==09) & (orig_ind==140 |
orig_ind==340 | orig_ind==640 | orig_ind==840) [aw=poids_ind], order(h) select(3)*/ /*les
habitants féminins originaires d'Europe du Sud ont comme 3 prénoms "inclassables" les plus
fréquents s, Cassandra et s*/
/*groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_prenom_ind==09) & (orig_ind==149 |
orig_ind==349 | orig_ind==649 | orig_ind==849) [aw=poids_ind], order(h) select(3)*/ /*les
habitants masculins originaires d'autre Europe ont comme 3 prénoms "inclassables" les plus
fréquents s, s et s*/
/*groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_prenom_ind==09) & (orig_ind==149 |
orig_ind==349 | orig_ind==649 | orig_ind==849) [aw=poids_ind], order(h) select(3)*/ /*les
habitants féminins originaires d'autre Europe ont comme 3 prénoms "inclassables" les plus
fréquents s, s et s*/
/*groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_prenom_ind==09) & (orig_ind==120 |
orig_ind==320 | orig_ind==620 | orig_ind==820) [aw=poids_ind], order(h) select(3)*/ /*les
habitants masculins originaires du Maghreb ont comme 3 prénoms "inclassables" les plus
fréquents Ilan, s et s*/
/*groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_prenom_ind==09) & (orig_ind==120 |
orig_ind==320 | orig_ind==620 | orig_ind==820) [aw=poids_ind], order(h) select(3)*/ /*les
habitants féminins originaires du Maghreb ont comme 3 prénoms "inclassables" les plus
fréquents Alyssa, s et s*/
/*groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_prenom_ind==09) & (orig_ind==129 |
orig_ind==329 | orig_ind==629) [aw=poids_ind], order(h) select(3)*/ /*les
habitants masculins originaires d'autre Afrique ont comme 3 prénoms "inclassables" les plus
fréquents s, s et s*/
/*groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_prenom_ind==09) & (orig_ind==129 |
orig_ind==329 | orig_ind==629) [aw=poids_ind], order(h) select(3)*/ /*les
habitants féminins originaires d'autre Afrique ont comme 3 prénoms "inclassables" les plus
fréquents s, Assa et s*/
/*groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_prenom_ind==09) & (orig_ind==134 |
orig_ind==334 | orig_ind==634) [aw=poids_ind], order(h) select(3)*/ /*les
habitants masculins originaires de Turquie ont comme 3 prénoms "inclassables" les plus
fréquents s, s et s*/
/*groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_prenom_ind==09) & (orig_ind==134 |
orig_ind==334 | orig_ind==634) [aw=poids_ind], order(h) select(3)*/ /*les
habitants féminins originaires de Turquie ont comme 3 prénoms "inclassables" les plus
fréquents Beyza, Dilek et Merve*/
/*groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_prenom_ind==09) & (orig_ind==130 |
orig_ind==330 | orig_ind==630) [aw=poids_ind], order(h) select(3)*/ /*les
habitants masculins originaires d'Asie ont comme 3 prénoms "inclassables" les plus
fréquents s, s et s*/
/*groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_prenom_ind==09) & (orig_ind==130 |
orig_ind==330 | orig_ind==630) [aw=poids_ind], order(h) select(3)*/ /*les
habitants féminins originaires d'Asie ont comme 3 prénoms "inclassables" les plus
fréquents s, s et s*/
/*groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_prenom_ind==09) & (orig_ind==150 |
orig_ind==350 | orig_ind==650) [aw=poids_ind], order(h) select(3)*/ /*les
habitants masculins originaires d'ailleurs ont comme 3 prénoms "inclassables" les plus
fréquents s, s et s*/
/*groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_prenom_ind==09) & (orig_ind==150 |
orig_ind==350 | orig_ind==650) [aw=poids_ind], order(h) select(3)*/ /*les
habitants féminins originaires d'ailleurs ont comme 3 prénoms "inclassables" les plus

```

```

fréquents s, s et s*/
/*groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_prenom_ind==09) & (orig_ind==.)
[aw=poids_ind], order(h) select(3)*/ /*les habitants masculins originaires de [non réponse]
ont comme 3 prénoms "inclassables" les plus fréquents -, - et -*/
/*groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_prenom_ind==09) & (orig_ind==.)
[aw=poids_ind], order(h) select(3)*/ /*les habitants féminins originaires de [non réponse]
ont comme 3 prénoms "inclassables" les plus fréquents -, - et -*/
/*groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_prenom_ind==09)
[aw=poids_ind], order(h) select(3)*/ /*les habitants masculins
ont comme 3 prénoms "inclassables" les plus fréquents Killian, s et Evan*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_prenom_ind==09)
[aw=poids_ind], order(h) select(3) /*les habitants féminins
ont comme 3 prénoms "inclassables" les plus fréquents Fiona, Lou et Mégane*/
save repro, replace

/*Chapitre 6: Analyses préliminaires sur les origines des prénoms en France*/

*Quels sont les prénoms que seule la nomenclature synthétique (mais pas la nomenclature de
l'INED) catégorise comme "latins"?
**enquête TeO
use repro
groups prenom_ind if (bon_champ==1 & orig_prenom_ind==2 & cat1_ined_rec!=2) [aw=poids_ind],
order(h) select(5) /*Patricia, Laura, Clara, Hugo et Sandra*/
**fichier INSEE des prénoms
use insee
groups prenom_ind if (orig_prenom_ind==2 & cat1_ined_rec!=2) [aw=nombre] ,
order(h) select(5) /*Patricia, Laura, Hugo, Maria et Anna*/

*Quels sont les prénoms que seule la nomenclature de l'INED (mais pas la nomenclature
synthétique) catégorise comme "internationaux ou anglophones"?
**enquête TeO
use repro
groups prenom_ind if (bon_champ==1 & orig_prenom_ind!=3 & cat1_ined_rec==3) [aw=poids_ind],
order(h) select(5) /*Patrick, Patricia, Anthony, Laura et Sarah*/
**fichier INSEE des prénoms
use insee
groups prenom_ind if (orig_prenom_ind!=3 & cat1_ined_rec==3) [aw=nombre] ,
order(h) select(5) /*Patrick, Patricia, Léa, Anthony et Sarah*/

*Quels sont les prénoms qu'1 seule nomenclature catégorise comme "arabo-musulmans"?
**enquête TeO
use repro
groups prenom_ind if (bon_champ==1 & orig_prenom_ind==5 & cat1_ined_rec!=5) [aw=poids_ind],
order(h) select(5) /*Nadia, Inès, Linda, Yannis et Sara*/
groups prenom_ind if (bon_champ==1 & orig_prenom_ind!=5 & cat1_ined_rec==5) [aw=poids_ind],
order(h) select(5) /*Lilou, Monia, Melyssa, Gurvan et Lilia*/
**fichier INSEE des prénoms
use insee
groups prenom_ind if (orig_prenom_ind==5 & cat1_ined_rec!=5) [aw=nombre] ,
order(h) select(5) /*Inès, Nadia, Lina, Nolan et Linda*/
groups prenom_ind if (orig_prenom_ind!=5 & cat1_ined_rec==5) [aw=nombre] ,
order(h) select(5) /*Lilou, Lila, Lyna, Louka et Lilia*/
save insee, replace

*Figure 6.10a. Répartition par origine des prénoms portés en France en 2008, par décennie de
naissance (nomenclature de l'INED), N = 32 041
use repro
tab decennais_ind cat1_ined_rec if bon_champ==1 & decennais_ind>=1950 & (regionnais_ind==11
| regionnais_ind==12) [aw=poids_ind], row nofreq missing

*Figure 6.10b. Répartition par origine des prénoms portés en France en 2008, par décennie de
naissance (nomenclature synthétique), N = 32 041
tab decennais_ind orig_prenom_ind if bon_champ==1 & decennais_ind>=1950 & (regionnais_ind==11
| regionnais_ind==12) [aw=poids_ind], row nofreq missing
save repro, replace

*Figure 6.11a. Répartition par origine des prénoms attribués en France depuis 1900, par année
de naissance (nomenclature de l'INED), N = 84 264 960
use insee
tab anais_ind cat1_ined_rec [aw=nombre], row nofreq missing
*Prénoms "inclassables" les plus fréquemment attribués en France à l'état-civil depuis les
années 2000
groups prenom_ind if sexe_ind==1 & anais_ind>=2000 & cat1_ined_rec==9 [aw=nombre], order(h)
select(6) /*prénoms masculins: Marceau, Milo, Lyam, Soan, Sasha*/
groups prenom_ind if sexe_ind==2 & anais_ind>=2000 & cat1_ined_rec==9 [aw=nombre], order(h)
select(6) /*prénoms féminins: Elea, Lya, Anae, Lylou et Sasha*/

```

```

*Figure 6.11b. Répartition par origine des prénoms attribués en France depuis 1900, par année
de naissance (nomenclature synthétique), N = 84 264 960
tab anais_ind orig_prenom_ind [aw=nombre], row nofreq missing
save insee, replace

```

```

*Figure 6.12a. Proportion de prénoms "arabo-musulmans" portés en France en 2008, par décennie
de naissance
/*Figure tirée de la Figure 6.10a-6.10b*/

```

```

*Figure 6.12b. Proportion de prénoms "arabo-musulmans" attribués en France depuis 1900, par
année de naissance
/*Figure tirée de la Figure 6.11a-6.11b*/

```

```

/*Chapitre 6: Origine du prénom et origine de son porteur en France: une analyse*/

```

```

*Tableau 6.13a. Origine du prénom (colonnes) selon l'origine géographique de son porteur
(lignes), en France en 2008 (enquête TeO), N = 42 208

```

```

use repro
*Je crée la variable orig_ind synthétique
generate orig_ind_synthetique = 1 if (orig_ind==500 | orig_ind==501 | orig_ind==502 |
orig_ind==503 | orig_ind==504) /*groupe "majoritaire"*/
replace orig_ind_synthetique = 2 if (orig_ind==212 | orig_ind==412 | orig_ind==712)
/*DOM*/
replace orig_ind_synthetique = 3 if (orig_ind==140 | orig_ind==340 | orig_ind==640 |
orig_ind==840) /*Europe du Sud*/
replace orig_ind_synthetique = 4 if (orig_ind==149 | orig_ind==349 | orig_ind==649 |
orig_ind==849) /*autre Europe*/
replace orig_ind_synthetique = 5 if (orig_ind==120 | orig_ind==320 | orig_ind==620 |
orig_ind==820) /*Maghreb*/
replace orig_ind_synthetique = 6 if (orig_ind==129 | orig_ind==329 | orig_ind==629)
/*autre Afrique*/
replace orig_ind_synthetique = 7 if (orig_ind==134 | orig_ind==334 | orig_ind==634)
/*Turquie*/
replace orig_ind_synthetique = 8 if (orig_ind==130 | orig_ind==330 | orig_ind==630)
/*Asie*/
replace orig_ind_synthetique = 9 if (orig_ind==150 | orig_ind==350 | orig_ind==650)
/*Ailleurs dans le monde*/
*J'applique des labels à la variable ORIG_IND_SYNTHETIQUE
label variable orig_ind_synthetique "origine synthétique"
label define orig_synth 1 "Groupe majoritaire" 2 "DOM" 3 "Europe Sud" 4 "autre Europe" 5
"Maghreb" 6 "autre Afrique" 7 "Turquie" 8 "Asie" 9 "ailleurs"
label values orig_ind_synthetique orig_synth
tab orig_ind_synthetique, missing

```

```

*Je crée le tableau 6.13a
tab orig_ind_synthetique if bon_champ==1 [aw=poids_ind], missing
tab orig_ind_synthetique cat1_ined_rec if bon_champ==1 [aw=poids_ind], cell nofreq missing
tab orig_ind_synthetique orig_prenom_ind if bon_champ==1 [aw=poids_ind], cell nofreq missing
/*non publié, avec nomenclature synthétique*/

```

```

*Tableau 6.13b. Origine géographique de l'individu (lignes) selon l'origine de son prénom
(colonnes), en France en 2008 (enquête TeO), N = 42 208

```

```

tab cat1_ined_rec if bon_champ==1 [aw=poids_ind], missing
tab orig_ind_synthetique cat1_ined_rec if bon_champ==1 [aw=poids_ind], cell nofreq missing

```

```

*Tableau 6.14a. Top 5 des prénoms selon l'origine géographique de leur porteur, en France en
2008 (enquête TeO), N = 42 208

```

```

tab orig_ind_synthetique if (bon_champ==1) [aw=poids_ind], missing
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind_synthetique==1) [aw=poids_ind],
order(h) select(5) /*les habitants masculins qui sont de la population "majoritaire"
ont comme 5 prénoms les plus fréquents Sébastien, Patrick, Michel, Eric et Philippe*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind_synthetique==1) [aw=poids_ind],
order(h) select(5) /*les habitants féminins qui sont de la population "majoritaire"
ont comme 5 prénoms les plus fréquents Catherine, Martine, Nathalie, Isabelle et Françoise*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind_synthetique==2) [aw=poids_ind],
order(h) select(5) /*les habitants masculins qui sont originaires de DOM (G1, G2 ou G3)
ont comme 5 prénoms les plus fréquents David, Patrick, Olivier, Thierry et Alain*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind_synthetique==2) [aw=poids_ind],
order(h) select(5) /*les habitants féminins qui sont originaires de DOM (G1, G2 ou G3)
ont comme 5 prénoms les plus fréquents Nathalie, Stéphanie, Véronique, Aurélie et Karine*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind_synthetique==3) [aw=poids_ind],
order(h) select(5) /*les habitants masculins qui sont originaires d'Europe du Sud (G1, G2,
G3 ou G4) ont comme 5 prénoms les plus fréquents José, Daniel, Manuel, David et Antonio*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind_synthetique==3) [aw=poids_ind],
order(h) select(5) /*les habitants féminins qui sont originaires d'Europe du Sud (G1, G2,

```

```

G3 ou G4) ont comme 5 prénoms les plus fréquents Maria, Isabelle, Marie-Carmen, Céline et Elisabeth*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind_synthetique==4) [aw=poids_ind],
order(h) select(5) /*les habitants masculins qui sont originaires d'autre Europe (G1, G2, G3 ou G4) ont comme 5 prénoms les plus fréquents Jacques, Jean-Claude, Michel, Pierre et Nicolas*/
/*groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind_synthetique==4) [aw=poids_ind],
order(h) select(5)*/ /*les habitants féminins qui sont originaires d'autre Europe (G1, G2, G3 ou G4) ont comme 5 prénoms les plus fréquents Catherine, Anne, s, Nadège et Sandra*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind_synthetique==5) [aw=poids_ind],
order(h) select(5) /*les habitants masculins qui sont originaires du Maghreb (G1, G2, G3 ou G4) ont comme 5 prénoms les plus fréquents Mohamed, Karim, Ahmed, Rachid et Farid*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind_synthetique==5) [aw=poids_ind],
order(h) select(5) /*les habitants féminins qui sont originaires du Maghreb (G1, G2, G3 ou G4) ont comme 5 prénoms les plus fréquents Fatima, Sarah, Nadia, Myriam et Samira*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind_synthetique==6) [aw=poids_ind],
order(h) select(5) /*les habitants masculins qui sont originaires d'autre Afrique (G1, G2 ou G3) ont comme 5 prénoms les plus fréquents Marcellin, Jordan, Mamadou, Mohamed et Kevin*/
/*groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind_synthetique==6) [aw=poids_ind],
order(h) select(5)*/ /*les habitants féminins qui sont originaires d'autre Afrique (G1, G2 ou G3) ont comme 5 prénoms les plus fréquents Aurélie, s, Aminata, Mariam et Fatoumata*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind_synthetique==7) [aw=poids_ind],
order(h) select(5) /*les habitants masculins qui sont originaires de Turquie (G1, G2 ou G3) ont comme 5 prénoms les plus fréquents Mehmet, Ali, Mustafa, Ibrahim et Murat*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind_synthetique==7) [aw=poids_ind],
order(h) select(5) /*les habitants féminins qui sont originaires de Turquie (G1, G2 ou G3) ont comme 5 prénoms les plus fréquents Murielle, Fatma, Ayse, Hatice et Emine*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind_synthetique==8) [aw=poids_ind],
order(h) select(5) /*les habitants masculins qui sont originaires d'Asie (G1, G2 ou G3) ont comme 5 prénoms les plus fréquents David, Nicolas, Vincent, Michel et Kevin*/
/*groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind_synthetique==8) [aw=poids_ind],
order(h) select(5)*/ /*les habitants féminins qui sont originaires d'Asie (G1, G2 ou G3) ont comme 5 prénoms les plus fréquents Marie, Sophie, s, Nina et Zara*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind_synthetique==9) [aw=poids_ind],
order(h) select(5) /*les habitants masculins qui sont originaires d'ailleurs (G1, G2 ou G3) ont comme 5 prénoms les plus fréquents Thomas, Mario, Paul, Georges et Jacques*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind_synthetique==9) [aw=poids_ind],
order(h) select(5) /*les habitants féminins qui sont originaires d'ailleurs (G1, G2 ou G3) ont comme 5 prénoms les plus fréquents Paulette, Emilie, Delphine, Sandra et Sylvie*/
/*groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind_synthetique==.) [aw=poids_ind],
order(h) select(5)*/ /*les habitants masculins dont on ignore les origines ont comme 5 prénoms les plus fréquents -, -, -, - et -*/
/*groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind_synthetique==.) [aw=poids_ind],
order(h) select(5)*/ /*les habitants féminins dont on ignore les origines ont comme 5 prénoms les plus fréquents -, -, -, - et -*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1) [aw=poids_ind],
order(h) select(5) /*les habitants masculins ont comme 5 prénoms les plus fréquents Sébastien, Patrick, Michel, Nicolas et Stéphane*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2) [aw=poids_ind],
order(h) select(5) /*les habitants féminins ont comme 5 prénoms les plus fréquents Catherine, Nathalie, Martine, Isabelle et Françoise*/

*Tableau 6.14b. Top 5 des prénoms selon leur origine, en France en 2008 (enquête TeO), N = 42 208
tab cat1_ined_rec if (bon_champ==1) [aw=poids_ind], missing
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & cat1_ined_rec==01) [aw=poids_ind],
order(h) select(5) /*les habitants masculins qui portent un prénom "français" ont comme 5 prénoms les plus fréquents Sébastien, Michel, Nicolas, Stéphane et Philippe*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & cat1_ined_rec==01) [aw=poids_ind],
order(h) select(5) /*les habitants féminins qui portent un prénom "français" ont comme 5 prénoms les plus fréquents Catherine, Nathalie, Martine, Isabelle et Françoise*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & cat1_ined_rec==02) [aw=poids_ind],
order(h) select(5) /*les habitants masculins qui portent un prénom "latin" ont comme 5 prénoms les plus fréquents Enzo, José, Manuel, Antonio et Teo*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & cat1_ined_rec==02) [aw=poids_ind],
order(h) select(5) /*les habitants féminins qui portent un prénom "latin" ont comme 5 prénoms les plus fréquents Sabrina, Inès, Lola, Luna et Amelia*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & cat1_ined_rec==03) [aw=poids_ind],
order(h) select(5) /*les habitants masculins qui portent un prénom "international ou anglophone" ont comme 5 prénoms les plus fréquents Patrick, Anthony, Kevin, Cédric et Jérémy*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & cat1_ined_rec==03) [aw=poids_ind],
order(h) select(5) /*les habitants féminins qui portent un prénom "international ou anglophone" ont comme 5 prénoms les plus fréquents Patricia, Audrey, Laura, Sarah et Vanessa*/
/*groups prenom_ind if (bon_champ==1 & sexe_ind==1 & cat1_ined_rec==04) [aw=poids_ind],
order(h) select(5)*/ /*les habitants masculins qui portent un prénom "d'autre Europe" ont comme 5 prénoms les plus fréquents Dimitri, Axel, s, Ivan et Kilian*/

```

```

/*groups prenom_ind if (bon_champ==1 & sexe_ind==2 & cat1_ined_rec==04) [aw=poids_ind],
order(h) select(5)*/ /*les habitants féminins qui portent un prénom "d'autre Europe"
ont comme 5 prénoms les plus fréquents Ingrid, Johanna, Tatiana, s et Karin*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & cat1_ined_rec==05) [aw=poids_ind],
order(h) select(5) /*les habitants masculins qui portent un prénom "arabo-musulman"
ont comme 5 prénoms les plus fréquents Mohamed, Karim, Yanis, Ahmed et Rachid*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & cat1_ined_rec==05) [aw=poids_ind],
order(h) select(5) /*les habitants féminins qui portent un prénom "arabo-musulman"
ont comme 5 prénoms les plus fréquents Fatima, Malika, Khadija, Samira et Leila*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & cat1_ined_rec==06) [aw=poids_ind],
order(h) select(5) /*les habitants masculins qui portent un prénom "africain"
ont comme 5 prénoms les plus fréquents Mamadou, Abdoulaye, Ibrahima, Amadou et Gloire*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & cat1_ined_rec==06) [aw=poids_ind],
order(h) select(5) /*les habitants féminins qui portent un prénom "africain"
ont comme 5 prénoms les plus fréquents Aminata, Fatoumata, Awa, Aya et Maina*/
/*groups prenom_ind if (bon_champ==1 & sexe_ind==1 & cat1_ined_rec==07) [aw=poids_ind],
order(h) select(5)*/ /*les habitants masculins qui portent un prénom "asiatique"
ont comme 5 prénoms les plus fréquents s, Loan, s, s et s*/
/*groups prenom_ind if (bon_champ==1 & sexe_ind==2 & cat1_ined_rec==07) [aw=poids_ind],
order(h) select(5)*/ /*les habitants féminins qui portent un prénom "asiatique"
ont comme 5 prénoms les plus fréquents Louane, s, s, s et s*/
/*groups prenom_ind if (bon_champ==1 & sexe_ind==1 & cat1_ined_rec==09) [aw=poids_ind],
order(h) select(5)*/ /*les habitants masculins qui portent un prénom "inclassable"
ont comme 5 prénoms les plus fréquents s, s, s, s et s*/
/*groups prenom_ind if (bon_champ==1 & sexe_ind==2 & cat1_ined_rec==09) [aw=poids_ind],
order(h) select(5)*/ /*les habitants féminins qui portent un prénom "inclassable"
ont comme 5 prénoms les plus fréquents s, s, s, s et s*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 ) [aw=poids_ind],
order(h) select(5) /*les habitants masculins
ont comme 5 prénoms les plus fréquents Sébastien, Patrick, Michel, Nicolas et Stéphane*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 ) [aw=poids_ind],
order(h) select(5) /*les habitants féminins
ont comme 5 prénoms les plus fréquents Catherine, Nathalie, Martine, Isabelle et Françoise*/

```

\*Tableau 6.15a. Origine du prénom (colonnes) selon l'origine religieuse de son porteur (lignes), en France en 2008 (enquête TeO), N = 42 208

```

tab orig_relig_ind if bon_champ==1 [aw=poids_ind], missing
tab orig_relig_ind orig_prenom_ind if bon_champ==1 [aw=poids_ind], cell nofreq missing
/*ORIG_RELIG_IND est (re)codée comme suit:
10=origine chrétienne
20=origine musulmane
30=origine juive
40=origine bouddhiste, hindouiste, autre, multi-religions
100=sans origine religieuse
. =Non réponse (structurelle), Refus de répondre, NSP*/

```

\*Tableau 6.15b. Origine religieuse de l'individu (lignes) selon l'origine de son prénom (colonnes), en France en 2008 (enquête TeO), N = 42 208

```

tab orig_prenom_ind if bon_champ==1 [aw=poids_ind], missing
tab orig_relig_ind orig_prenom_ind if bon_champ==1 [aw=poids_ind], cell nofreq missing

```

\*Prénoms les plus portés par les individus d'origine juive

```

groups prenom_ind if bon_champ==1 & sexe_ind==1 & orig_relig_ind==30 [aw=poids_ind], order(h)
select(5) /*Alain, David, Samuel, Raphaël et Sébastien*/
groups prenom_ind if bon_champ==1 & sexe_ind==2 & orig_relig_ind==30 [aw=poids_ind], order(h)
select(5) /*Catherine, Monique, Johanna, Véronique et Florence*/

```

\*Prénoms "juifs" les plus portés par les individus

```

groups prenom_ind if bon_champ==1 & sexe_ind==1 & orig_prenom_ind==8 [aw=poids_ind], order(h)
select(5) /*Samuel, Aaron, Elie, Joachim et Nathanaël*/
groups prenom_ind if bon_champ==1 & sexe_ind==2 & orig_prenom_ind==8 [aw=poids_ind], order(h)
select(5) /*Sarah, Myriam, Judith, Yaël et Yaëlle*/
save repro, replace

```

/\*Chapitre 6: Origine du prénom et origine de son porteur en France: une synthèse\*/

\*Figure 6.18. Prévalence des prénoms dans leur groupe d'origine et hors de leur groupe d'origine, en France en 2008 (enquête TeO)

/\*Figure tirée des Tableaux 6.13a et 6.15a\*/

/\*Chapitre 6: Résumé\*/

```

*****
/*CHAPITRE 7. Essai de reproduction du premier résultat spectaculaire:
   "Nicolas"*/
*****

/*Chapitre 7: Analyses préliminaires sur le prénom "Nicolas"*/
order ident noi poids_ind sexe_ind anais_ind prenom_ind orig_prenom_ind cat1_ined_rec
lienego_ind regionnais_ind residence_ind group2_ind regionnaispere_ind regionnaismere_ind
orig_ind autreparent natpniv2 natmniv2 n_indper n_indmer lnpl_gr lnp2_gr lnml_gr lnm2_gr
relig_ind r_impvie orig_relig_ind x_vufri
sort ident noi

*Figure 7.1a. Part des habitants masculins en 2008 qui se prénomment "Nicolas", par décennie
de naissance, N = 16 006
use repro
generate nicolas = 1 if (sexe_ind==1 & prenom_ind=="NICOLAS")
replace nicolas = 0 if (sexe_ind==1 & prenom_ind!="NICOLAS")
tab nicolas , missing /*1'enquête TeO compte 377 hommes prénommés
"Nicolas", contre 40 947 hommes prénommés autrement (données non pondérées)*/
tab nicolas [aw=poids_ind], missing /*1'enquête TeO compte 617 hommes prénommés
"Nicolas", contre 40 715 hommes prénommés autrement (données pondérées)*/
tab nicolas if bon_champ==1 & decennais_ind>=1950 & (regionnais_ind==11 |
regionnais_ind==12) [aw=poids_ind], missing
tab decennais_ind nicolas if bon_champ==1 & decennais_ind>=1950 & (regionnais_ind==11 |
regionnais_ind==12) [aw=poids_ind], row nofreq missing
save repro, replace

*Figure 7.1b.Part des nouveau-nés masculins qui se voient attribuer le prénom "Nicolas", par
décennie de naissance, N = 42 887 293
use insee
generate nicolas = 1 if (sexe_ind==1 & prenom_ind=="NICOLAS")
replace nicolas = 0 if (sexe_ind==1 & prenom_ind!="NICOLAS")
tab nicolas [aw=nombre], missing
tab decennais_ind nicolas [aw=nombre], row nofreq
save insee, replace

*Je supprime le Fichier INSEE des prénoms
erase insee.dta
*Dans le répertoire du CASD, ne restent que le présent programme (PROG_REPRO) et la table
REPRO

/*Chapitre 7: Les prénoms portés en France en 2008, par origine et par "génération" de leur
porteur*/

*Tableau 7.2. Top 20 des prénoms des habitants du groupe "majoritaire" selon qu'ils sont
enquêtés ou enfants d'enquêté, en France en 2008 (enquête TeO), N = 31 775
use repro
tab orig_ind if bon_champ==1 [aw=poids_ind], missing

groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==502) [aw=poids_ind], order(h)
select(20) /*Top 20 pondéré des prénoms des H enquêtés groupe "majoritaire"*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==502) [aw=poids_ind], order(h)
select(20) /*Top 20 pondéré des prénoms des F enquêtées groupe "majoritaire"*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==503) [aw=poids_ind], order(h)
select(20) /*Top 20 pondéré des prénoms des H enfants d'enquêté groupe "majoritaire"*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==503) [aw=poids_ind], order(h)
select(20) /*Top 20 pondéré des prénoms des F enfants d'enquêté groupe "majoritaire"*/

groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==502) , order(h)
select(95) /*Top non pondéré des prénoms des H enquêtés groupe "majoritaire"*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==502) , order(h)
select(107) /*Top non pondéré des prénoms des F enquêtées groupe "majoritaire"*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==503) , order(h)
select(72) /*Top non pondéré des prénoms des H enfants d'enquêté groupe "majoritaire"*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==503) , order(h)
select(67) /*Top non pondéré des prénoms des F enfants d'enquêté groupe "majoritaire"*/

tab prenom_ind cat1_ined_rec if (prenom_ind=="SEBASTIEN" | prenom_ind=="PATRICK" |
prenom_ind=="MICHEL" | prenom_ind=="PHILIPPE" | prenom_ind=="ERIC" |
prenom_ind=="STEPHANE" | prenom_ind=="THIERRY" | prenom_ind=="FREDERIC" | prenom_ind=="ALAIN"
| prenom_ind=="LAURENT" | prenom_ind=="DIDIER" | prenom_ind=="BRUNO" |
prenom_ind=="CHRISTOPHE" | prenom_ind=="NICOLAS" | prenom_ind=="CHRISTIAN" |
prenom_ind=="DAVID" | prenom_ind=="DOMINIQUE" | prenom_ind=="DANIEL" |
prenom_ind=="PASCAL" | prenom_ind=="JULIEN") /*d'après la nomenclature de l'INED, 1
prénom du Top 20 des prénoms des H enquêtés groupe "majoritaire" est d'une origine

```

```

autre que "français": "PATRICK" ("international ou anglophone")*/
tab prenom_ind catl_ined_rec if (prenom_ind=="CATHERINE" | prenom_ind=="MARTINE" |
prenom_ind=="NATHALIE" | prenom_ind=="ISABELLE" | prenom_ind=="FRANCOISE" |
prenom_ind=="SANDRINE" | prenom_ind=="SYLVIE" | prenom_ind=="VALERIE" |
prenom_ind=="PATRICIA" | prenom_ind=="LAURENCE" | prenom_ind=="CHRISTINE" |
prenom_ind=="VERONIQUE" | prenom_ind=="MICHELE" | prenom_ind=="ELODIE" |
prenom_ind=="CELINE" | prenom_ind=="SOPHIE" | prenom_ind=="CORINNE" |
prenom_ind=="ELISABETH" | prenom_ind=="DOMINIQUE" | prenom_ind=="DELPHINE") /*d'après la
nomenclature de l'INED, 1 prénom du Top 20 des prénoms des F enquêtées groupe
"majoritaire" est d'une origine autre que "français": "PATRICIA" ("international ou
anglophone")*/
tab prenom_ind catl_ined_rec if (prenom_ind=="MAXIME" | prenom_ind=="THOMAS" |
prenom_ind=="JULIEN" | prenom_ind=="NICOLAS" | prenom_ind=="LUCAS" |
prenom_ind=="ALEXIS" | prenom_ind=="PIERRE" | prenom_ind=="CLEMENT" | prenom_ind=="TOM"
| prenom_ind=="ADRIEN" | prenom_ind=="FLORIAN" | prenom_ind=="ANTOINE" |
prenom_ind=="ROMAIN" | prenom_ind=="ARTHUR" | prenom_ind=="GUILLAUME" |
prenom_ind=="QUENTIN" | prenom_ind=="ANTHONY" | prenom_ind=="LEO" | prenom_ind=="THEO"
| prenom_ind=="VINCENT") /*d'après la nomenclature de l'INED, 4 prénoms du Top 20 des prénoms
des H enfants d'enquête groupe "majoritaire" sont d'une origine autre que "français": "TOM",
"ANTHONY" "LEO" et "THEO" ("international ou anglophone")*/
tab prenom_ind catl_ined_rec if (prenom_ind=="MANON" | prenom_ind=="PAULINE" |
prenom_ind=="ANAIIS" | prenom_ind=="MARIE" | prenom_ind=="OCEANE" |
prenom_ind=="CAMILLE" | prenom_ind=="MARINE" | prenom_ind=="LEA" | prenom_ind=="LAURA"
| prenom_ind=="MATHILDE" | prenom_ind=="JUSTINE" | prenom_ind=="JULIE" |
prenom_ind=="CLARA" | prenom_ind=="CHLOE" | prenom_ind=="EMMA" |
prenom_ind=="ELODIE" | prenom_ind=="MARION" | prenom_ind=="CLAIRE" |
prenom_ind=="LUCIE" | prenom_ind=="CLEMENCE") /*d'après la nomenclature de l'INED, 2
prénoms du Top 20 des prénoms des F enfants d'enquête groupe "majoritaire" sont d'une origine
autre que "français": "LEA", "LAURA", "CLARA" et "EMMA" ("international ou anglophone")*/

```

\*Tableau 7.3. Top 20 des prénoms des habitants originaires d'Europe du Sud selon leur "génération", en France en 2008 (enquête TeO), N = 2 704

```

groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==140) [aw=poids_ind], order(h)
select(20) /*Top 20 pondéré des prénoms des H G1 Europe du Sud*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==140) [aw=poids_ind], order(h)
select(20) /*Top 20 pondéré des prénoms des F G1 Europe du Sud*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==340) [aw=poids_ind], order(h)
select(20) /*Top 20 pondéré des prénoms des H G2 Europe du Sud*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==340) [aw=poids_ind], order(h)
select(20) /*Top 20 pondéré des prénoms des F G2 Europe du Sud*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==640) [aw=poids_ind], order(h)
select(20) /*Top 20 pondéré des prénoms des H G3 Europe du Sud*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==640) [aw=poids_ind], order(h)
select(20) /*Top 20 pondéré des prénoms des F G3 Europe du Sud*/

```

```

groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==140) , order(h)
select(21) /*Top non pondéré des prénoms des H G1 Europe du Sud*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==140) , order(h)
select(22) /*Top non pondéré des prénoms des F G1 Europe du Sud*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==340) , order(h)
select(101) /*Top non pondéré des prénoms des H G2 Europe du Sud*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==340) , order(h)
select(101) /*Top non pondéré des prénoms des F G2 Europe du Sud*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==640) , order(h)
select(79) /*Top non pondéré des prénoms des H G3 Europe du Sud*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==640) , order(h)
select(78) /*Top non pondéré des prénoms des F G3 Europe du Sud*/

```

```

tab prenom_ind catl_ined_rec if (prenom_ind=="JOSE" | prenom_ind=="ANTONIO" |
prenom_ind=="MANUEL" | prenom_ind=="CARLOS" | prenom_ind=="ANTOINE" | prenom_ind=="MARIO"
| prenom_ind=="FERNANDO" | prenom_ind=="ALBERTO" | prenom_ind=="JOAQUIM" |
prenom_ind=="JOAO" | prenom_ind=="JORGE" | prenom_ind=="LUIS" |
prenom_ind=="FRANCISCO" | prenom_ind=="RAFAEL" | prenom_ind=="FERNAND" |
prenom_ind=="MIGUEL" | prenom_ind=="JOSE MANUEL" | prenom_ind=="JUAN" |
prenom_ind=="DANIEL" | prenom_ind=="") /*d'après la nomenclature de l'INED, 3
prénoms du Top 20 des prénoms des H G1 Europe du Sud sont d'une origine autre que "latin":
"ANTOINE", "FERNAND" et "DANIEL"*/
tab prenom_ind catl_ined_rec if (prenom_ind=="MARIA" | prenom_ind=="MARIE" |
prenom_ind=="ISABELLE" | prenom_ind=="ISABEL" | prenom_ind=="FATIMA" | prenom_ind=="ROSA"
| prenom_ind=="LUCIA" | prenom_ind=="PAULA" | prenom_ind=="MARIE HELENE" |
prenom_ind=="DEOLINDA" | prenom_ind=="GLORIA" |
prenom_ind=="MARIE JOSE" | prenom_ind=="NATALIA" | prenom_ind=="MARIA ISABEL" |
prenom_ind=="JOSEPHINE" | prenom_ind=="CARMEN" | prenom_ind=="ADELAIDE" |
prenom_ind=="ANA" | prenom_ind=="") /*d'après la nomenclature de l'INED, 10
prénoms du Top 20 des prénoms des F G1 Europe du Sud sont d'une origine autre que "latin":
"MARIA", "MARIE", "ISABELLE", "FATIMA", "ROSA", "MARIE-HELENE", "GLORIA", "MARIE-JOSE",
"JOSEPHINE" et "ADELAIDE"*/

```

```

tab prenom_ind cat1_ined_rec if (prenom_ind=="DANIEL" | prenom_ind=="DAVID" |
prenom_ind=="PHILIPPE" | prenom_ind=="MICHEL" | prenom_ind=="STEPHANE" |
prenom_ind=="GILBERT" | prenom_ind=="FREDERIC" | prenom_ind=="PIERRE" |
prenom_ind=="CHRISTOPHE" | prenom_ind=="DOMINIQUE" | prenom_ind=="NICOLAS" |
prenom_ind=="BRUNO" | prenom_ind=="ALEXANDRE" | prenom_ind=="SEBASTIEN" |
prenom_ind=="AIME" | prenom_ind=="PATRICK" | prenom_ind=="ALBERT" |
prenom_ind=="JOSE" | prenom_ind=="CHRISTIAN" | prenom_ind=="JOSEPH") /*d'après la
nomenclature de l'INED, 19 prénoms du Top 20 des prénoms des H G2 Europe du Sud sont d'une
origine autre que "latin": tous sauf "JOSE"*/
tab prenom_ind cat1_ined_rec if (prenom_ind=="MARIE CARMEN" | prenom_ind=="ELISABETH" |
prenom_ind=="EVELYNE" | prenom_ind=="SYLVIE" | prenom_ind=="ISABELLE" | prenom_ind=="ANNIE"
| prenom_ind=="MARYSE" | prenom_ind==" | prenom_ind=="MARIE JEANNE" |
prenom_ind=="SANDRINE" | prenom_ind=="NATHALIE" | prenom_ind=="STEPHANIE" |
prenom_ind=="CHRISTINE" | prenom_ind=="CELINE" | prenom_ind=="SANDRA" |
prenom_ind=="PATRICIA" | prenom_ind=="MELANIE" | prenom_ind=="SOPHIE" |
prenom_ind=="FRANCOISE" | prenom_ind=="BEATRICE") /*d'après la nomenclature de l'INED, 20
prénoms du Top 20 des prénoms des F G2 Europe du Sud sont d'une origine autre que "latin":
tous*/
tab prenom_ind cat1_ined_rec if (prenom_ind=="JEAN BAPTISTE" | prenom_ind=="SAMUEL" |
prenom_ind=="THOMAS" | prenom_ind=="RICHARD" | prenom_ind=="ENZO" |
prenom_ind=="PIERRE" | prenom_ind=="EMMANUEL" | prenom_ind=="FRANCOIS" | prenom_ind=="LUCAS"
| prenom_ind=="NICOLAS" | prenom_ind==" | prenom_ind=="JULIEN" |
prenom_ind=="ANTHONY" | prenom_ind=="ADRIEN" | prenom_ind=="KEVIN" |
prenom_ind=="ALEXANDRE" | prenom_ind=="DAMIEN" | prenom_ind=="MAXIME" |
prenom_ind=="HUGO" | prenom_ind=="ALEXIS") /*d'après la nomenclature de l'INED, 19
prénoms du Top 20 des prénoms des H G3 Europe du Sud sont d'une origine autre que "latin":
tous sauf "ENZO"*/
tab prenom_ind cat1_ined_rec if (prenom_ind=="ANNE LAURE" | prenom_ind=="EMILIE" |
prenom_ind=="CARINNE" | prenom_ind=="VIRGINIE" | prenom_ind=="CELINE" |
prenom_ind=="JOHANNA" | prenom_ind=="MAEVA" | prenom_ind=="MAGALI" |
prenom_ind=="SEVERINE" | prenom_ind=="CHARLOTTE" | prenom_ind=="MAUDE" |
prenom_ind=="LAURA" | prenom_ind=="SOPHIE" | prenom_ind=="LEA" |
prenom_ind=="CAROLINE" | prenom_ind=="JULIE" | prenom_ind=="MARIE" |
prenom_ind=="CAMILLE" | prenom_ind==" | prenom_ind=="CLARA") /*d'après la
nomenclature de l'INED, 20 prénoms du Top 20 des prénoms des F G3 Europe du Sud sont d'une
origine autre que "latin": tous*/

```

\*Tableau 7.4. Top 20 des prénoms des habitants originaires du Maghreb selon leur "génération", en France en 2008 (enquête TeO), N = 3 155

```

groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==120) [aw=poids_ind], order(h)
select(20) /*Top 20 pondéré des prénoms des H G1 Maghreb*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==120) [aw=poids_ind], order(h)
select(20) /*Top 20 pondéré des prénoms des F G1 Maghreb*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==320) [aw=poids_ind], order(h)
select(20) /*Top 20 pondéré des prénoms des H G2 Maghreb*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==320) [aw=poids_ind], order(h)
select(20) /*Top 20 pondéré des prénoms des F G2 Maghreb*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==620) [aw=poids_ind], order(h)
select(20) /*Top 20 pondéré des prénoms des H G3 Maghreb*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==620) [aw=poids_ind], order(h)
select(20) /*Top 20 pondéré des prénoms des F G3 Maghreb*/

```

```

groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==120) , order(h)
select(39) /*Top non pondéré des prénoms des H G1 Maghreb*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==120) , order(h)
select(48) /*Top non pondéré des prénoms des F G1 Maghreb*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==320) , order(h)
select(118) /*Top non pondéré des prénoms des H G2 Maghreb*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==320) , order(h)
select(124) /*Top non pondéré des prénoms des F G2 Maghreb*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==620) , order(h)
select(21) /*Top non pondéré des prénoms des H G3 Maghreb*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==620) , order(h)
select(22) /*Top non pondéré des prénoms des F G3 Maghreb*/

```

```

tab prenom_ind cat1_ined_rec if (prenom_ind=="MOHAMED" | prenom_ind=="AHMED" |
prenom_ind=="RACHID" | prenom_ind=="MOHAMMED" | prenom_ind=="ABDELKADER" |
prenom_ind=="ABDELLAH" | prenom_ind=="KHALID" | prenom_ind=="MOURAD" | prenom_ind=="MUSTAPHA"
| prenom_ind=="ALI" | prenom_ind=="BRAHIM" | prenom_ind=="HAMID" | prenom_ind=="HICHAM"
| prenom_ind=="SAID" | prenom_ind=="OMAR" | prenom_ind=="WALID" | prenom_ind=="NABIL"
| prenom_ind=="DRISS" | prenom_ind=="HASSAN" | prenom_ind=="KARIM") /*d'après la
nomenclature de l'INED, 0 prénom du Top 20 des prénoms des H G1 Maghreb sont d'une origine
autre que "arabo-musulman": ils sont tous "arabo-musulmans"*/
tab prenom_ind cat1_ined_rec if (prenom_ind=="FATIMA" | prenom_ind=="KHADIJA" |
prenom_ind=="SAMIRA" | prenom_ind=="FATIHA" | prenom_ind=="AÏCHA" |
prenom_ind=="MALIKA" | prenom_ind=="NAÏMA" | prenom_ind=="NADIA" | prenom_ind=="ZOHRA"
| prenom_ind=="KARIMA" | prenom_ind=="SAMIA" | prenom_ind=="HOURIA" |

```



```

prenom_ind=="RACHIDA" | prenom_ind=="MERIEM" | prenom_ind=="SAIDA" | prenom_ind=="FATMA" |
prenom_ind=="BOUCHRA" | prenom_ind=="YAMINA" | prenom_ind=="JAMILA" | prenom_ind=="NAJAT")
/*d'après la nomenclature de l'INED, 1 prénom du Top 20 des prénoms des F G1 Maghreb sont
d'une origine autre que "arabo-musulman": "NADIA"*/
tab prenom_ind cat1_ined_rec if (prenom_ind=="MOHAMED" | prenom_ind=="KARIM" |
prenom_ind=="FARID" | prenom_ind=="MEHDI" | prenom_ind=="RACHID" |
prenom_ind=="SAMIR" | prenom_ind=="KAMEL" | prenom_ind=="ADAM" | prenom_ind=="DANIEL"
| prenom_ind=="SOFIANE" | prenom_ind=="YANIS" | prenom_ind=="YASSINE" | prenom_ind=="SAID"
| prenom_ind=="AHMED" | prenom_ind=="AMINE" | prenom_ind=="YOUSSEF" | prenom_ind=="YACINE"
| prenom_ind=="NORDINE" | prenom_ind=="ALI" | prenom_ind=="SAMY") /*d'après la
nomenclature de l'INED, 2 prénoms du Top 20 des prénoms des H G2 Maghreb sont d'une origine
autre que "arabo-musulman": "ADAM" et "DANIEL"*/
tab prenom_ind cat1_ined_rec if (prenom_ind=="SARAH" | prenom_ind=="NADIA" |
prenom_ind=="MYRIAM" | prenom_ind=="FATIMA" | prenom_ind=="YAMINA" | prenom_ind=="INES"
| prenom_ind=="MALIKA" | prenom_ind=="SONIA" | prenom_ind=="SABRINA" | prenom_ind=="NORA"
| prenom_ind=="SAMIRA" | prenom_ind=="LEILA" | prenom_ind=="HOURLIA" | prenom_ind=="ANISSA"
| prenom_ind=="SAMIA" | prenom_ind=="YASMINA" | prenom_ind=="AICHA" | prenom_ind=="LINA"
| prenom_ind=="YASMINE" | prenom_ind=="KARIMA") /*d'après la nomenclature de l'INED, 7
prénoms du Top 20 des prénoms des F G2 Maghreb sont d'une origine autre que "arabo-musulman":
"SARAH", "NADIA", "MYRIAM", "INES", "SONIA", "SABRINA" et "LINA"*/
tab prenom_ind cat1_ined_rec if (prenom_ind=="KARIM" | prenom_ind=="NASSIM" |
prenom_ind=="FARES" | prenom_ind=="NICOLAS" | prenom_ind=="JEREMY" |
prenom_ind=="MOURAD" | prenom_ind=="YANIS" | prenom_ind=="MEHDI" | prenom_ind=="KEVIN"
| prenom_ind=="YACINE" | prenom_ind=="ILYES" | prenom_ind=="RAYAN" | prenom_ind=="BENOIT"
| prenom_ind=="DRISS" | prenom_ind=="MOHAMED" | prenom_ind=="RAYANE" | prenom_ind=="SOFIAN"
| prenom_ind=="ADAM" | prenom_ind=="") /*d'après la
nomenclature de l'INED, 6 prénoms du Top 20 des prénoms des H G3 Maghreb sont d'une origine
autre que "arabo-musulman": "NICOLAS", "JEREMY", "KEVIN", "BENOIT", "ADAM" et "s"*/
tab prenom_ind cat1_ined_rec if (prenom_ind=="SARAH" | prenom_ind=="SOPHIA" |
prenom_ind=="SABRINA" | prenom_ind=="LINA" | prenom_ind=="INES" |
prenom_ind=="SOFIA" | prenom_ind=="CELIA" | prenom_ind=="JESSICA" | prenom_ind=="LEILA"
| prenom_ind=="MYRIAM" | prenom_ind=="ANAISS" | prenom_ind=="VIRGINIE" |
prenom_ind=="MELISSA" | prenom_ind=="LILIA" | prenom_ind=="AICHA" | prenom_ind=="LISA" |
prenom_ind=="EMMA" | prenom_ind=="") /*d'après la nomenclature de l'INED, 17 prénoms du Top 20 des prénoms des F G3 Maghreb sont
d'une origine autre que "arabo-musulman": tous sauf "LEILA", "LILIA" et "AICHA"*/

```

/\*Chapitre 7: La reproduction "au plus près" du tableau page 3\*/

\*Tableau 7.5a: champ retenu, données pondérées, N = 42 208

```

tab sexe_ind if (bon_champ==1) [aw=poids_ind], missing
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==140) [aw=poids_ind], order(h)
select(3) /*Top 3 des prénoms des H G1 Eur Sud: José, Antonio et Manuel*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==340) [aw=poids_ind], order(h)
select(3) /*Top 3 des prénoms des H G2 Eur Sud: Daniel, David et Philippe*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==640) [aw=poids_ind], order(h)
select(3) /*Top 3 des prénoms des H G3 Eur Sud: Jean-Baptiste, Samuel et Thomas*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==140) [aw=poids_ind], order(h)
select(3) /*Top 3 des prénoms des F G1 Eur Sud: Maria, Marie et Isabelle*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==340) [aw=poids_ind], order(h)
select(3) /*Top 3 des prénoms des F G2 Eur Sud: Marie-Carmen, Elisabeth et Evelyne*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==640) [aw=poids_ind], order(h)
select(3) /*Top 3 des prénoms des F G3 Eur Sud: Anne-Laure, Emilie et Carinne*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==120) [aw=poids_ind], order(h)
select(3) /*Top 3 des prénoms des H G1 Maghreb: Mohamed, Ahmed et Rachid*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==320) [aw=poids_ind], order(h)
select(3) /*Top 3 des prénoms des H G2 Maghreb: Mohamed, Karim et Farid*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==620) [aw=poids_ind], order(h)
select(3) /*Top 3 des prénoms des H G3 Maghreb: Karim, Nassim et Farès*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==120) [aw=poids_ind], order(h)
select(3) /*Top 3 des prénoms des F G1 Maghreb: Fatima, Khadija et Samira*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==320) [aw=poids_ind], order(h)
select(3) /*Top 3 des prénoms des F G2 Maghreb: Sarah, Nadia et Myriam*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==620) [aw=poids_ind], order(h)
select(3) /*Top 3 des prénoms des F G3 Maghreb: Sarah, Sophia et Sabrina*/

```

\*Tableau 7.5b: champ retenu, données non pondérées, N = 42 208

```

groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==140) , order(h)
select(3) /*Top 3 des prénoms des H G1 Eur Sud: José, Antonio et Manuel*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==340) , order(h)
select(3) /*Top 3 des prénoms des H G2 Eur Sud: David, Alexandre et Philippe*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==640) , order(h)
select(3) /*Top 3 des prénoms des H G3 Eur Sud: Thomas, Enzo et Nicolas*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==140) , order(h)
select(3) /*Top 3 des prénoms des F G1 Eur Sud: Maria, Marie et Isabelle*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==340) , order(h)

```

```

select(3) /*Top 3 des prénoms des F G2 Eur Sud: Isabelle, Sylvie et Nathalie*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==640) , order(h)
select(3) /*Top 3 des prénoms des F G3 Eur Sud: Laura, Léa et Julie*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==120) , order(h)
select(3) /*Top 3 des prénoms des H G1 Maghreb: Mohamed, Ahmed et Rachid*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==320) , order(h)
select(3) /*Top 3 des prénoms des H G2 Maghreb: Mohamed, Karim et Mehdi*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==620) , order(h)
select(3) /*Top 3 des prénoms des H G3 Maghreb: Yanis, Mehdi et Nicolas*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==120) , order(h)
select(3) /*Top 3 des prénoms des F G1 Maghreb: Fatima, Fatiha et Khadija*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==320) , order(h)
select(3) /*Top 3 des prénoms des F G2 Maghreb: Sarah, Myriam et Nadia*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==620) , order(h)
select(3) /*Top 3 des prénoms des F G3 Maghreb: Sarah, Inès et Lina*/

```

\*Tableau 7.5c: champ maximal, données pondérées, N = 83 045

```

tab sexe_ind if (lienego_ind>=0 & lienego_ind<=201) [aw=poids_ind], missing
groups prenom_ind if (lienego_ind>=0 & lienego_ind<=201 & sexe_ind==1 & orig_ind==140)
[aw=poids_ind], order(h) select(3) /*Top 3 des prénoms des H G1 Eur Sud: José, Manuel et
Antonio*/
groups prenom_ind if (lienego_ind>=0 & lienego_ind<=201 & sexe_ind==1 & orig_ind==340)
[aw=poids_ind], order(h) select(3) /*Top 3 des prénoms des H G2 Eur Sud: Pierre, François et
Philippe*/
groups prenom_ind if (lienego_ind>=0 & lienego_ind<=201 & sexe_ind==1 & orig_ind==640)
[aw=poids_ind], order(h) select(3) /*Top 3 des prénoms des H G3 Eur Sud: Sébastien, Maxime et
Thomas*/
groups prenom_ind if (lienego_ind>=0 & lienego_ind<=201 & sexe_ind==2 & orig_ind==140)
[aw=poids_ind], order(h) select(3) /*Top 3 des prénoms des F G1 Eur Sud: Maria, Maria Inès et
Carmelle*/
groups prenom_ind if (lienego_ind>=0 & lienego_ind<=201 & sexe_ind==2 & orig_ind==340)
[aw=poids_ind], order(h) select(3) /*Top 3 des prénoms des F G2 Eur Sud: Isabelle, Christine
et Elisabeth*/
groups prenom_ind if (lienego_ind>=0 & lienego_ind<=201 & sexe_ind==2 & orig_ind==640)
[aw=poids_ind], order(h) select(3) /*Top 3 des prénoms des F G3 Eur Sud: Elodie, Mélanie et
Virginie*/
groups prenom_ind if (lienego_ind>=0 & lienego_ind<=201 & sexe_ind==1 & orig_ind==120)
[aw=poids_ind], order(h) select(3) /*Top 3 des prénoms des H G1 Maghreb: Mohamed, Ahmed et
Rachid*/
groups prenom_ind if (lienego_ind>=0 & lienego_ind<=201 & sexe_ind==1 & orig_ind==320)
[aw=poids_ind], order(h) select(3) /*Top 3 des prénoms des H G2 Maghreb: Mohamed, Karim et
Mehdi*/
groups prenom_ind if (lienego_ind>=0 & lienego_ind<=201 & sexe_ind==1 & orig_ind==620)
[aw=poids_ind], order(h) select(3) /*Top 3 des prénoms des H G3 Maghreb: Yanis, Nicolas et
Karim*/
groups prenom_ind if (lienego_ind>=0 & lienego_ind<=201 & sexe_ind==2 & orig_ind==120)
[aw=poids_ind], order(h) select(3) /*Top 3 des prénoms des F G1 Maghreb: Fatima, Khadija et
Aicha*/
groups prenom_ind if (lienego_ind>=0 & lienego_ind<=201 & sexe_ind==2 & orig_ind==320)
[aw=poids_ind], order(h) select(3) /*Top 3 des prénoms des F G2 Maghreb: Nadia, Sarah et
Fatima*/
groups prenom_ind if (lienego_ind>=0 & lienego_ind<=201 & sexe_ind==2 & orig_ind==620)
[aw=poids_ind], order(h) select(3) /*Top 3 des prénoms des F G3 Maghreb: Inès, Sarah et
Audrey*/

```

\*Tableau 7.5d: champ maximal, données non pondérées, N = 83 045

```

groups prenom_ind if (lienego_ind>=0 & lienego_ind<=201 & sexe_ind==1 & orig_ind==140)
, order(h) select(3) /*Top 3 des prénoms des H G1 Eur Sud: José, Manuel et Antonio*/
groups prenom_ind if (lienego_ind>=0 & lienego_ind<=201 & sexe_ind==1 & orig_ind==340)
, order(h) select(3) /*Top 3 des prénoms des H G2 Eur Sud: David, Alexandre et Philippe*/
groups prenom_ind if (lienego_ind>=0 & lienego_ind<=201 & sexe_ind==1 & orig_ind==640)
, order(h) select(3) /*Top 3 des prénoms des H G3 Eur Sud: Thomas, Nicolas et Lucas*/
groups prenom_ind if (lienego_ind>=0 & lienego_ind<=201 & sexe_ind==2 & orig_ind==140)
, order(h) select(3) /*Top 3 des prénoms des F G1 Eur Sud: Maria, Marie et Rosa*/
groups prenom_ind if (lienego_ind>=0 & lienego_ind<=201 & sexe_ind==2 & orig_ind==340)
, order(h) select(3) /*Top 3 des prénoms des F G2 Eur Sud: Sandrine, Nathalie et Sandra*/
groups prenom_ind if (lienego_ind>=0 & lienego_ind<=201 & sexe_ind==2 & orig_ind==640)
, order(h) select(3) /*Top 3 des prénoms des F G3 Eur Sud: Laura, Léa et Julie*/
groups prenom_ind if (lienego_ind>=0 & lienego_ind<=201 & sexe_ind==1 & orig_ind==120)
, order(h) select(3) /*Top 3 des prénoms des H G1 Maghreb: Mohamed, Ahmed et Ali*/
groups prenom_ind if (lienego_ind>=0 & lienego_ind<=201 & sexe_ind==1 & orig_ind==320)
, order(h) select(3) /*Top 3 des prénoms des H G2 Maghreb: Mohamed, Karim et Mehdi*/
groups prenom_ind if (lienego_ind>=0 & lienego_ind<=201 & sexe_ind==1 & orig_ind==620)
, order(h) select(3) /*Top 3 des prénoms des H G3 Maghreb: Yanis, Mehdi et Mohamed*/
groups prenom_ind if (lienego_ind>=0 & lienego_ind<=201 & sexe_ind==2 & orig_ind==120)
, order(h) select(3) /*Top 3 des prénoms des F G1 Maghreb: Fatima, Aicha et Malika*/
groups prenom_ind if (lienego_ind>=0 & lienego_ind<=201 & sexe_ind==2 & orig_ind==320)

```

```

, order(h) select(3) /*Top 3 des prénoms des F G2 Maghreb: Nadia, Sarah et Fatima*/
groups prenom_ind if (lienego_ind>=0 & lienego_ind<=201 & sexe_ind==2 & orig_ind==620)
, order(h) select(3) /*Top 3 des prénoms des F G3 Maghreb: Sarah, Inès et Lina*/

*Tableau 7.5e: champ enquêtés et tous leurs enfants vivants, données pondérées, N = 52 903
tab sexe_ind if (lienego_ind==0 | lienego_ind==2 | lienego_ind==201) [aw=poids_ind],
missing
groups prenom_ind if ((lienego_ind==0 | lienego_ind==2 | lienego_ind==201) & sexe_ind==1 &
orig_ind==140) [aw=poids_ind], order(h) select(3) /*Top 3 des prénoms des H G1 Eur Sud: José,
Antonio et Manuel*/
groups prenom_ind if ((lienego_ind==0 | lienego_ind==2 | lienego_ind==201) & sexe_ind==1 &
orig_ind==340) [aw=poids_ind], order(h) select(3) /*Top 3 des prénoms des H G2 Eur Sud: David,
Daniel et Sébastien*/
groups prenom_ind if ((lienego_ind==0 | lienego_ind==2 | lienego_ind==201) & sexe_ind==1 &
orig_ind==640) [aw=poids_ind], order(h) select(3) /*Top 3 des prénoms des H G3 Eur Sud:
Sébastien, Maxime et François*/
groups prenom_ind if ((lienego_ind==0 | lienego_ind==2 | lienego_ind==201) & sexe_ind==2 &
orig_ind==140) [aw=poids_ind], order(h) select(3) /*Top 3 des prénoms des F G1 Eur Sud: Maria,
Marie et Isabelle*/
groups prenom_ind if ((lienego_ind==0 | lienego_ind==2 | lienego_ind==201) & sexe_ind==2 &
orig_ind==340) [aw=poids_ind], order(h) select(3) /*Top 3 des prénoms des F G2 Eur Sud:
Elisabeth, Sylvie et Marie Carmen*/
groups prenom_ind if ((lienego_ind==0 | lienego_ind==2 | lienego_ind==201) & sexe_ind==2 &
orig_ind==640) [aw=poids_ind], order(h) select(3) /*Top 3 des prénoms des F G3 Eur Sud:
Elodie, Mélanie et Virginie*/
groups prenom_ind if ((lienego_ind==0 | lienego_ind==2 | lienego_ind==201) & sexe_ind==1 &
orig_ind==120) [aw=poids_ind], order(h) select(3) /*Top 3 des prénoms des H G1 Maghreb:
Mohamed, Ahmed et Rachid*/
groups prenom_ind if ((lienego_ind==0 | lienego_ind==2 | lienego_ind==201) & sexe_ind==1 &
orig_ind==320) [aw=poids_ind], order(h) select(3) /*Top 3 des prénoms des H G2 Maghreb:
Mohamed, Karim et Samir*/
groups prenom_ind if ((lienego_ind==0 | lienego_ind==2 | lienego_ind==201) & sexe_ind==1 &
orig_ind==620) [aw=poids_ind], order(h) select(3) /*Top 3 des prénoms des H G3 Maghreb:
Nicolas, Yanis et Karim*/
groups prenom_ind if ((lienego_ind==0 | lienego_ind==2 | lienego_ind==201) & sexe_ind==2 &
orig_ind==120) [aw=poids_ind], order(h) select(3) /*Top 3 des prénoms des F G1 Maghreb:
Fatima, Khadija et Samira*/
groups prenom_ind if ((lienego_ind==0 | lienego_ind==2 | lienego_ind==201) & sexe_ind==2 &
orig_ind==320) [aw=poids_ind], order(h) select(3) /*Top 3 des prénoms des F G2 Maghreb: Sarah,
Nadia et Fatima*/
groups prenom_ind if ((lienego_ind==0 | lienego_ind==2 | lienego_ind==201) & sexe_ind==2 &
orig_ind==620) [aw=poids_ind], order(h) select(3) /*Top 3 des prénoms des F G3 Maghreb: Inès,
Sarah et Noelia*/

*Tableau 7.5f: champ enquêtés et tous leurs enfants vivants, données non pondérées, N = 52
903
groups prenom_ind if ((lienego_ind==0 | lienego_ind==2 | lienego_ind==201) & sexe_ind==1 &
orig_ind==140) , order(h) select(3) /*Top 3 des prénoms des H G1 Eur Sud: José,
Antonio et Manuel*/
groups prenom_ind if ((lienego_ind==0 | lienego_ind==2 | lienego_ind==201) & sexe_ind==1 &
orig_ind==340) , order(h) select(3) /*Top 3 des prénoms des H G2 Eur Sud: David,
Alexandre et Philippe*/
groups prenom_ind if ((lienego_ind==0 | lienego_ind==2 | lienego_ind==201) & sexe_ind==1 &
orig_ind==640) , order(h) select(3) /*Top 3 des prénoms des H G3 Eur Sud:
Thomas, Nicolas et Lucas*/
groups prenom_ind if ((lienego_ind==0 | lienego_ind==2 | lienego_ind==201) & sexe_ind==2 &
orig_ind==140) , order(h) select(3) /*Top 3 des prénoms des F G1 Eur Sud: Maria,
Marie et Isabelle*/
groups prenom_ind if ((lienego_ind==0 | lienego_ind==2 | lienego_ind==201) & sexe_ind==2 &
orig_ind==340) , order(h) select(3) /*Top 3 des prénoms des F G2 Eur Sud:
Sandrine, Sandra et Isabelle*/
groups prenom_ind if ((lienego_ind==0 | lienego_ind==2 | lienego_ind==201) & sexe_ind==2 &
orig_ind==640) , order(h) select(3) /*Top 3 des prénoms des F G3 Eur Sud: Laura,
Léa et Julie*/
groups prenom_ind if ((lienego_ind==0 | lienego_ind==2 | lienego_ind==201) & sexe_ind==1 &
orig_ind==120) , order(h) select(3) /*Top 3 des prénoms des H G1 Maghreb:
Mohamed, Ahmed et Rachid*/
groups prenom_ind if ((lienego_ind==0 | lienego_ind==2 | lienego_ind==201) & sexe_ind==1 &
orig_ind==320) , order(h) select(3) /*Top 3 des prénoms des H G2 Maghreb:
Mohamed, Karim et Mehdi*/
groups prenom_ind if ((lienego_ind==0 | lienego_ind==2 | lienego_ind==201) & sexe_ind==1 &
orig_ind==620) , order(h) select(3) /*Top 3 des prénoms des H G3 Maghreb: Yanis,
Mehdi et Nicolas*/
groups prenom_ind if ((lienego_ind==0 | lienego_ind==2 | lienego_ind==201) & sexe_ind==2 &
orig_ind==120) , order(h) select(3) /*Top 3 des prénoms des F G1 Maghreb:
Fatima, Fatiha et Khadija*/
groups prenom_ind if ((lienego_ind==0 | lienego_ind==2 | lienego_ind==201) & sexe_ind==2 &

```

```

orig_ind==320) , order(h) select(3) /*Top 3 des prénoms des F G2 Maghreb: Sarah,
Nadia et Fatima*/
groups prenom_ind if ((lienego_ind==0 | lienego_ind==2 | lienego_ind==201) & sexe_ind==2 &
orig_ind==620) , order(h) select(3) /*Top 3 des prénoms des F G3 Maghreb: Sarah,
Inès et Lina*/

```

/\*Chapitre 7: A la recherche de petits-fils d'immigrés du Maghreb prénommés "Nicolas": auprès des experts des migrations et chez les couples "mixtes"\*/

/\*Chapitre 7: A la recherche de petits-fils d'immigrés du Maghreb prénommés "Nicolas": dans l'annuaire et sur les réseaux sociaux\*/

```

generate nicolasmaghreb = 1 if (sexe_ind==1 & prenom_ind=="NICOLAS" & (orig_ind==120 |
orig_ind==320 | orig_ind==620 | orig_ind==820))
/*tab orig_ind nicolasmaghreb if bon_champ==1 [aw=poids_ind], missing*/ /*parmi les
hommes du champ retenu (données pondérées), l'enquête TeO compte 10 "Nicolas" originaires
du Maghreb (0 G1 et 3 G2 et 6 G3)*/
bysort nicolasmaghreb: summarize anais_ind if bon_champ==1 [aw=poids_ind] /*ces "Nicolas"
originaires du Maghreb sont nés de 1967 à 2001, en moyenne en 1983*/
tab orig_ind nicolasmaghreb if bon_champ==1 , missing /*parmi les
hommes du champ retenu (données non pondérées), l'enquête TeO compte 17 "Nicolas" originaires
du Maghreb (0 G1 et 7 G2 et 10 G3)*/

```

```

generate yanismaghreb = 1 if (sexe_ind==1 & prenom_ind=="YANIS" & (orig_ind==120 |
orig_ind==320 | orig_ind==620 | orig_ind==820))
/*tab orig_ind yanismaghreb if bon_champ==1 [aw=poids_ind], missing*/ /*parmi les
hommes du champ retenu (données pondérées), l'enquête TeO compte 12 "Yanis" originaires du
Maghreb (0 G1 et 8 G2 et 3 G3)*/
bysort yanismaghreb: summarize anais_ind if bon_champ==1 [aw=poids_ind] /*ces "Yanis"
originaires du Maghreb sont nés de 1994 à 2007, en moyenne en 2003*/

```

```

generate nicolaseuropesud = 1 if (sexe_ind==1 & prenom_ind=="NICOLAS" & (orig_ind==140 |
orig_ind==340 | orig_ind==640 | orig_ind==840))
/*tab orig_ind nicolaseuropesud if bon_champ==1 [aw=poids_ind], missing*/ /*parmi les
hommes du champ retenu (données pondérées), l'enquête TeO compte 17 "Nicolas" originaires
d'Europe du Sud (0 G1 et 12 G2 et 5 G3)*/
bysort nicolaseuropesud: summarize anais_ind if bon_champ==1 [aw=poids_ind] /*ces "Nicolas"
originaires d'Europe du Sud sont nés de 1967 à 2007, en moyenne en 1987*/

```

/\*Chapitre 7: A la recherche de petits-fils d'immigrés du Maghreb prénommés "Nicolas": dans les médias\*/

/\*Chapitre 7: Quelle erreur l'enquête TeO pourrait-elle contenir?\*/

```

bysort nicolasmaghreb: summarize poids_ind if bon_champ==1 /*sur le champ retenu (non
pondéré), les 17 "Nicolas" originaires du Maghreb ont en moyenne un poids de 625, contre 1 112
en moyenne chez les autres individus*/
bysort nicolasmaghreb: summarize poids_ind /*sur le champ maximal (non
pondéré), les 19 "Nicolas" originaires du Maghreb ont en moyenne un poids de 833, contre 1 157
en moyenne chez les autres individus*/

```

/\*Chapitre 7: Les petits-fils d'immigrés du Maghreb prénommés "Nicolas" seraient-ils, en réalité, des petits-fils de pieds-noirs?\*/

```

*Tableau 7.10. Prénoms de l'entourage maghrébin des petits-fils d'immigrés du Maghreb
prénommés "Nicolas" et "Yanis", en France en 2008 (enquête TeO), N = 97
generate nicolasg3maghreb = 1 if (sexe_ind==1 & prenom_ind=="NICOLAS" & orig_ind==620)
tab nicolasg3maghreb, missing /*sur le champ maximal non pondéré, l'enquête TeO compte 11
"Nicolas" petits-fils d'immigrés du Maghreb*/
sort nicolasg3maghreb
egen nicolasg3maghreb_entourage = max(nicolasg3maghreb), by(ident)
replace nicolasg3maghreb_entourage = . if (nicolasg3maghreb_entourage==1 &
nicolasg3maghreb==1)
sort nicolasg3maghreb_entourage ident noi
tab orig_ind cat1_ined_rec if (nicolasg3maghreb_entourage==1 & (orig_ind==120 |
orig_ind==320 | orig_ind==620)), missing /*sur le champ maximal non pondéré, l'entourage
maghrébin des 11 "Nicolas" petits-fils d'immigrés du Maghreb comprend 0 grand-parent G1
Maghreb, 11 parents G2 Maghreb, et 19 frères et soeurs G3 Maghreb, soit au total 30
individus*/

generate yanisg3maghreb = 1 if (sexe_ind==1 & prenom_ind=="YANIS" & orig_ind==620)
tab yanisg3maghreb, missing /*sur le champ maximal non pondéré, l'enquête TeO compte 23
"Yanis" petits-fils d'immigrés du Maghreb*/

```

```

    sort yanisg3maghreb
    egen yanisg3maghreb_entourage = max(yanisg3maghreb), by(ident)
    replace yanisg3maghreb_entourage = . if (yanisg3maghreb_entourage==1 & yanisg3maghreb==1)
    sort yanisg3maghreb_entourage ident noi
    tab orig_ind cat1 ined rec if (yanisg3maghreb_entourage==1 & (orig_ind==120 |
orig_ind==320 | orig_ind==620)), missing /*sur le champ maximal non pondéré, l'entourage
maghrébin des 23 "Yanis" petits-fils d'immigrés du Maghreb comprend 3 grands-parents G1
Maghreb, 35 parents G2 Maghreb, et 29 frères et soeurs G3 Maghreb, soit au total 67
individus*/

*Combien d'enquêtés enfants de rapatriés d'Afrique du Nord et enfants d'immigrés du Maghreb
ont au moins un parent qui leur parlait dans leur enfance en arabe ou en berbère (comme 1e ou
2e langue)
generate arabe_ou_berbere = 1 if (lnp1_gr=="31" | lnp1_gr=="32" | lnp2_gr=="31" |
lnp2_gr=="32" | lnm1_gr=="31" | lnm1_gr=="32" | lnm2_gr=="31" | lnm2_gr=="32")
    tab arabe_ou_berbere if (lienego_ind==0) & (group2==51) & (regionnaisp2==20 |
regionnaism2==20) [aw=poids_ind], missing /* 5 % des enquêtés enfants de rapatriés d'Afrique
du Nord ont au moins un parent qui leur parlait dans leur enfance en arabe ou en berbère
(comme 1e ou 2e langue)*/
    tab arabe_ou_berbere if (lienego_ind==0) & (orig_ind==320)
[aw=poids_ind], missing /*65 % des enquêtés enfants d'immigrés du Maghreb ont au
moins un parent qui leur parlait dans leur enfance en arabe ou en berbère (comme 1e ou 2e
langue)*/

*Combien d'enquêtés enfants de rapatriés d'Afrique du Nord et enfants d'immigrés du Maghreb
sont d'origine musulmane
    tab orig_relig_ind if (lienego_ind==0) & (group2==51) & (regionnaisp2==20 |
regionnaism2==20) [aw=poids_ind], missing /* 6 % des enquêtés enfants de rapatriés d'Afrique
du Nord sont d'origine musulmane*/
    tab orig_relig_ind if (lienego_ind==0) & (orig_ind==320)
[aw=poids_ind], missing /*77 % des enquêtés enfants d'immigrés du Maghreb sont
d'origine musulmane*/

*Combien d'enquêtés enfants de rapatriés d'Afrique du Nord et enfants d'immigrés du Maghreb se
déclarent "plutôt pas d'accord" ou "pas d'accord du tout" avec le fait qu'on les voit comme
des "Français"
    tab x_vufri if (lienego_ind==0) & (group2==51) & (regionnaisp2==20 | regionnaism2==20)
[aw=poids_ind], missing /* 5 % des enquêtés enfants de rapatriés d'Afrique du Nord se
déclarent "plutôt pas d'accord" ou "pas d'accord du tout" avec le fait qu'on les voit comme
des "Français"*/
    tab x_vufri if (lienego_ind==0) & (orig_ind==320)
[aw=poids_ind], missing /*39 % des enquêtés enfants d'immigrés du Maghreb se
déclarent "plutôt pas d'accord" ou "pas d'accord du tout" avec le fait qu'on les voit comme
des "Français"*/

*Tableau 7.11. Profils des 11 petits-fils d'immigrés du Maghreb prénommés "Nicolas", d'après
l'enquête TeO
**1e critère: Langue(s) que le G1 Maghreb parlait au G2 Maghreb dans son enfance
list ident noi orig_ind lienego_ind orig_enq autrep parent orig_ca orig_pc
orig_pereenq lnp1_gr lnp2_gr orig_mereenq lnm1_gr lnm2_gr if (nicolasg3maghreb==1)
**2e critère: Religion du G1 Maghreb
list ident noi orig_ind lienego_ind orig_enq autrep parent orig_ca orig_pc
orig_pereenq relperel orig_mereenq relmerel if (nicolasg3maghreb==1)
**3e critère: Religion du G2 Maghreb
list ident noi orig_ind lienego_ind orig_enq relegol autrep parent orig_ca relcal orig_pc relpc1
if (nicolasg3maghreb==1)
**4e critère: Opinion de l'enquêté parent de "Nicolas" sur: "On me voit comme un Français"
egen x_vufri total = max(x_vufri), by(ident)
list ident noi orig_ind lienego_ind orig_enq x_vufri_total
if (nicolasg3maghreb==1)

*Tableau 7.12. Profils des 23 petits-fils d'immigrés du Maghreb prénommés "Yanis", d'après
l'enquête TeO
**1e critère: Langue(s) que le G1 Maghreb parlait au G2 Maghreb dans son enfance
list ident noi orig_ind lienego_ind orig_enq autrep parent orig_ca orig_pc
orig_pereenq lnp1_gr lnp2_gr orig_mereenq lnm1_gr lnm2_gr if (yanisg3maghreb==1)
**2e critère: Religion du G1 Maghreb
list ident noi orig_ind lienego_ind orig_enq autrep parent orig_ca orig_pc
orig_pereenq relperel orig_mereenq relmerel if (yanisg3maghreb==1)
**3e critère: Religion du G2 Maghreb
list ident noi orig_ind lienego_ind orig_enq relegol autrep parent orig_ca relcal orig_pc relpc1
if (yanisg3maghreb==1)
**4e critère: Opinion de l'enquêté parent de "Yanis" sur: "On me voit comme un Français"
list ident noi orig_ind lienego_ind orig_enq x_vufri_total
if (yanisg3maghreb==1)

*Année de naissance des petits-fils d'immigrés du Maghreb prénommés "Nicolas" et "Yanis"

```

```

bysort nicolasg3maghreb: summarize anais_ind [aw=poids_ind] /*les 11 "Nicolas" petits-fils
d'immigrés du Maghreb sont nés de 1981 à 2001, en moyenne en 1987*/
bysort yanisg3maghreb : summarize anais_ind [aw=poids_ind] /*les 23 "Yanis" petits-fils
d'immigrés du Maghreb sont nés de 1996 à 2007, en moyenne en 2004*/

```

/\*Chapitre 7: Combien de descendants d'immigrés du Maghreb de l'enquête TeO sont, en réalité, des descendants de pieds-noirs?\*/

\*Tableau 7.13. 26% des petits-enfants d'immigrés du Maghreb de l'enquête TeO sont vraisemblablement des petits-enfants de pieds-noirs, N = 42 208

```

*J'estime d'abord, parmi les individus que TeO catégorise comme immigrés ou descendants
d'immigrés du Maghreb, lesquels sont mal et bien catégorisés
**1e critère: les parents originaires du Maghreb de l'enquêté qui est G1 ou G2 Maghreb ne lui
parlaient, dans son enfance, ni arabe ni berbère
generate descendantrapatrie_1 = 1 if (orig_enq==120 | orig_enq==320) & lienego_ind==0 &
(orig_ind==120 | orig_ind==320) & ((orig_pereenq==120 & orig_mereenq!=120 & lnp1_gr!="31" &
lnp1_gr!="32" & lnp2_gr!="31" & lnp2_gr!="32") | (orig_mereenq==120 & orig_pereenq!=120 &
lnm1_gr!="31" & lnm1_gr!="32" & lnm2_gr!="31" & lnm2_gr!="32") | (orig_pereenq==120 &
orig_mereenq==120 & lnp1_gr!="31" & lnp1_gr!="32" & lnp2_gr!="31" & lnp2_gr!="32" &
lnm1_gr!="31" & lnm1_gr!="32" & lnm2_gr!="31" & lnm2_gr!="32")) /*si l'individu est un enquêté
G1 ou G2 Maghreb dont le(s) parent(s) originaire(s) du Maghreb ne lui parlai(en)t ni arabe ni
berbère, alors il est possiblement rapatrié ou enfant de rapatriés plutôt qu'immigré ou enfant
d'immigrés*/
    tab descendantrapatrie_1, missing
**2e critère: les parents originaires du Maghreb de l'enquêté qui est G1 ou G2 Maghreb ne sont
pas musulmans
generate descendantrapatrie_2 = 1 if (orig_enq==120 | orig_enq==320) & lienego_ind==0 &
(orig_ind==120 | orig_ind==320) & ((orig_pereenq==120 & orig_mereenq!=120 & relperel!="20") |
(orig_mereenq==120 & orig_pereenq!=120 & relmerel!="20") | (orig_pereenq==120 &
orig_mereenq==120 & relperel!="20" & relmerel!="20")) /*si l'individu est un enquêté G1 ou G2
Maghreb dont le(s) parent(s) originaire(s) du Maghreb n'est (ne sont) pas musulman(s), alors
il est possiblement rapatrié ou enfant de rapatriés plutôt qu'immigré ou enfant d'immigrés*/
    tab descendantrapatrie_2, missing
**3e critère: l'enquêté qui est G1 ou G2 Maghreb n'est pas musulman non plus
generate descendantrapatrie_3 = 1 if (orig_enq==120 | orig_enq==320) & lienego_ind==0 &
(orig_ind==120 | orig_ind==320) & (relig_ind!=20) /*si l'individu est un enquêté G1 ou G2
Maghreb mais non musulman, alors il est possiblement rapatrié ou enfant de rapatriés plutôt
qu'immigré ou enfant d'immigrés*/
    tab descendantrapatrie_3, missing
**4e critère: l'enquêté qui est G1 ou G2 Maghreb se déclare "plutôt d'accord" ou "tout à fait
d'accord" avec l'idée que "on le voit comme un Français"
generate descendantrapatrie_4 = 1 if (orig_enq==120 | orig_enq==320) & lienego_ind==0 &
(orig_ind==120 | orig_ind==320) & (x_vufri==1 | x_vufri==2) /*si l'individu est un enquêté G1
ou G2 Maghreb d'accord avec l'idée que "on le voit comme un Français", alors il est
possiblement rapatrié ou enfant de rapatriés plutôt qu'immigré ou enfant d'immigrés*/
    tab descendantrapatrie_4, missing
**Si un enquêté remplit tous les critères renseignés et est donc vraisemblablement rapatrié ou
enfant de rapatriés plutôt qu'immigré ou enfant d'immigrés, son entourage familial est
recatégorisé en conséquence
egen descendantrapat_1 = max(descendantrapatrie_1), by(ident)
egen descendantrapat_2 = max(descendantrapatrie_2), by(ident)
egen descendantrapat_3 = max(descendantrapatrie_3), by(ident)
egen descendantrapat_4 = max(descendantrapatrie_4), by(ident)
generate immigr_vrai_ou_rapatrie = 1 if (orig_ind==120 | orig_ind==620 |
orig_ind==820) & (descendantrapat_1==1 & descendantrapat_2==1 & descendantrapat_3==1 &
descendantrapat_4==1)
    replace immigr_vrai_ou_rapatrie = 0 if (orig_ind==120 | orig_ind==320 | orig_ind==620 |
orig_ind==820) & (descendantrapat_1==. | descendantrapat_2==. | descendantrapat_3==. |
descendantrapat_4==.)
    tab orig_ind immigr_vrai_ou_rapatrie if bon_champ==1 [aw=poids_ind], missing /*sur le
champ retenu pondéré, parmi les 3 155 personnes originaires du Maghreb, 11 G1 Maghreb sur 1
043 sont probablement rapatriés, 175 G2 Maghreb sur 1 777 sont probablement enfants de
rapatriés, et 87 G3 Maghreb sur 335 sont probablement petits-enfants de rapatriés*/
/*IMMIGRE_VRAI_OU_RAPATRIE est codée comme suit:
0 = vrai immigré (ou descendant d'immigrés) du Maghreb
1 = probable rapatrié (ou descendant de rapatriés) d'Afrique du Nord*/

```

```

*J'estime ensuite quels individus TeO catégorise comme rapatriés ou descendants de rapatriés
d'Afrique du Nord
generate rapatrie_afn_selon_teo = 0 if
(orig_ind==501 | orig_ind==502 | orig_ind==503 | orig_ind==504)
/*l'individu est du groupe "majoritaire"*/
    replace rapatrie_afn_selon_teo = 1 if lienego_ind==0
& (orig_ind==501 | orig_ind==502 | orig_ind==503 | orig_ind==504) & group2_ind==54 &
regionnais_ind==20

```

```

/*1'enquêté est un rapatrié d'Afrique du Nord*/
  replace rapatrie_afn_selon_teo = 2 if lienego_ind==0
  & (orig_ind==501 | orig_ind==502 | orig_ind==503 | orig_ind==504) & group2_ind==51 &
  (regionnaispere_ind==20 | regionnaismere_ind==20)
/*1'enquêté est un enfant de rapatriés d'Afrique du Nord*/
  replace rapatrie_afn_selon_teo = 1 if lienego_ind==1
  & (orig_ind==501 | orig_ind==502 | orig_ind==503 | orig_ind==504) & group2_ind==54 &
  regionnais_ind==20
/*le conjoint actuel de l'enquêté est un rapatrié d'Afrique du Nord*/
  replace rapatrie_afn_selon_teo = 2 if lienego_ind==1
  & (orig_ind==501 | orig_ind==502 | orig_ind==503 | orig_ind==504) & group2_ind==51 &
  (regionnaispere_ind==20 | regionnaismere_ind==20)
/*le conjoint actuel de l'enquêté est un enfant de rapatriés d'Afrique du Nord*/
  replace rapatrie_afn_selon_teo = 2 if (lienego_ind==2 | lienego_ind==200 | lienego_ind==201)
  & (orig_ind==501 | orig_ind==502 | orig_ind==503 | orig_ind==504) & ((group2==54 &
  regionnais2==20) | (autreparent==0 & group2pc==54 & regionnaispc2==20) | (autreparent==1 &
  group2ca==54 & regionnaisca2==20))
/*1'enfant de l'enquêté est un enfant de rapatriés d'Afrique du Nord*/
  replace rapatrie_afn_selon_teo = 2 if (lienego_ind==2 | lienego_ind==200 | lienego_ind==201)
  & (orig_ind==501 | orig_ind==502 | orig_ind==503 | orig_ind==504) & ((group2==51 &
  regionnaisp2==20) | (group2==51 & regionnaism2==20) | (autreparent==0 & group2pc==51 &
  regionnaispcp2==20) | (autreparent==0 & group2pc==51 & regionnaispcm2==20) | (autreparent==1 &
  group2ca==51 & regionnaiscap2==20) | (autreparent==1 & group2ca==51 & regionnaiscam2==20))
/*1'enfant de l'enquêté est un petit-enfant de rapatriés d'Afrique du Nord*/
  replace rapatrie_afn_selon_teo = 1 if lienego_ind==3
  & (orig_ind==501 | orig_ind==502 | orig_ind==503 | orig_ind==504) & group2==51 &
  regionnais_ind==20
/*le parent de l'enquêté est un rapatrié d'Afrique du Nord*/
  replace rapatrie_afn_selon_teo = 1 if lienego_ind==10
  & (orig_ind==501 | orig_ind==502 | orig_ind==503 | orig_ind==504) & group2==54 &
  regionnais2==20
/*le frère ou la soeur de l'enquêté est un rapatrié d'Afrique du Nord*/
  replace rapatrie_afn_selon_teo = 2 if lienego_ind==10
  & (orig_ind==501 | orig_ind==502 | orig_ind==503 | orig_ind==504) & group2==51
/*le frère ou la soeur de l'enquêté est un enfant de rapatriés d'Afrique du Nord*/
  tab rapatrie_afn_selon_teo if bon_champ==1 [aw=poids_ind], missing /*sur le champ retenu
pondéré, parmi les 31 775 personnes du groupe "majoritaire", 181 individus sont catégorisés
comme rapatriés d'Afrique du Nord, et 1 150 sont catégorisés comme enfant ou petit-enfant de
rapatriés d'Afrique du Nord, contre 30 443 personnes du groupe "majoritaire" qui sont
catégorisées comme ni l'un ni l'autre (et 10 433 non "majoritaires")*/
/*RAPATRIE_AFN_SELON_TEO est codée comme suit:
0 = ni rapatrié ni descendant de rapatriés d'Afrique du nord
1 = rapatrié d'Afrique du nord
2 = enfant ou petit-enfant de rapatriés d'Afrique du nord
. = hors champ (non "majoritaire")*/

*Je crée ensuite la variable ligne du tableau 7.13
generate orig_immigre_rapatrie = 0 if (orig_ind==120 | orig_ind==320 | orig_ind==620 |
orig_ind==820) & (immigre_vrai_ou_rapatrie==0)
  replace orig_immigre_rapatrie = 1 if (orig_ind==120 | orig_ind==320 | orig_ind==620 |
orig_ind==820) & (immigre_vrai_ou_rapatrie==1)
  replace orig_immigre_rapatrie = 2 if (orig_ind==501 | orig_ind==502 | orig_ind==503 |
orig_ind==504) & (rapatrie_afn_selon_teo==1 | rapatrie_afn_selon_teo==2)
  replace orig_immigre_rapatrie = 3 if ((orig_ind==501 | orig_ind==502 | orig_ind==503 |
orig_ind==504) & (rapatrie_afn_selon_teo==0)) | (orig_ind!=120 & orig_ind!=320 & orig_ind!=620
& orig_ind!=820 & orig_ind!=501 & orig_ind!=502 & orig_ind!=503 & orig_ind!=504)
  tab orig_immigre_rapatrie if bon_champ==1 [aw=poids_ind], missing /*sur le champ retenu
pondéré, 2 881 individus sont originaires du Maghreb et bien catégorisés, 273 originaires du
Maghreb et mal catégorisés, 1 332 catégorisés par TeO comme rapatriés ou descendants de
rapatriés, et 37 722 catégorisés autrement*/
/*ORIG_IMMIGRE_RAPATRIE est codée comme suit:
0 = individu que TeO catégorise comme immigré ou descendant d'immigrés du Maghreb et
qui est BIEN catégorisé (probablement immigré du Maghreb ou descendant d'immigrés du
Maghreb)
1 = individu que TeO catégorise comme immigré ou descendant d'immigrés du Maghreb
mais qui est MAL catégorisé (probablement rapatrié d'Afrique du nord ou descendant de
rapatriés d'Afrique du nord)
2 = individu que TeO catégorise comme rapatrié ou descendant de rapatriés d'Afrique du Nord
3 = individu que TeO catégorise autrement (ni immigré ou descendant d'immigrés du Maghreb, ni
rapatrié ou descendant de rapatriés d'Afrique du nord)
. = hors champ*/

*Je crée enfin la variable colonne du tableau 7.13
generate orig_ind_4 = 0 if (orig_ind==500)
/*G0*/
  replace orig_ind_4 = 1 if (orig_ind==120 | orig_ind==129 | orig_ind==130 | orig_ind==134 |
orig_ind==140 | orig_ind==149 | orig_ind==150 | orig_ind==212 | orig_ind==501) /*G1, immigrés

```

```

et rapatriés*/
  replace orig_ind_4 = 2 if (orig_ind==320 | orig_ind==329 | orig_ind==330 | orig_ind==334 |
  orig_ind==340 | orig_ind==349 | orig_ind==350 | orig_ind==412 | orig_ind==502) /*G2, enfants
d'immigrés et de rapatriés*/
  replace orig_ind_4 = 3 if (orig_ind==620 | orig_ind==629 | orig_ind==630 | orig_ind==634 |
  orig_ind==640 | orig_ind==649 | orig_ind==650 | orig_ind==712 | orig_ind==503) /*G3, petits-
enfants d'immigrés et de rapatriés*/
  replace orig_ind_4 = 4 if (orig_ind==820 |
  orig_ind==840 | orig_ind==849 | orig_ind==504) /*G4*/
  replace orig_ind_4 = 1 if (orig_ind==502 & rapatrié_afn_selon_teo==1) /*les enquêtés
rapatriés sont G1*/
  replace orig_ind_4 = 2 if (orig_ind==502 & rapatrié_afn_selon_teo==2) /*les enquêtés enfants
de rapatriés sont G2*/
  /*tab orig_ind_4 if bon_champ==1 [aw=poids_ind], missing*/ /*sur le champ retenu pondéré, 3
665 individus sont G1, 29 566 sont G2, et 8 977 sont G3*/

*Je peux enfin tracer le tableau 7.13, N = 42 208
tab orig_immigre_rapatrié orig_ind_4 if bon_champ==1 [aw=poids_ind], missing
tab orig_immigre_rapatrié orig_ind_4 if (bon_champ==1 & nicolasg3maghreb==1), missing /*le
champ retenu non pondéré ne comprend que 10 des 11 "Nicolas" G3 Maghreb (car le 11e est enfant
d'un enquêté majoritaire et d'un conjoint d'enquêté G2 Maghreb), et parmi eux, 8/10 sont
probablement descendants de rapatriés d'Afrique du Nord*/

/*Chapitre 7: La reproduction "au plus près" du tableau p. 3, sans les erreurs*/

*Tableau 7.14. Top 20 des prénoms des petits-enfants d'immigrés du Maghreb, selon qu'ils sont
effectivement petits-enfants d'immigrés du Maghreb ou plus probablement petits-enfants de
rapatriés d'Afrique du nord, en France en 2008 (enquête TeO), N = 335 (87 / 248)
  tab orig_ind orig_immigre_rapatrié if (bon_champ==1) [aw=poids_ind], missing
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==620) & (orig_immigre_rapatrié==0
| orig_immigre_rapatrié==1) [aw=poids_ind], order(h) select(20) /*Top 20 pondéré des prénoms
des H G3 Maghreb - total*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==620) & (orig_immigre_rapatrié==0
| orig_immigre_rapatrié==1) [aw=poids_ind], order(h) select(20) /*Top 20 pondéré des prénoms
des F G3 Maghreb - total*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==620) & (orig_immigre_rapatrié==1)
[aw=poids_ind], order(h) select(20) /*Top 20 pondéré des prénoms des H G3 Maghreb
probablement petits-enfants de rapatriés*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==620) & (orig_immigre_rapatrié==1)
[aw=poids_ind], order(h) select(20) /*Top 20 pondéré des prénoms des F G3 Maghreb
probablement petits-enfants de rapatriés*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==620) & (orig_immigre_rapatrié==0)
[aw=poids_ind], order(h) select(20) /*Top 20 pondéré des prénoms des H G3 Maghreb
probablement petits-enfants d'immigrés*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==620) & (orig_immigre_rapatrié==0)
[aw=poids_ind], order(h) select(20) /*Top 20 pondéré des prénoms des F G3 Maghreb
probablement petits-enfants d'immigrés*/

groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==620) & (orig_immigre_rapatrié==0
| orig_immigre_rapatrié==1) , order(h) select(21) /*Top non pondéré des prénoms
des H G3 Maghreb - total*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==620) & (orig_immigre_rapatrié==0
| orig_immigre_rapatrié==1) , order(h) select(22) /*Top non pondéré des prénoms
des F G3 Maghreb - total*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==620) & (orig_immigre_rapatrié==1)
, order(h) select(4) /*Top non pondéré des prénoms des H G3 Maghreb probablement petits-
enfants de rapatriés*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==620) & (orig_immigre_rapatrié==1)
, order(h) select(0) /*Top non pondéré des prénoms des F G3 Maghreb probablement petits-
enfants de rapatriés*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==620) & (orig_immigre_rapatrié==0)
, order(h) select(16) /*Top non pondéré des prénoms des H G3 Maghreb probablement petits-
enfants d'immigrés*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==620) & (orig_immigre_rapatrié==0)
, order(h) select(15) /*Top non pondéré des prénoms des F G3 Maghreb probablement petits-
enfants d'immigrés*/

tab prenom_ind cat1_ined_rec if (prenom_ind=="KARIM" | prenom_ind=="NASSIM" |
prenom_ind=="FARES" | prenom_ind=="NICOLAS" | prenom_ind=="JEREMY" | prenom_ind=="MOURAD" |
prenom_ind=="YANIS" | prenom_ind=="MEHDI" | prenom_ind=="KEVIN" | prenom_ind=="YACINE"
| prenom_ind=="ILYES" | prenom_ind=="RAYAN" | prenom_ind=="BENOIT" |
prenom_ind=="DRISS" | prenom_ind=="MOHAMED" | prenom_ind=="RAYANE" | prenom_ind=="SOFIAN"
| prenom_ind=="ADAM" | prenom_ind=="") /*d'après la
nomenclature de l'INED, 6 prénoms du Top 20 des prénoms des H G3 Maghreb sont d'une origine
autre que "arabo-musulman": "NICOLAS", "JEREMY", "KEVIN", "BENOIT", "ADAM" et "s"*/
tab prenom_ind cat1_ined_rec if (prenom_ind=="SARAH" | prenom_ind=="SOPHIA" |

```



```

prenom_ind=="SABRINA" | prenom_ind=="LINA" | prenom_ind=="INES" | prenom_ind=="SOFIA" |
prenom_ind=="CELIA" | prenom_ind=="JESSICA" | prenom_ind=="LEILA" | prenom_ind=="MYRIAM"
| prenom_ind=="ANAIIS" | prenom_ind=="VIRGINIE" | prenom_ind=="MELISSA" |
prenom_ind=="LILIA" | prenom_ind=="AICHA" | prenom_ind=="LISA" | prenom_ind=="EMMA"
| prenom_ind==" | prenom_ind=="LOLA" | prenom_ind=="") /*d'après la
nomenclature de l'INED, 17 prénoms du Top 20 des prénoms des F G3 Maghreb sont d'une origine
autre que "arabo-musulman": tous sauf "LEILA", "LILIA" et "AICHA"*/
tab prenom_ind cat1_ined_rec if (prenom_ind=="NICOLAS" | prenom_ind=="JEREMY" |
prenom_ind=="ANTONY" | prenom_ind=="ROMAIN" | prenom_ind=="THOMAS" | prenom_ind=="ETHAN" |
prenom_ind=="BENOIT" | prenom_ind=="THEO" | prenom_ind=="MARTIN" | prenom_ind=="YANNICK"
| prenom_ind=="GUILLAUME" | prenom_ind=="OSCAR" | prenom_ind=="AURELIEN" |
prenom_ind=="JONATHAN" | prenom_ind==" | prenom_ind==" | prenom_ind=="
| prenom_ind=="GAETAN" | prenom_ind=="TRISTAN" | prenom_ind=="ELIOTT") /*d'après la
nomenclature de l'INED, 20 prénoms du Top 20 des prénoms des H G3 Maghreb sont d'une origine
autre que "arabo-musulman": tous*/
tab prenom_ind cat1_ined_rec if (prenom_ind==" | prenom_ind=="VIRGINIE" |
prenom_ind=="JESSICA" | prenom_ind=="MARLENE" | prenom_ind=="CLARA" | prenom_ind=="CARLA" |
prenom_ind=="LUDIVINE" | prenom_ind=="LOUNA" | prenom_ind=="CAMILLE" | prenom_ind=="
| prenom_ind=="RACHELLE" | prenom_ind=="MAUD" | prenom_ind=="SABRINA" |
prenom_ind=="KATHLEEN" | prenom_ind==" | prenom_ind==" | prenom_ind=="ALIZEE"
| prenom_ind==" | prenom_ind=="LISA" | prenom_ind=="LOLA") /*d'après la
nomenclature de l'INED, 19 prénoms du Top 20 des prénoms des F G3 Maghreb sont d'une origine
autre que "arabo-musulman": tous sauf "LOUNA"*/
tab prenom_ind cat1_ined_rec if (prenom_ind=="KARIM" | prenom_ind=="NASSIM" |
prenom_ind=="FARES" | prenom_ind=="MOURAD" | prenom_ind=="YANIS" | prenom_ind=="MEHDI" |
prenom_ind=="KEVIN" | prenom_ind=="YACINE" | prenom_ind=="ILYES" | prenom_ind=="RAYAN"
| prenom_ind=="MOHAMED" | prenom_ind=="RAYANE" | prenom_ind=="ADAM" | prenom_ind=="
| prenom_ind==" | prenom_ind=="MARWAN" | prenom_ind=="RYAD" |
prenom_ind=="SOFIAN" | prenom_ind=="DRISS" | prenom_ind=="YOUNES") /*d'après la
nomenclature de l'INED, 3 prénoms du Top 20 des prénoms des H G3 Maghreb sont d'une origine
autre que "arabo-musulman": "KEVIN", "ADAM" et "s"*/
tab prenom_ind cat1_ined_rec if (prenom_ind=="SARAH" | prenom_ind=="SOPHIA" |
prenom_ind=="LINA" | prenom_ind=="SABRINA" | prenom_ind=="INES" | prenom_ind=="SOFIA" |
prenom_ind=="CELIA" | prenom_ind=="LEILA" | prenom_ind=="MYRIAM" | prenom_ind=="ANAIIS"
| prenom_ind=="LILIA" | prenom_ind=="AICHA" | prenom_ind==" |
prenom_ind=="YASMINE" | prenom_ind==" | prenom_ind=="MELISSA" | prenom_ind=="KAINA"
| prenom_ind=="FARAH" | prenom_ind=="JASMINE" | prenom_ind=="MAEVA") /*d'après la
nomenclature de l'INED, 12 prénoms du Top 20 des prénoms des F G3 Maghreb sont d'une origine
autre que "arabo-musulman": "SARAH", "SOPHIA", "LINA", "SABRINA", "INES", "SOFIA", "CELIA",
"MYRIAM", "ANAIIS", "s", "MELISSA" et "MAEVA"*/

```

\*Quelles sont les origines religieuses de ces petits-enfants d'immigrés du Maghreb, selon qu'ils sont effectivement petits-enfants d'immigrés du Maghreb ou plus probablement petits-enfants de rapatriés d'Afrique du Nord?

```

tab orig_relig_ind orig_immigre_rapatrie if bon_champ==1 & orig_ind==620 &
(orig_immigre_rapatrie==0 | orig_immigre_rapatrie==1) [aw=poids_ind], nofreq col missing

```

\*Tableau 7.15. Top 20 des prénoms des habitants originaires du Maghreb selon leur "génération", en France en 2008 (enquête TeO), N = 2 881

```

tab orig_ind orig_immigre_rapatrie if (bon_champ==1) [aw=poids_ind], missing
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==120 & orig_immigre_rapatrie==0)
[aw=poids_ind], order(h) select(20) /*Top 20 pondéré des prénoms des H G1 Maghreb, dissociés des
rapatriés*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==120 & orig_immigre_rapatrie==0)
[aw=poids_ind], order(h) select(20) /*Top 20 pondéré des prénoms des F G1 Maghreb, dissociés des
rapatriés*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==320 & orig_immigre_rapatrie==0)
[aw=poids_ind], order(h) select(20) /*Top 20 pondéré des prénoms des H G2 Maghreb, dissociés des
descendants de rapatriés*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==320 & orig_immigre_rapatrie==0)
[aw=poids_ind], order(h) select(20) /*Top 20 pondéré des prénoms des F G2 Maghreb, dissociés des
descendants de rapatriés*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==620 & orig_immigre_rapatrie==0)
[aw=poids_ind], order(h) select(20) /*Top 20 pondéré des prénoms des H G3 Maghreb, dissociés des
descendants de rapatriés*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==620 & orig_immigre_rapatrie==0)
[aw=poids_ind], order(h) select(20) /*Top 20 pondéré des prénoms des F G3 Maghreb, dissociés des
descendants de rapatriés*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==120 & orig_immigre_rapatrie==0)
, order(h) select(39) /*Top non pondéré des prénoms des H G1 Maghreb, dissociés des
rapatriés*/
groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==120 & orig_immigre_rapatrie==0)
, order(h) select(47) /*Top non pondéré des prénoms des F G1 Maghreb, dissociés des
rapatriés*/
groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==320 & orig_immigre_rapatrie==0)
, order(h) select(102) /*Top non pondéré des prénoms des H G2 Maghreb, dissociés des

```

```

descendants de rapatriés*/
  groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==320 & orig_immigre_rapatrie==0)
, order(h) select(111) /*Top non pondéré des prénoms des F G2 Maghreb, dissociés des
descendants de rapatriés*/
  groups prenom_ind if (bon_champ==1 & sexe_ind==1 & orig_ind==620 & orig_immigre_rapatrie==0)
, order(h) select(17) /*Top non pondéré des prénoms des H G3 Maghreb, dissociés des
descendants de rapatriés*/
  groups prenom_ind if (bon_champ==1 & sexe_ind==2 & orig_ind==620 & orig_immigre_rapatrie==0)
, order(h) select(16) /*Top non pondéré des prénoms des F G3 Maghreb, dissociés des
descendants de rapatriés*/

tab prenom_ind cat1_ined_rec if (prenom_ind=="MOHAMED" | prenom_ind=="AHMED" |
prenom_ind=="RACHID" | prenom_ind=="MOHAMMED" | prenom_ind=="ABDELKADER" |
prenom_ind=="ABDELLAH" | prenom_ind=="KHALID" | prenom_ind=="MOURAD" | prenom_ind=="MUSTAPHA"
| prenom_ind=="ALI" | prenom_ind=="BRAHIM" | prenom_ind=="HAMID" | prenom_ind=="HICHAM" |
prenom_ind=="SAID" | prenom_ind=="OMAR" | prenom_ind=="WALID" | prenom_ind=="NABIL" |
prenom_ind=="DRISS" | prenom_ind=="HASSAN" | prenom_ind=="KARIM") /*d'après la
nomenclature de l'INED, 0 prénom du Top 20 des prénoms des H G1 Maghreb sont d'une origine
autre que "arabo-musulman": ils sont tous "arabo-musulmans"*/
tab prenom_ind cat1_ined_rec if (prenom_ind=="FATIMA" | prenom_ind=="KHADIJA" |
prenom_ind=="SAMIRA" | prenom_ind=="FATIHA" | prenom_ind=="AICHA" |
prenom_ind=="MALIKA" | prenom_ind=="NADIA" | prenom_ind=="ZOHRA" | prenom_ind=="NAIMA"
| prenom_ind=="YASSINE" | prenom_ind=="KARIMA" | prenom_ind=="HOURLIA" | prenom_ind=="RACHIDA" |
prenom_ind=="MERIEM" | prenom_ind=="SAIDA" | prenom_ind=="FATMA" | prenom_ind=="BOUCHRA" |
prenom_ind=="YAMINA" | prenom_ind=="JAMILA" | prenom_ind=="NAJAT") /*d'après la
nomenclature de l'INED, 1 prénom du Top 20 des prénoms des F G1 Maghreb sont d'une origine
autre que "arabo-musulman": "NADIA"*/
tab prenom_ind cat1_ined_rec if (prenom_ind=="MOHAMED" | prenom_ind=="KARIM" |
prenom_ind=="FARID" | prenom_ind=="RACHID" | prenom_ind=="MEHDI" | prenom_ind=="SAMIR"
| prenom_ind=="KAMEL" | prenom_ind=="ADAM" | prenom_ind=="SOFIANE" | prenom_ind=="YANIS"
| prenom_ind=="SAMIA" | prenom_ind=="SAID" | prenom_ind=="AMINE" | prenom_ind=="YOUSSEF"
| prenom_ind=="AHMED" | prenom_ind=="YACINE" | prenom_ind=="NORDINE" | prenom_ind=="SAMY"
| prenom_ind=="HAKIM" | prenom_ind=="ABDELKADER") /*d'après la nomenclature de l'INED, 1
prénom du Top 20 des prénoms des H G2 Maghreb sont d'une origine autre que "arabo-musulman":
"ADAM"*/
tab prenom_ind cat1_ined_rec if (prenom_ind=="NADIA" | prenom_ind=="SARAH" |
prenom_ind=="MYRIAM" | prenom_ind=="FATIMA" | prenom_ind=="YAMINA" | prenom_ind=="INES"
| prenom_ind=="SONIA" | prenom_ind=="MALIKA" | prenom_ind=="SABRINA" | prenom_ind=="NORA"
| prenom_ind=="SAMIRA" | prenom_ind=="ANISSA" | prenom_ind=="LEILA" | prenom_ind=="HOURLIA"
| prenom_ind=="SAMIA" | prenom_ind=="YASMINA" | prenom_ind=="LINA" | prenom_ind=="YASMINE"
| prenom_ind=="AICHA" | prenom_ind=="KARIMA") /*d'après la nomenclature de l'INED, 7
prénoms du Top 20 des prénoms des F G2 Maghreb sont d'une origine autre que "arabo-musulman":
"NADIA", "SARAH", "MYRIAM", "INES", "SONIA", "SABRINA" et "LINA" */
tab prenom_ind cat1_ined_rec if (prenom_ind=="KARIM" | prenom_ind=="NASSIM" |
prenom_ind=="FARES" | prenom_ind=="MOURAD" | prenom_ind=="YANIS" | prenom_ind=="MEHDI"
| prenom_ind=="KEVIN" | prenom_ind=="YACINE" | prenom_ind=="ILYES" | prenom_ind=="RAYAN"
| prenom_ind=="MOHAMED" | prenom_ind=="RAYANE" | prenom_ind=="ADAM" | prenom_ind=="
| prenom_ind=="
| prenom_ind=="MARWAN" | prenom_ind=="RYAD" | prenom_ind=="SOFIAN"
| prenom_ind=="DRISS" | prenom_ind=="YOUNES") /*d'après la nomenclature de l'INED, 3
prénoms du Top 20 des prénoms des H G3 Maghreb sont d'une origine autre que "arabo-musulman":
"KEVIN", "ADAM" et "s"*/
tab prenom_ind cat1_ined_rec if (prenom_ind=="SARAH" | prenom_ind=="SOPHIA" |
prenom_ind=="LINA" | prenom_ind=="SABRINA" | prenom_ind=="INES" | prenom_ind=="SOFIA"
| prenom_ind=="CELIA" | prenom_ind=="LEILA" | prenom_ind=="MYRIAM" | prenom_ind=="ANAISS"
| prenom_ind=="LILIA" | prenom_ind=="AICHA" | prenom_ind=="
| prenom_ind=="YASMINE"
| prenom_ind=="
| prenom_ind=="MELISSA" | prenom_ind=="KAINA" | prenom_ind=="FARAH"
| prenom_ind=="JASMINE" | prenom_ind=="MAEVA") /*d'après la nomenclature de l'INED, 12
prénoms du Top 20 des prénoms des F G3 Maghreb sont d'une origine autre que "arabo-musulman":
"SARAH", "SOPHIA", "LINA", "SABRINA", "INES", "SOFIA", "CELIA", "MYRIAM", "ANAISS", "s",
"MELISSA" et "MAEVA"*/

```

\*Je re-crée la variable d'origine (géographique) des individus, corrigée de sa confusion entre immigrés (et descendants d'immigrés) du Maghreb et certains rapatriés (et descendants de rapatriés) d'Afrique du nord

```

generate orig_ind_corrigeepiedsnoirs = orig_ind
replace orig_ind_corrigeepiedsnoirs = . if (orig_ind==120 & orig_immigre_rapatrie==1)
replace orig_ind_corrigeepiedsnoirs = . if (orig_ind==320 & orig_immigre_rapatrie==1)
replace orig_ind_corrigeepiedsnoirs = . if (orig_ind==620 & orig_immigre_rapatrie==1)
replace orig_ind_corrigeepiedsnoirs = . if (orig_ind==820 & orig_immigre_rapatrie==1)
replace orig_ind_corrigeepiedsnoirs = . if (orig_immigre_rapatrie==.)
tab orig_ind orig_ind_corrigeepiedsnoirs if bon_champ==1 [aw=poids_ind], missing
*J'applique les mêmes labels à la variable ORIG_IND_CORRIGEE_PIEDSNOIRS qu'à la variable
ORIG_IND
label variable orig_ind_corrigeepiedsnoirs "origine"
label values orig_ind_corrigeepiedsnoirs orig

```

\*Je crée la variable ORIG\_IND\_CORRIGEE\_PIEDSNOIRS synthétique

```

generate orig_ind_corrige_piedsnoirs_s = 1 if (orig_ind_corrige_piedsnoirs==500 |
orig_ind_corrige_piedsnoirs==501 | orig_ind_corrige_piedsnoirs==502 |
orig_ind_corrige_piedsnoirs==503 | orig_ind_corrige_piedsnoirs==504) /*groupe
"majoritaire"*/
  replace orig_ind_corrige_piedsnoirs_s = 2 if (orig_ind_corrige_piedsnoirs==212 |
orig_ind_corrige_piedsnoirs==412 | orig_ind_corrige_piedsnoirs==712)
/*DOM*/
  replace orig_ind_corrige_piedsnoirs_s = 3 if (orig_ind_corrige_piedsnoirs==140 |
orig_ind_corrige_piedsnoirs==340 | orig_ind_corrige_piedsnoirs==640 |
orig_ind_corrige_piedsnoirs==840) /*Europe du Sud*/
  replace orig_ind_corrige_piedsnoirs_s = 4 if (orig_ind_corrige_piedsnoirs==149 |
orig_ind_corrige_piedsnoirs==349 | orig_ind_corrige_piedsnoirs==649 |
orig_ind_corrige_piedsnoirs==849) /*autre Europe*/
  replace orig_ind_corrige_piedsnoirs_s = 5 if (orig_ind_corrige_piedsnoirs==120 |
orig_ind_corrige_piedsnoirs==320 | orig_ind_corrige_piedsnoirs==620 |
orig_ind_corrige_piedsnoirs==820) /*Maghreb*/
  replace orig_ind_corrige_piedsnoirs_s = 6 if (orig_ind_corrige_piedsnoirs==129 |
orig_ind_corrige_piedsnoirs==329 | orig_ind_corrige_piedsnoirs==629)
/*autre Afrique*/
  replace orig_ind_corrige_piedsnoirs_s = 7 if (orig_ind_corrige_piedsnoirs==134 |
orig_ind_corrige_piedsnoirs==334 | orig_ind_corrige_piedsnoirs==634)
/*Turquie*/
  replace orig_ind_corrige_piedsnoirs_s = 8 if (orig_ind_corrige_piedsnoirs==130 |
orig_ind_corrige_piedsnoirs==330 | orig_ind_corrige_piedsnoirs==630)
/*Asie*/
  replace orig_ind_corrige_piedsnoirs_s = 9 if (orig_ind_corrige_piedsnoirs==150 |
orig_ind_corrige_piedsnoirs==350 | orig_ind_corrige_piedsnoirs==650)
/*Ailleurs dans le monde*/
*J'applique des labels à la variable ORIG_IND_CORRIGE_PIEDSNOIRS_SYNTHETIQUE
label variable orig_ind_corrige_piedsnoirs_s "origine synthétique"
  label values orig_ind_corrige_piedsnoirs_s orig_synth

*Je recalcule les Tableaux 6.13a-b, corrigés de leur confusion entre immigrés (et descendants
d'immigrés) du Maghreb et certains rapatriés (et descendants de rapatriés) d'Afrique du nord,
N = 52 903
tab orig_ind_corrige_piedsnoirs_s cat1_ined_rec if bon_champ==1 [aw=poids_ind], cell nofreq
missing /*en métropole en 2008, si l'on sépare les rapatriés et descendants de rapatriés
d'Afrique du nord des immigrés et descendants d'immigrés du Maghreb, alors 5 % (et non pas 10
%) des immigrés et descendants d'immigrés du Maghreb portent un prénom "français", et 80 % (et
non pas 74 %) portent un prénom "arabo-musulman"*/

/*Chapitre 7: Résumé*/
order ident noi poids_ind sexe_ind anais_ind prenom_ind orig_prenom_ind cat1_ined_rec
lienego_ind regionnais_ind residence_ind group2_ind regionnaispere_ind regionnaismere_ind
orig_ind_corrige_piedsnoirs autrepere natpniv2 natmniv2 n_indper n_indmer lnp1_gr
lnp2_gr lnm1_gr lnm2_gr relig_ind r_impvie orig_relig_ind x_vufri
sort ident noi

save repro, replace

*****
/*CHAPITRE 8. Essai de reproduction du second résultat spectaculaire:
le "23 %"*/
*****

/*Chapitre 8: La reproduction "au plus près" de la figure 2, page 3*/

*Tableaux 8.1. Proportion de prénoms spécifiques par origine et "génération", en France en
2008 (enquête TeO)

*Tableau 8.1a: champ retenu, N = 42 208
tab orig_ind if (bon_champ==1) [aw=poids_ind], missing
tab orig_ind cat1_ined_rec if (bon_champ==1)
[aw=poids_ind], row nofreq /*données pondérées*/
tab orig_ind if (bon_champ==1) , missing
tab orig_ind cat1_ined_rec if (bon_champ==1)
, row nofreq /*données non pondérées*/
tab orig_ind_corrige_piedsnoirs if (bon_champ==1) [aw=poids_ind], missing
tab orig_ind_corrige_piedsnoirs cat1_ined_rec if (bon_champ==1)
[aw=poids_ind], row nofreq /*données pondérées sans pieds-noirs*/
tab orig_ind_corrige_piedsnoirs if (bon_champ==1) , missing
tab orig_ind_corrige_piedsnoirs cat1_ined_rec if (bon_champ==1)
, row nofreq /*données non pondérées sans pieds-noirs*/

```

```

*Tableau 8.1b: champ "maximal", N = 83 045
tab orig_ind if (lienego_ind==0 | lienego_ind==2 | lienego_ind==201) [aw=poids_ind], missing
tab orig_ind cat1_ined_rec
[aw=poids_ind], row nofreq /*données pondérées*/
tab orig_ind , missing
tab orig_ind cat1_ined_rec
, row nofreq /*données non pondérées*/

*Tableau 8.1c: champ "enquêtés et tous leurs enfants vivants", N = 52 902
tab orig_ind if (lienego_ind==0 | lienego_ind==2 | lienego_ind==201) [aw=poids_ind], missing
tab orig_ind cat1_ined_rec if (lienego_ind==0 | lienego_ind==2 |
lienego_ind==201) [aw=poids_ind], row nofreq /*données pondérées*/
tab orig_ind if (lienego_ind==0 | lienego_ind==2 | lienego_ind==201) , missing
tab orig_ind cat1_ined_rec if (lienego_ind==0 | lienego_ind==2 |
lienego_ind==201) , row nofreq /*données non pondérées*/

/*Chapitre 8: Analyses complémentaires: proportion de prénoms spécifiques par origine et par
génération*/

*Tableau 8.2a: Proportion de prénoms spécifiques par origine et "génération" en France en 2008
(enquête TeO), N = 42 208
*Tableau 8.2b: Proportion de prénoms "français" par origine et "génération" en France en 2008
(enquête TeO), N = 42 208
*Tableau 8.2c: Proportion de prénoms "internationaux ou anglophones" par origine et
"génération" en France en 2008 (enquête TeO), N = 42 208
tab orig_ind_corrige_piedsnoirs cat1_ined_rec if (bon_champ==1) [aw=poids_ind], cell nofreq
/*données pondérées sans pieds-noirs*/

/*Chapitre 8: Résumé*/

*****
/*CHAPITRE 9. Essai de reproduction des autres résultats*/
*****

/*Chapitre 9: Reproduction "au plus près" de la figure 1 (page 2)*/

*Figure 9.1b. Proportion de nouveau-nés recevant un prénom "français" par année de naissance -
France, population majoritaire (%), N = 32 185
tab orig_ind_corrige_piedsnoirs cat1_ined_rec if (bon_champ==1 & (regionnais_ind==11 |
regionnais_ind==12)) [aw=poids_ind], missing /*32 185*/
tab orig_ind_corrige_piedsnoirs cat1_ined_rec if (bon_champ==1 & (regionnais_ind==11 |
regionnais_ind==12) & quinquennais_ind==1945) [aw=poids_ind], cell nofreq missing /*1945-
1949*/
tab orig_ind_corrige_piedsnoirs cat1_ined_rec if (bon_champ==1 & (regionnais_ind==11 |
regionnais_ind==12) & quinquennais_ind==1950) [aw=poids_ind], cell nofreq missing /*1950-
1954*/
tab orig_ind_corrige_piedsnoirs cat1_ined_rec if (bon_champ==1 & (regionnais_ind==11 |
regionnais_ind==12) & quinquennais_ind==1955) [aw=poids_ind], cell nofreq missing /*1955-
1959*/
tab orig_ind_corrige_piedsnoirs cat1_ined_rec if (bon_champ==1 & (regionnais_ind==11 |
regionnais_ind==12) & quinquennais_ind==1960) [aw=poids_ind], cell nofreq missing /*1960-
1964*/
tab orig_ind_corrige_piedsnoirs cat1_ined_rec if (bon_champ==1 & (regionnais_ind==11 |
regionnais_ind==12) & quinquennais_ind==1965) [aw=poids_ind], cell nofreq missing /*1965-
1969*/
tab orig_ind_corrige_piedsnoirs cat1_ined_rec if (bon_champ==1 & (regionnais_ind==11 |
regionnais_ind==12) & quinquennais_ind==1970) [aw=poids_ind], cell nofreq missing /*1970-
1974*/
tab orig_ind_corrige_piedsnoirs cat1_ined_rec if (bon_champ==1 & (regionnais_ind==11 |
regionnais_ind==12) & quinquennais_ind==1975) [aw=poids_ind], cell nofreq missing /*1975-
1979*/
tab orig_ind_corrige_piedsnoirs cat1_ined_rec if (bon_champ==1 & (regionnais_ind==11 |
regionnais_ind==12) & quinquennais_ind==1980) [aw=poids_ind], cell nofreq missing /*1980-
1984*/
tab orig_ind_corrige_piedsnoirs cat1_ined_rec if (bon_champ==1 & (regionnais_ind==11 |
regionnais_ind==12) & quinquennais_ind==1985) [aw=poids_ind], cell nofreq missing /*1985-
1989*/
tab orig_ind_corrige_piedsnoirs cat1_ined_rec if (bon_champ==1 & (regionnais_ind==11 |
regionnais_ind==12) & quinquennais_ind==1990) [aw=poids_ind], cell nofreq missing /*1990-
1994*/
tab orig_ind_corrige_piedsnoirs cat1_ined_rec if (bon_champ==1 & (regionnais_ind==11 |

```

```

regionnais_ind==12) & quinquennais_ind==1995) [aw=poids_ind], cell nofreq missing /*1995-
1999*/
tab orig_ind_corrigee_piedsnoirs cat1_ined_rec if (bon_champ==1 & (regionnais_ind==11 |
regionnais_ind==12) & quinquennais_ind==2000) [aw=poids_ind], cell nofreq missing /*2000-
2004*/
tab orig_ind_corrigee_piedsnoirs cat1_ined_rec if (bon_champ==1 & (regionnais_ind==11 |
regionnais_ind==12) & quinquennais_ind==2005) [aw=poids_ind], cell nofreq missing /*2005-
2009*/

/*Chapitre 9: Analyses complémentaires autour de la figure 1 (page 2)*/

*Figure 9.2. Proportion de nouveau-nés recevant un prénom "français" par année de naissance,
dans le groupe "majoritaire" et dans le groupe non "majoritaire", en France en 2008 (enquête
TeO), N = 32 185
tab orig_ind_corrigee_piedsnoirs cat1_ined_rec if (bon_champ==1 & (regionnais_ind==11 |
regionnais_ind==12)) [aw=poids_ind], missing /*32 185*/

/*Chapitre 9: Reproduction "au plus près" de la figure 3 (page 4)*/

*Je commence par associer le degré de religiosité de l'enquêté à tous les membres de son
entourage familial
egen r_impvie_tous = max(r_impvie), by(ident)
label values r_impvie_tous religiosite
tab r_impvie_tous, missing

*Figure 9.3x [non publié]. Proportion de prénoms arabo-musulmans chez les petits-enfants
d'immigrés du Maghreb selon la religiosité du parent, les petits-enfants de pieds-noirs étant
conservés parmi les petits-enfants d'immigrés du Maghreb, N = 1 490
tab r_impvie_tous cat1_ined_rec if (bon_champ==1 & orig_ind==620) [aw=poids_ind], cell nofreq
missing
*Sur ces 1 490 petits-enfants d'immigrés maghrébins (les petits-enfants de pieds-noirs étant
conservés parmi les petits-enfants d'immigrés du Maghreb), combien sont d'origine musulmane,
chrétienne et sans origine religieuse?
tab orig_relig_ind if (bon_champ==1 & orig_ind==620) [aw=poids_ind],
missing

*Figure 9.3b. Proportion de prénoms arabo-musulmans chez les petits-enfants
d'immigrés du Maghreb selon la religiosité du parent, les petits-enfants de pieds-noirs étant
exclus, N = 1 111
tab r_impvie_tous cat1_ined_rec if (bon_champ==1 & orig_ind_corrigee_piedsnoirs==620)
[aw=poids_ind], cell nofreq missing
*Sur ces 1 111 petits-enfants d'immigrés maghrébins (les petits-enfants de pieds-noirs étant
exclus), combien sont d'origine musulmane, chrétienne et sans origine religieuse?
tab orig_relig_ind if (bon_champ==1 & orig_ind_corrigee_piedsnoirs==620)
[aw=poids_ind], missing

/*Chapitre 9: Analyses complémentaires autour de la figure 3 (page 4)*/

*Figure 9.3y [non publié]. Proportion de prénoms latins chez les petits-enfants d'immigrés
d'Europe du Sud selon la religiosité du parent, N = 2 626
tab r_impvie_tous cat1_ined_rec if (bon_champ==1 & orig_ind_corrigee_piedsnoirs==640)
[aw=poids_ind], cell nofreq missing
*Sur ces 2 626 petits-enfants d'immigrés d'Europe du Sud, combien sont d'origine musulmane,
chrétienne et sans origine religieuse?
tab orig_relig_ind if (bon_champ==1 & orig_ind_corrigee_piedsnoirs==640)
[aw=poids_ind], missing

*Figure 9.3c. Proportion de prénoms spécifiques chez les petits-enfants d'immigrés du Maghreb
et d'Europe du Sud, selon la religiosité du parent, en France en 2008 (enquête TeO)
*Figure qui réunit les Figures 9.3b et 9.3y

/*Chapitre 9: Résumé*/

*****
/*CHAPITRE 10. Une seule option: la rétractation*/
*****

*****
/*CHAPITRE 11. Erreurs de bonne foi ou fraude scientifique?*/

```

\*\*\*\*\*

/\*Chapitre 11: Note de bas de page sur affiliation religieuse et religiosité\*/

\*Proportion d'enquêtés pour qui la religion a "beaucoup d'importance" dans leur vie, selon la religion et l'année de naissance, N = 21 232

tab relig\_ind r\_impvie if (lienego\_ind==0 & decennais\_ind==1950) [aw=poids\_ind], cell nofreq missing /\*années 1950\*/

tab relig\_ind r\_impvie if (lienego\_ind==0 & decennais\_ind==1960) [aw=poids\_ind], cell nofreq missing /\*années 1960\*/

tab relig\_ind r\_impvie if (lienego\_ind==0 & decennais\_ind==1970) [aw=poids\_ind], cell nofreq missing /\*années 1970\*/

tab relig\_ind r\_impvie if (lienego\_ind==0 & decennais\_ind==1980) [aw=poids\_ind], cell nofreq missing /\*années 1980\*/

tab relig\_ind r\_impvie if (lienego\_ind==0 & decennais\_ind==1990) [aw=poids\_ind], cell nofreq missing /\*années 1990\*/

\*\*\*\*\*

/\*ANNEXE 4B. Les programmes informatiques communiqués par les auteurs\*/

\*\*\*\*\*

/\*

/\*Je prépare une table contenant les variables CAT0 et CAT1 sur le champ maximal\*/

cd "C:\Users\Public\Documents"

use repro, clear

duplicates report prenom\_ind /\*il y a 65 046 doublons en surplus\*/

duplicates drop prenom\_ind, force /\*ces 65 046 doublons en surplus sont éliminés\*/

keep prenom\_ind cat0 cat1

label drop \_all

sort prenom\_ind

rename prenom\_ind prenom

save "C:\Users\Public\Documents\les 3 programmes ined\tables SAS et autres\cat" /\*cette table CAT, je l'enregistre dans le dossier "les 3 programmes ined"\*/

/\*Je convertis cette table en SAS\*/

cd "C:\Users\Public\Documents\les 3 programmes ined\tables SAS et autres"

export sasxport cat, vallabfile(none)

/\*Ensuite, sous SAS, j'importe les variables CAT0 et CAT1 de la table SAS repro aux tables SAS indiv2\_prenom, enfants2\_prenom et thl2\_prenom\*/

/\*J'ai désormais les 3 tables au bon format pour faire tourner les 3 programmes R des auteurs et de l'INED\*/

\*/

\*\*\*\*\*

\*\*\*\*\*

/\*\*\*\*\*\* FIN \*\*\*\*\*\*/

\*\*\*\*\*

\*\*\*\*\*

/\*

log close

\*/

## Annexe 6A : Une histoire comparée des prénoms en Europe et au Maghreb

Cette Annexe 6A, associée au Chapitre 6, p. 25, comprend une revue de littérature sur l'histoire comparée des prénoms en Europe et au Maghreb.<sup>88</sup> L'histoire des prénoms n'est pas seulement une formidable source d'informations sur les langues, les religions, l'organisation sociale et bien d'autres aspects des populations humaines du monde actuel comme du passé : elle permet aussi de donner une signification aussi précise que possible aux catégories de prénoms distinguées par Coulmont et Simon (2019).

### ❖ L'onomastique

Si les noms communs ont tous une signification (un « contenu sémantique », comme disent les linguistes), tel n'est pas le cas de tous les noms propres. C'est pourquoi une discipline, l'onomastique,<sup>89</sup> étudie spécifiquement les noms propres (Eichler et al. 1995 ; Nicolaisen 2015 ; Hough 2016a), et plus particulièrement :

- les noms de lieux (toponomastique ou toponymie), comme les noms de villes, fleuves et rivières,<sup>90</sup> montagnes, rues, etc. ;
- les noms de personnes (anthroponomastique ou anthroponymie), comme les prénoms et noms de famille des individus, et les noms des groupes (ethnonymes) ;
- les noms de personnages littéraires et divers autres noms, comme les noms d'animaux domestiques,<sup>91</sup> les noms de marque, de bateaux, de trains, d'avions, de planètes, d'étoiles et de constellations ; et enfin
- l'usage social des noms (socio-onomastique).

### ❖ La structure du nom personnel

De nos jours dans les pays occidentaux, le nom personnel est composé de 2 éléments : d'une part un prénom (ou plusieurs), et d'autre part un nom de famille héréditaire, généralement transmis par le père à tous ses enfants. Or non seulement il n'en a pas toujours été ainsi en Europe, mais surtout il n'en va pas de même partout ailleurs — très loin de là. En schématisant, on peut considérer qu'il existe dans le monde 4 grands types de systèmes de nom personnel (Hanks et Parkin 2016, p. 214-215) :

- i. **système à 1 seul nom** : dans l'Europe du haut Moyen Âge (5<sup>e</sup>-10<sup>e</sup> siècles) et dans certaines populations jusqu'à nos jours, comme les Birmans (ou dans une certaine

---

<sup>88</sup> Les thèses soutenues dans le présent document n'engagent personne d'autre que moi et je suis seul responsable de toute erreur. Je tiens toutefois à remercier chaleureusement des proches et plusieurs chercheurs de m'avoir aidé dans la revue de littérature présentée dans cette Annexe 6A :

- sur l'Europe : Michaël Gasperoni, historien spécialiste d'onomastique italienne (et juive) ;
- sur le Maghreb et le monde arabe : Corinne Gomez-Le Chevanton, historienne spécialiste des Européens en Algérie coloniale ; Manel Khelifa, arabophone ; Mohamed Lahrouchi, linguiste spécialiste des langues berbères ; Pierre Larcher, linguiste spécialiste de l'arabe ;
- sur le monde turc : Olivier Bouquet, historien spécialiste d'onomastique ottomane ; Élise Massicard, politologue spécialiste de la Turquie contemporaine ; Özgür Türesay, historien de l'empire ottoman ;
- sur le monde juif : Alexandre Beider, linguiste spécialiste d'onomastique juive.

<sup>89</sup> Si elle est relativement méconnue, même par les chercheurs en sciences sociales, l'onomastique est une discipline très riche, qui permet notamment d'étudier la plupart des domaines de la linguistique : phonologie, morphologie, syntaxe, sémantique, sociolinguistique et psycholinguistique (Eichler et al. 1995 ; Zgusta 1995 ; MacKenzie 2018). Sur l'onomastique en France, voir Fabre (1995) et Mulon (1995).

<sup>90</sup> L'étude des noms de fleuves et rivières a permis de produire des connaissances précieuses, notamment en préhistoire de l'Europe : « A considerable number of river names in western and central Europe show remarkable similarity (e.g., Esera in Spain, Isère in France, Yser in Belgium, Isar in Bavaria, Jizera in the Czech Republic) and are the only evidence of a pre-Celtic Indo-European population of those regions » (Zgusta 2014).

<sup>91</sup> Pour d'intéressants développements sur les noms propres d'animaux dans la tradition arabe, voir Larcher 2013, p. 309-310.

mesure les Thaïs), le nom personnel se réduit à un seul nom attribué à la naissance, qu'on peut considérer comme l'équivalent, pour nous, d'un prénom ;

- ii. **système à 1 prénom et 1 nom de famille héréditaire** : dans la plus grande partie de l'Europe depuis environ le 15<sup>e</sup> siècle, comme en Chine depuis plusieurs siècles *avant* l'ère commune (Creamer 1995), le nom personnel est composé d'un prénom attribué à la naissance et d'un nom de famille héréditaire<sup>92</sup> ; le prénom précède le nom de famille dans la plus grande partie de l'Europe et plus généralement dans le monde indo-européen (de l'Irlande au Bangladesh), tandis que le nom de famille précède le prénom en Chine et plus généralement en Asie orientale (Corée, Japon, Vietnam) mais aussi en Hongrie ;
- iii. **système à 1 prénom et 1 patronyme non héréditaire** : chez les Islandais et les Malais jusqu'à nos jours, le nom personnel est composé d'un prénom attribué à la naissance, et d'un « patronyme » non héréditaire, c'est-à-dire d'un nom dérivé du prénom du père et qui est transmis à ses enfants mais qu'eux-mêmes ne transmettront pas à leurs enfants (puisque leurs enfants porteront, à leur tour, un nom dérivé du prénom de leur propre père)<sup>93</sup> ;
- iv. **système comprenant jusqu'à 5 noms** : dans le monde arabe (de la Mauritanie à l'Irak), le nom personnel était jusqu'au 19<sup>e</sup> siècle composé de jusqu'à 5 éléments (Beeston 1971 ; Grangaud et Michel 2010 ; Larcher 2013), parfois abrégés :
  - le prénom (*ism*), composé d'un nom commun, d'un adjectif ou d'un verbe de la langue arabe ;
  - le « teknonyme » (Larcher 2013, p. 306) ou indication de paternité ou de maternité (*kunya*), composé de Abu [père de] + *ism* du 1<sup>er</sup> fils ou autre terme (pour les hommes), ou de Umm [mère de] + *ism* du 1<sup>er</sup> fils ou autre terme (pour les femmes) ;
  - le surnom ou nom honorifique (*laqab*), attribué au cours de la vie ;
  - la chaîne de « patronymes » (*nasab*), composée de ibn/ben [fils de] + *ism* du père (pour les hommes), ou de bint [fille de] + *ism* du père (pour les femmes), éventuellement jusqu'au nom du légendaire fondateur de la tribu, ou jusqu'à l'origine géographique ou professionnelle ou l'affiliation à une école juridique (*madhhab*)<sup>94</sup> ;
  - le nom de clan ou de tribu ou de ville ou province d'origine (*nisba*), terminant en -i ou -î.

Par conséquent, les prénoms maghrébins et plus généralement arabes s'inscrivaient traditionnellement (d'au moins le 7<sup>e</sup> siècle jusqu'au 19<sup>e</sup> siècle) dans un système onomastique beaucoup plus complexe que celui des sociétés française et européennes (depuis les environs du 12<sup>e</sup> siècle). Et pour cause : au moins jusqu'à la colonisation, le système onomastique arabe était, dans le monde, celui qui combinait le plus d'éléments.

---

<sup>92</sup> En Europe, la Péninsule Ibérique pratique une variante de ce système, puisque le nom personnel y est composé d'un prénom et de 2 noms de famille héréditaires : prénom + 1<sup>er</sup> nom de famille du père + 1<sup>er</sup> nom de famille de la mère en Espagne depuis les environs du 16<sup>e</sup> siècle (et dans les anciennes colonies espagnoles) ; et prénom + 1<sup>er</sup> nom de famille de la mère + 1<sup>er</sup> nom de famille du père au Portugal (et au Brésil) (Mateos et Tucker 2008).

<sup>93</sup> Ainsi en Islande, le patronyme est constitué du prénom du père auquel on ajoute -son pour les garçons et -dóttir pour les filles ; le fils et la fille d'un homme prénommé Sven portent respectivement les patronymes Svensson et Svendsdóttir. En Malaisie, le patronyme est constitué de bin + prénom du père (pour les hommes) ou de binte + prénom du père (pour les femmes) (Jones et Phillips 1995). Le monde chrétien orthodoxe (Russie, Belarus, Ukraine, Bulgarie, Grèce) pratique une variante de ce système, puisque le nom personnel y est composé d'un prénom et d'un patronyme non héréditaire, auxquels s'ajoute aussi un nom de famille héréditaire.

<sup>94</sup> Dans beaucoup de pays arabes, un individu peut aussi être identifié par son prénom suivi de celui de son père et de celui de son grand-père paternel, sans qu'il n'y ait plus de ibn/ben ou de bint entre ces noms.



En Europe, depuis les environs du 12<sup>e</sup> au 15<sup>e</sup> siècles, le nom personnel de la plupart des habitants est composé de 2 éléments : un prénom attribué à la naissance, et un nom de famille héréditaire (Lawson 2016, p. 176).<sup>95</sup> En effet la christianisation des prénoms à partir du haut Moyen Âge avait beaucoup réduit la taille du stock de prénoms, c'est-à-dire le nombre de prénoms différents qui sont attribués au moins une fois. Et dans un contexte d'essor de la population et d'urbanisation à partir des environs de l'an 1000, les doublons se sont multipliés, conduisant à l'addition progressive du surnom, lui-même progressivement devenu nom de famille héréditaire pour mettre en avant une qualification professionnelle ou divers droits politiques (Lawson 2016, p. 172-173 ; Brylla 2016, p. 237 ; Saarelma 2009, p. 202).<sup>96</sup>

Par contraste, au Maghreb, le nom personnel des Berbères comporte, avant le 7<sup>e</sup> siècle, 3 éléments : un prénom attribué à la naissance ; un patronyme composé de Ou [fils de] et du prénom du père (voire du prénom du père du père, etc.) ; et une désignation ethnique souvent précédée de Aït [ceux de] (Yermèche 2013, p. 48-50). Puis à partir du 7<sup>e</sup> siècle, l'islamisation du Maghreb arabise la structure du nom personnel des habitants (Yermèche 2013, p. 54-55). Les Berbères non arabisés conservent leur système à 3 éléments (aussi contenus dans les noms arabes), mais les Berbères arabisés adoptent les dénominations arabes : les particules berbères Ou [fils de] et Aït [descendants de] sont remplacées par les particules filiales arabes Ben et Ould [fils de] et par les particules ethniques arabes Beni et Ouled [enfants de]. En outre, du 7<sup>e</sup> au 19<sup>e</sup> siècles, certaines couches de lettrés du Maghreb adoptent les 5 éléments du nom personnel arabe. Comme le résume un spécialiste de l'Algérie coloniale, jusqu'à l'organisation de l'état civil des indigènes algériens de confession musulmane (1882),

« les indigènes algériens avaient un système de désignation et d'identification des individus totalement différent de celui utilisé par les Français, fondé sur le prénom et le nom. Le système en vigueur chez les indigènes algériens était beaucoup plus complexe. Ainsi chaque individu avait une dénomination qui comprenait un nom propre ou prénom (*El ism*), suivi parfois par l'indication de paternité ou de maternité, la *kunya* (*bou* ou *oum*), lui-même suivi par le *laqab* (surnom souvent en relation avec la profession qui s'est transformée à la longue en nom) auxquels s'ajoutent le prénom du père précédé d'un *ben* et enfin du nom de la tribu ou du lieu d'origine. Ainsi l'identité du chef de l'insurrection dans le sud-ouest algérien était : *Mohamed* (prénom) *Ben El Arbi* (fils de El Arbi) *Bou Ammama* (surnom) des *Ouled Sidi Cheikh* (tribu) » (Kateb 2001, p. 110-111).

Ce n'est qu'avec la colonisation de l'Afrique du Nord par la France, et précisément avec l'établissement de l'état civil des indigènes d'Algérie (1882), que les habitants passent d'un système de nomination arabo-berbère lignager (comprenant jusqu'à 5 éléments) à un système européen individuel (comprenant un prénom et un nom de famille), les noms de tribus disparaissant dans le nouveau système de nomination (Yermèche 2013, p. 59). Depuis les indépendances (étalées de 1956 à 1962), les pays du Maghreb ont tous trois conservé l'état civil hérité de la colonisation, y compris la structure du nom personnel utilisée en France et en Occident. Mais comme cette pratique ne correspondait pas à l'usage local, certains habitants ignoraient jusqu'à récemment quel pouvait bien être leur « nom de famille ».

<sup>95</sup> Sur la « genèse médiévale de l'anthroponymie moderne » en France, voir la somme de Bourin 1990, Bourin et Chareille 1992 et 1995, Beck 1997, Bourin et Chareille 2002 et Chareille 2008.

<sup>96</sup> En Europe, la plupart des noms de famille (héréditaires), qui datent des environs du 12<sup>e</sup> siècle, proviennent de patronymes (non héréditaires), de noms de lieux, de noms de métiers et de surnoms (sobriquets). Selon certaines sources, parmi les noms de famille portés en France aujourd'hui, la plupart seraient originaires de l'ancien français (environ 34 % des noms de famille) ou d'origine latine (30 %) ou germanique (21 %), loin devant les noms d'origine gauloise (3 %), hébraïque (3 %), grecque (2 %), flamande (2 %), basque (1 %), etc. Au Maghreb, la plupart des noms de famille enregistrés à l'état civil ont été inventés suite à l'introduction de l'état civil, que ces noms de famille soient issus de la *kunya*, du *laqab*, du *nasab* ou de la *nisba*.

## ❖ Le lexique des prénoms

Si toutes les sociétés humaines ne connaissent pas de nom de famille, loin de là, il semble que toutes connaissent le prénom (Alford 1988). On admet généralement que ce prénom a deux fonctions : faire référence à un individu plutôt qu'un autre (fonction dénotative), et associer son porteur à d'autres individus (fonction connotative) (Coulmont 2017a, p. 1). C'est ce qui permet d'inférer du prénom d'un individu son sexe (presque toujours<sup>97</sup>), l'origine géographique, linguistique, religieuse ou sociale de sa famille (souvent), sa position dans sa fratrie et les circonstances de sa naissance (parfois), son âge (depuis le 20<sup>e</sup> siècle), certains goûts esthétiques (Boyer 2016, p. 12) et plus généralement diverses caractéristiques stéréotypiques des porteurs des divers prénoms (Lawson 1995).

En Europe depuis les environs du 12<sup>e</sup> siècle, le prénom de la plupart des habitants est un prénom chrétien, c'est-à-dire un prénom du Nouveau Testament ou le prénom d'un saint ou martyr chrétien (Lawson 2016, p. 176), la plupart des prénoms de l'Ancien Testament étant en revanche abandonnés (Leibring 2016, p. 206). Parfois, le prénom chrétien attribué au nouveau-né correspond à celui du saint du jour de sa naissance ou du saint patron de son lieu de naissance (région, ville, quartier, etc.) (Saarelma 2009, p. 201). C'est aussi depuis les environs du 12<sup>e</sup> siècle que le prénom féminin « Maria » devient particulièrement fréquent en Europe du Sud, puis en Allemagne et en Scandinavie (Leibring 2016, p. 206). C'est depuis les environs de cette période aussi que le prénom latin « Iohannes » a donné les différents cognats « Jean » en français, « Giovanni » en italien, « Juan » en espagnol et « João » en portugais, ainsi que « John » en anglais, « Hans » en allemand, etc. (Leibring 2016, p. 204). En France, le stock de prénoms, initialement à dominante germanique, est ainsi remplacé du 12<sup>e</sup> au 15<sup>e</sup> siècles par un stock de prénoms du Nouveau Testament et de saints et martyrs chrétiens (Bozon 1987, p. 84-85), ces prénoms chrétiens pouvant être d'origine hébraïque, grecque ou latine. Par exemple, le prénom français « Jean » provient du vieux français « Jehan », qui provient comme on vient de le voir du latin « Iohannes » ou « Ioannes », mais ce prénom latin provient lui-même du grec « Ἰωάννης », qui provient lui-même de l'hébreu « Yohanan » (יְהוֹנָן) ou « Yehohanan » (יְהוֹחָנָן), qui en hébreu classique signifie « Dieu est bon ».

Au Maghreb à partir du 7<sup>e</sup> siècle, les prénoms berbères sont remplacés par des prénoms arabes (Yermèche 2013, p. 54-55). Ainsi les membres des dynasties berbères qui ont régné sur l'Afrique du Nord (Almoravides du 11<sup>e</sup> au 12<sup>e</sup> siècles, Almohades du 12<sup>e</sup> au 13<sup>e</sup> siècles, etc.) portaient-ils déjà des prénoms arabes.<sup>98</sup> Par conséquent, au Maghreb depuis plusieurs siècles,

---

<sup>97</sup> Les prénoms « épiciènes », qui ne permettent pas d'inférer le sexe du porteur du prénom, semblent assez rares dans la plupart des langues. En français on compte « Claude », « Dominique », « Camille », « Ange » ou « Léonide ». Certaines publications avancent que lorsqu'un prénom relativement fréquent est donné autant aux garçons qu'aux filles, les parents tendent à l'abandonner pour leurs fils si bien que le prénom épiciène tend à devenir féminin : ainsi aux États-Unis, 'American parents have the legal right to give any name. This allows formerly male names like Ashley, Beverly, and Madison to shift into names for girls. Because of sexism, shifts from female to male are rare' (Lawson 2016, p. 188). Toutefois, le [fichier INSEE des prénoms](#) ne confirme pas ce scénario en France depuis 1900, ni pour les « Claude » ni les « Dominique » ni les « Camille ». Mais parmi les prénoms les plus souvent attribués en Pennsylvanie en 1990, les prénoms masculins se terminant par une lettre « féminine » (comme -a, -e ou -i) sont moins fréquents que les prénoms féminins se terminant par une lettre « masculine » (autres lettres, principalement des consonnes) (Barry et Harper 2000). Ces phénomènes, s'ils étaient effectivement observés sur un grand nombre de pays, pourraient éventuellement s'expliquer par la même raison qui fait que dans l'Occident contemporain les femmes peuvent porter des pantalons mais les hommes ne peuvent pas porter de jupe : si l'origine féminine du prénom ou du vêtement est trop évidente, le fait qu'un homme le porte serait pour lui déshonorant. Notons toutefois qu'en Israël, 'A current trend [...] is to give children monosyllabic Hebrew names that are not gendered, nor particularly religious in meaning, and are homophonous with unrelated foreign counterparts (Gai, Ron, Tom). This trend reflects developments in Western society, namely the blurring of gender roles and cultural boundaries. It is a Hebrew expression of a more universalistic egalitarian identity of the modern Israeli' (Lawson 2016, p. 181).

<sup>98</sup> Pour une utilisation des prénoms portés à l'époque médiévale de l'Espagne à l'Iran (en passant par la Tunisie, l'Égypte, la Syrie et l'Irak) pour retracer l'histoire quantitative de la conversion à l'islam, voir Bulliet 1979.

le stock de prénoms portés par les Berbères ne se distingue presque plus du stock de prénoms portés par les Arabes. Concrètement, les populations berbères ont adopté des prénoms arabes soit tels quels, soit sous une forme berbérisée, c'est-à-dire selon une variante qui, par érosion linguistique, s'est adaptée aux contraintes linguistiques et notamment phonétiques des langues berbères. Par exemple le prénom arabe « Mohamed » est devenu, sous sa forme berbérisée, « Mohand », « Abou » est devenu « Abbou », « Amina » « Mina », « Ismaïl » « Smaïn », « Issa » « Issou », « Hassan » « Lahcen », etc.<sup>99</sup>

Au Maghreb, au fur et à mesure que, du 7<sup>e</sup> au 19<sup>e</sup> siècles, les prénoms s'arabisent, ils s'islamisent aussi : non seulement se diffusent les prénoms arabes des personnages du Coran et des proches du prophète Mohamed, mais en outre se multiplient les prénoms religieux ou théophores<sup>100</sup> « à base de *Eddine* comme *Djameleddine*, *Seïfeddine*, *Zineddine*, *Abideddine*, à base de *Ellah* comme *Ferdjellah*, *Saadallah*, *Fatahellah*, *Razkellah* ou encore à base de *Abd* comme *Abdellah*, *Abdelkarim*, *Abdasamed*, *Abdelislam* ainsi qu'un vocabulaire relatif à la pratique de la religion tel que *Achour*, *Laïd*, *Belaïd*, *Ramdane*, *Chabane*, *Mabrouk* » (Yermèche 2013, p. 55). Mais en remplacement des prénoms berbères apparaissent aussi des prénoms arabes profanes comme « Dehbia », « Djouhra », « Fatma », « Lyes », « Nedjma », « Seghir » ou « Sofiane » (Yermèche 2013, p. 56). De nos jours, seule une minorité de Berbères portent des prénoms spécifiques : la plupart d'entre eux portent des prénoms de fonds lexical arabe qui ont été berbérisés, d'autres portent des prénoms berbères (Yermèche 2013, p. 48-50). Il peut s'agir de prénoms berbères attribués il y a quelques décennies, avant les indépendances, comme « Ijja » (celle qui a une odeur douce), « Idir » et « Iddr » (le vivant, respectivement en kabyle et en tachelhit) ; ou de prénoms attribués plus récemment, à la suite des revendications identitaires berbères des années 1980, comme « Massinissa » (du nom d'un chef berbère des 3<sup>e</sup> et 2<sup>e</sup> siècles avant l'ère commune, premier roi de Numidie), « Yugurtha » (du nom d'un roi de Numidie du 2<sup>e</sup> siècle avant l'ère commune qui résista contre l'expansion romaine) et « Kahina » (du nom d'une reine berbère du 7<sup>e</sup> siècle qui résista contre l'expansion arabe).

Suite à l'établissement de l'état civil des indigènes d'Algérie (1882), les prénoms arabes (et berbères) sont translittérés de l'écriture arabe vers l'alphabet latin, avant d'être réarabisés suite à l'indépendance (1962). Mais pendant la période coloniale au Maghreb, qui atteint sa durée la plus longue en Algérie (1830-1962), les stocks des prénoms portés par les Arabes et les Berbères d'une part, par les Français et les Européens d'autre part, restent presque totalement disjoints : presque aucun colon français ou européen ne se voit attribuer de prénom arabe,<sup>101</sup> et presque aucun indigène musulman ne se voit attribuer de prénom français ou européen.

« Durant toute la période coloniale, le prénom, au-delà de la signification sociale, symbolique ou rituelle, a toujours traduit le souci profond d'une famille algérienne de se démarquer de l'ordre français qui voulait faire d'elle par la contrainte une famille française. Pour se préserver de cette assimilation forcée, les Algériens allaient puiser dans le fond historique antérieur les prénoms les plus répandus dans les récits de l'Islam ; il fallait à tout prix se démarquer de l'autre : du Chrétien » (Varro et Lesbet 1986, p. 144).

<sup>99</sup> Les prénoms « Mohand » et « Mina » font partie de la liste d'environ 800 prénoms masculins et féminins arabes et berbères utilisés par l'Insee pour chiffrer les recensements 1954, 1962 et 1968 (« liste des prénoms musulmans les plus usuels relevés sur les registres d'état civil en Algérie ») (Couto 2014, p. 363-370).

<sup>100</sup> Des prénoms « théophores » sont des prénoms qui incluent le nom d'une divinité (encadré 6.17).

<sup>101</sup> Pour une liste de prénoms d'Européens d'Algérie concessionnaires de la préfecture de Constantine de 1854 à 1927 (sous-série 2M), voir Gomez-Le Chevanton 2021, p. 399-412. Ajoutons que d'après des enquêtes orales menées auprès des couples formés dans l'Algérie coloniale entre un homme français et une femme espagnole, italienne, maltaise ou juive, une fois ces épouses naturalisées, elles francisaient le plus souvent leur prénom : « Francesca » devient « Françoise », « Enriqueta » devient « Henriette » et « Sultane » devient « Reine » (Colonna et Taraud 2006).

C'est ainsi que dans la période coloniale tardive, au cours des recensements de 1954 et de 1962 (mais aussi 1968), l'appareil statistique français peut utiliser le prénom (et le nom) des résidents d'Algérie pour distinguer les indigènes musulmans, qui portent des prénoms arabes ou berbères, des Français et des Européens d'Algérie, qui portent des prénoms chrétiens ou israélites.<sup>102</sup> Si la fiabilité de cet indicateur n'est pas contestée (même par ceux qui contestent la légitimité politique de son usage), c'est que **les stocks de prénoms des indigènes d'Algérie et des colons sont presque totalement disjoints**. Quand la chanteuse Joséphine Baker adopte en Dordogne deux enfants trouvés dans l'Algérie en guerre (1956), son fils de parents berbères se prénomme « Brahim » (et il sera renommé « Brian »), tandis que sa fille de parents pieds-noirs se prénomme « Marianne » (Denéchère 2011). De même, en France en 1992, « Pour les rapatriés originaires du Maroc, qui figurent dans l'échantillon, presque tous avaient un prénom français » (Tribalat et al. 1996, p. 162).

Au Maghreb depuis les indépendances et les politiques d'arabisation menées à l'époque du nationalisme panarabe, non seulement la langue de l'enseignement est devenue l'arabe (au détriment du français et surtout des langues berbères), mais diverses dispositions juridiques et pratiques administratives (listes plus ou moins informelles de prénoms autorisés) ont cherché à interdire, empêcher ou rendre difficile l'attribution aux nouveau-nés de prénoms non-arabes et avant tout berbères, au moins jusqu'aux années 2000 (Maroc) si ce n'est jusqu'à nos jours (Algérie) (Tilmatine 2015, p. 13). Ainsi en Algérie, l'article 64 de l'Ordonnance n°70-20 du 19 février 1970 relative à l'état civil dispose : « Les prénoms doivent être de consonance algérienne ; il peut en être autrement pour les enfants nés de parents appartenant à une confession non musulmane » ; « Sont interdits tous les prénoms autres que ceux consacrés par l'usage ou par la tradition ».

Quant à l'évolution du stock des prénoms attribués au Maghreb depuis les indépendances, elle ne semble pas avoir à ce jour été étudiée dans le détail. Il reste toutefois d'usage, dans certaines familles, de prénommer son premier fils « Mohamed » (du prénom du prophète de l'islam) et sa première fille « Fatima » (du prénom de la fille préférée du prophète Mohamed). Plus généralement, il reste d'usage de prénommer ses enfants en rappel des prénoms de certains aïeux. D'après certaines observations datant des années 1980, « Paradoxalement, la francophonie n'a jamais été aussi répandue que depuis l'indépendance : c'est l'introduction, par l'intermédiaire de quelques familles algériennes francophones, de prénoms à résonance non-"islamique" en rupture avec la connotation religieuse mais sans que cela fasse "trop

---

<sup>102</sup> Au cours du recensement de 1954, « les « personnes nées en Algérie ayant à la fois un nom et un prénom à consonance "arabe ou berbère" sont [...] classées en "musulmans originaires d'Algérie" et ceux qui ont "un prénom chrétien ou israélite" en "Français de naissance originaires d'Algérie" ». Le même principe sera appliqué au recensement de 1962, alors que l'Algérie vient d'accéder à l'indépendance. [...] Comme en 1954, le code distingue les Français des Algériens selon le lieu de naissance et la consonance arabe ou berbère de ses noms et prénoms : « On classera en 01 : "Français musulmans d'Algérie" les personnes nées en Algérie qui ont à la fois un nom et un prénom à consonance arabe ou berbère ». » (Escafré-Dublet et al. 2018, p. 50). Et pour le recensement de 1968, « l'INSEE a exceptionnellement décidé de classer les personnes françaises provenant de l'ancienne colonie et ayant « un nom et un prénom à consonance arabe ou berbère » sous l'intitulé « musulmans originaires d'Algérie ». [...] Les personnes nées françaises en Algérie ayant un nom et un prénom à consonance arabe ou berbère sont classées sous l'intitulé « musulmans originaires d'Algérie », qu'elles aient accédé ou non, par naturalisation, reconnaissance ou réintégration, à la nationalité de plein droit. » (Couto 2013, p. 95). De fait, dans les consignes de chiffrage du recensement général de la population de 1968 (RP 1968, Code n°25 bis), pour distinguer — parmi les personnes nées en Algérie — les musulmans algériens des non-musulmans, figurent d'une part une liste d'environ 800 prénoms masculins et féminins arabes et berbères dont les porteurs sont à catégoriser comme « musulmans » (« liste des prénoms musulmans les plus usuels relevés sur les registres d'état civil en Algérie », code DN=30), d'autre part une liste d'une cinquantaine de noms juifs dont les porteurs sont à catégoriser comme « non musulmans » (« noms israélites d'origine arabe ou berbère », code DN=31) (Couto 2014, p. 363-370).

français". C'est l'ère des Ludmilla, Morgan, Saly, Médy, Fériel, etc... » (Varro et Lesbet 1986, p. 144). En Algérie dans les années 1980 certains couples francophones choisissent effectivement pour leurs enfants des prénoms dits « passe-partout »<sup>103</sup> : « Samy, Karim, Merouane, Malik, etc... pour les garçons », et « Fériel, Lydia, Sabrina etc... pour les filles » (Varro et Lesbet 1986, p. 147). Toutefois, en Algérie dans les années 1980,

« Si le père algérien annonce aux siens que son fils s'appelle Jean-Pierre, il sera rejeté sans attendre de savoir que ce dernier parle arabe et sera un fervent croyant de l'Islam, tandis que s'il dit qu'il s'appelle Mohamed, Réda ou Khallil, il sera accepté sans que personne ne cherche à savoir si le petit mange du porc, ne parle que le français et ne connaît même pas l'existence de Dieu » (Varro et Lesbet 1986, p. 147).

Au Maroc sont parfois attribués, dans les années 1990, des prénoms arabes jusqu'alors inconnus comme les prénoms féminins « Ihâm » (inspiration) et « Ahlâm » (rêves), et qui sont inspirés des personnages de feuilletons télévisés égyptiens, et dans les années 2000 des prénoms latins dérivés de feuilletons latino-américains (comme « Maria »).

### **Encadré 6.1. L'arabisation des prénoms dans le monde musulman**

Le prestige associé à la langue arabe (la langue du texte sacré) ainsi qu'au prophète de l'islam et à ses premiers compagnons et successeurs (porteurs de prénoms arabes) a conduit au fil des siècles les populations islamisées à adopter le stock de prénoms arabes, même lorsque ces populations islamisées n'ont pas été arabisées. C'est le cas chez les Berbères du Maghreb, qui depuis environ un millénaire portent largement des prénoms d'origine arabe (voir ci-dessus). C'est le cas aussi chez les turcophones et les kurdophones de Turquie, qui des années 1950 aux années 2010 attribuaient environ 50 % de prénoms d'origine arabe, et seulement environ 25 % de prénoms d'origine turque (encadré 6.2). C'est le cas aussi chez les Persans chiites d'Iran, qui des années 1960 aux années 1980 attribuaient des prénoms en grande majorité d'origine arabe, et seulement environ 25 % de prénoms d'origine persane non islamique (Habibi 1992). Et c'est le cas, plus généralement, dans le monde musulman.

Ainsi en Asie du sud (Pakistan, Inde et Bangladesh), si les prénoms des sikhs sont d'origine penjabi, et si les prénoms des hindous sont d'origines très variées (indo-européennes et particulièrement sanskrite, ou dravidiennes et notamment tamoule), les prénoms des musulmans sont pour la plupart d'origine arabe, même si quelques-uns sont d'origine persane (Hanks et Parkin 2016, p. 226 ; pour un algorithme permettant d'inférer l'affiliation religieuse probable à partir des noms de personnes d'Asie du sud, voir Susewind 2015). De même en Asie du sud-est, les prénoms des Malais sont pour la plupart d'origine arabe (Jones et Phillips 1995) ; par exemple, dans les décennies 2000 et 2010 les prénoms les plus attribués par les Malais de Malaisie sont « Mohamed » pour les garçons et « Nur » (lumière, en arabe) pour les filles (Robustova 2016, p. 218). En Indonésie aussi, les prénoms des Minangkabaus de Sumatra sont pour la plupart d'origine arabe (Jones et Phillips 1995), et de 1910 à 2010 environ 40 % des prénoms attribués à Java sont d'origine exclusivement

<sup>103</sup> Conformément à la littérature qualitative sur les prénoms attribués par les couples « mixtes » (Annexe 7, p. 327), j'appelle « passe-partout » des prénoms qui « ne sonnent pas trop arabe » tout en pouvant passer pour l'être : il peut s'agir de prénoms arabes homophones de prénoms anglophones (« Samy » est homophone de Sammy », etc.), ou de prénoms d'origine « mixte » ou ambiguë (arabe mais pas seulement), ou encore de prénoms non arabes (« anglophones » ou « latins » voire « français ») mais qui peuvent avoir une racine arabe crédible. De fait, la situation linguistique du Maghreb permet toutes sortes d'hybridations, puisque coexistent deux variétés d'arabe, classique et dialectale (dont les systèmes phonologiques ne coïncident pas entièrement), ainsi que les langues berbères et le français.



arabe, 20 % supplémentaires étant d'origine à la fois arabe et javanaise (Kuipers et Askuri 2017, p. 37).

L'arabisation du stock de prénoms attribués dans le monde musulman va de pair avec son adaptation à la phonologie des langues locales : ainsi le prénom du prophète est-il adopté sous la forme « Mohamed » au Maghreb et en Égypte, « Mamadou » en Afrique de l'ouest, « Maxamed » en somali, « Mehmed » en turc ottoman et en bosniaque et « Mehmet » en turc moderne, « Muhamet » en albanais, « Mokhmad » en tchéchène, « Mihemed » en kurde, « Mohamad » en persan et en pachtoune, « Muhammad » en penjabi et en bengali, « Mukhamed » en kazakh, « Muhamad » en malais et en indonésien, etc. Le grand nombre de variantes orthographiques du prénom du prophète de l'islam provient donc non seulement de la variété des translittérations possibles de l'écriture arabe vers d'autres écritures comme l'alphabet latin, mais aussi de la variété des formes que prend le prénom lui-même dans les nombreuses langues des parents qui l'attribuent. En Indonésie du début du 20<sup>e</sup> siècle jusqu'au début du 21<sup>e</sup>, « Muhammad » a remplacé « Mat » (Kuipers et Askuri 2017, p. 42).

Le processus d'arabisation des prénoms dans le monde musulman — avec diverses nuances régionales, sectaires, etc. — semble connaître un regain depuis l'essor de la religiosité islamique et de l'islam politique à partir des années 1970. Ainsi au Pakistan, les prénoms ourdous ou persans reculent depuis les années 1970 au profit des prénoms arabes (Rahman 2013), et certains membres des minorités religieuses (chrétiens, parsis, sikhs) modifient leur prénom pour éviter les discriminations et les persécutions (Rahman 2014). De même en Indonésie : si jusqu'aux années 1980 les Javanais attribuaient à la majorité de leurs enfants des prénoms d'origine javanaise (ou sanskrite) (Jones et Phillips 1995), depuis quelques décennies ils délaissent de plus en plus les prénoms javanais au profit de prénoms d'origine arabe, qui signalent leur piété islamique (Kuipers et Askuri 2017, p. 36-37 et 45). En outre en Indonésie, des gouvernements locaux envisagent d'interdire l'attribution de prénoms occidentaux (Llewellyn 2018). Ainsi l'arabisation des prénoms dans le monde musulman reflète-t-elle l'essor de la religiosité musulmane dans ces pays, mais aussi — et plus particulièrement — l'essor d'une variante fondamentaliste de l'islam, que les pays de la péninsule arabique (Arabie saoudite, Qatar, Émirats arabes unis) ont diffusée jusqu'en Indonésie en finançant des écoles coraniques (*madrasas*) qui nourrissent des attitudes singulièrement liberticides à l'égard des minorités religieuses (non musulmans, chiites, soufis, ahmadyyas, etc.).

Notons enfin que l'origine linguistique des prénoms attribués peut révéler l'état de l'opinion publique, en l'absence de sondages. Par exemple en Iran des années 1960 aux années 1980, les types de prénoms attribués par les parents persanophones chiites d'une ville moyenne comme Hamadan révèlent des mouvements d'opinion : alors que de 1963 à 1973, dans l'esprit des réformes du shah, les parents attribuaient de moins en moins de prénoms islamiques, de 1973 à 1979 ils se sont mis à en attribuer de plus en plus (au détriment des prénoms persans non islamiques promus par le shah), préfigurant ainsi la Révolution islamique ; puis de 1983 à 1988, une fois le régime islamique établi mais mis en difficulté par la guerre contre l'Irak, les parents ont de nouveau attribué de moins en moins de prénoms islamiques, au profit de prénoms non islamiques persans comme arabes (Habibi 1992, p. 256).

## Encadré 6.2. Les prénoms en Turquie

Si l'islamisation a conduit à l'arabisation partielle de la structure du nom personnel des Berbères (voir ci-dessus), elle a eu des effets similaires chez les Turcs. Le nom personnel des élites turques des époques seldjoukide (du 11<sup>e</sup> au 13<sup>e</sup> siècles) et ottomane (du 13<sup>e</sup> au 20<sup>e</sup>

siècles), malgré toutes ses variations (au fil du temps et entre lieux et milieux sociaux), emprunte en partie la structure du nom personnel arabe : il peut être composé du *ism* (premier nom, attribué à la naissance et porté par les membres de toute la société), du *mahlas* (nom de plume, porté par une étroite élite), du *lakab* (dernière partie du nom, conférée par l'usage en référence à l'origine géographique ou ethnique, aux caractéristiques physiques ou morales de l'individu, à ses titres honorifiques, etc.), et de la *kunya* (seulement sous les Seldjoukides) (Bouquet 2010, p. 215-219). Ce n'est que sous la République (1934) que l'État turc impose à ses citoyens de porter un prénom et un nom de famille héréditaire (Bouquet et Fliche 2013).

Comme plus généralement dans le monde musulman (encadré 6.1), l'islamisation des Turcs a aussi conduit au remplacement des prénoms préislamiques par des prénoms arabes (comme par exemple « Ali ») et par des prénoms arabes turquisés. Ainsi « Mohamed » a donné « Mehmed » puis « Mehmet », « Ahmed » a donné « Ahmet », « Hassan » « Hasan », « Hussein » « Hüseyin », « Omar » « Ömer », et pour les prénoms féminins « Aïcha » a donné « Aise », « Farida » « Feride », « Fatima » « Fatma », « Khadija » « Hatice », « Zaynab » « Zeynep », etc. Ainsi pendant la période ottomane (1299-1922), les prénoms turcs préislamiques ne sont plus attribués, et en Turquie de 1950 à 2010 environ 50 % des prénoms (féminins) attribués sont d'origine arabe (Sabuncu et Coulmont 2011).

Certains prénoms turcs sans origine arabe ont été de nouveau attribués depuis l'essor du nationalisme turc (au début du 20<sup>e</sup> siècle) et la République turque (1923) : en Turquie de 1950 à 2010, environ 25 % des prénoms (féminins) attribués sont d'origine turque (et non arabe, persane, etc.) (Sabuncu et Coulmont 2011). Parmi ces prénoms turcs attribués en Turquie depuis le début du 20<sup>e</sup> siècle, certains sont des prénoms associés à des légendes ou à des personnages turcs de la période préislamique : « Oguz » du nom d'Oghuz Khan, et « Alper » du nom d'Alper Tunga, deux héros légendaires des Turcs ; « Asena », du nom d'une louve de la mythologie turque ; « Mete » (courageux), du nom du fondateur de l'empire Xiongnu au 3<sup>e</sup> siècle avant l'ère commune ; ou « Göltekin », du nom d'un général du khaganat turc du 6<sup>e</sup> au 8<sup>e</sup> siècles. Mais la plupart des prénoms turcs attribués depuis le début du 20<sup>e</sup> siècle sont des prénoms issus de termes turcs sémantiquement transparents (c'est-à-dire, de signification évidente pour un turcophone), inventés depuis le 20<sup>e</sup> siècle plutôt que repris de prénoms turcs préislamiques : c'est le cas des prénoms masculins « Demir » (fer), « Güçlü » (fort), « Övünç » (fierté), « Savas » (guerre) ou « Seçkin » (élite), et de divers prénoms épiciques : « Özge » (autre), « Umut » (espoir), etc. Parmi ces prénoms turcs, certains sont plutôt connotés comme de gauche ou alévis, comme les prénoms masculins « Baris » (paix), « Devrim » (révolution) et « Ulas » (atteints) ou comme les prénoms épiciques « Deniz » (mer) et « Özgür » (libre), tandis que d'autres sont plutôt connotés comme de droite, comme les prénoms « Alp » (courageux), « Alparslan » (lion héroïque, 2<sup>e</sup> sultan de la dynastie seldjoukide, au 11<sup>e</sup> siècle) et « Eren » (saint). Il en va de même de certains prénoms turcs d'origine arabe, qui peuvent être connotés de gauche ou alévis comme « Haydar » (lion), « Cafer » ou « Muharrem », ou de droite comme les prénoms terminant en -din tels « Necmeddin » (étoile de la religion), « Fahrettin » (fierté de la religion), etc. Toutefois, peu de prénoms sont spécifiques aux alévis, à ceci près que des alévis n'attribueraient pas les prénoms « Osman » ou « Omer ».

Ainsi la trajectoire historique des prénoms des Turcs ressemble-t-elle à celle des prénoms des Finnois. Tout comme les prénoms turcs ont été remplacés par des prénoms arabes avec l'islamisation, au Moyen Âge les prénoms vieux-finnois ont été remplacés par des prénoms chrétiens, car l'Église catholique ne comptait pas de saints finnois (Saarelma 2009, p. 202). Et de même que les prénoms turcs d'origine non arabe ni musulmane n'ont été de nouveau attribués qu'avec le mouvement nationaliste du début du 20<sup>e</sup> siècle, les prénoms finnois d'origine non chrétienne (eux aussi sémantiquement transparents) n'ont été de

nouveau attribués qu'avec le mouvement nationaliste de la fin du 19<sup>e</sup> siècle (Eskola et Hämäläinen 2019 ; Leibring 2016, p. 208-209).

Ces observations concernent les turcophones de Turquie, mais qu'en est-il des kurdophones ? Jusque dans les années 1930 en Turquie, les prénoms des Turcs et des Kurdes ne diffèrent que peu : ce sont des prénoms musulmans d'origine principalement arabe et persane (Aslan 2009, p. 4). Mais à partir de l'escalade du conflit armé entre l'État turc et le Parti des travailleurs du Kurdistan (PKK), l'État interdit de parler kurde en public ou en privé (1983-1991) et les autorités locales rendent plus difficile qu'auparavant l'attribution aux nouveau-nés de prénoms kurdes (Aslan 2009, p. 5-6). Une fois que le PKK a abandonné la lutte armée et que l'État central (y compris la Cour de cassation) a assoupli les restrictions à l'encontre des expressions de la langue et de la culture kurdes (années 1990 et surtout 2000), des militants kurdes proposent de kurdifier les prénoms des Kurdes, notamment à l'aide de prénoms kurdes contenant les lettres q, x et w — qui n'existent pas dans l'alphabet turc —, tandis que les pouvoirs publics locaux (agents d'état civil, procureurs, juges, policiers et gendarmes) cherchent à empêcher ces parents d'attribuer à leurs enfants des prénoms kurdes (Aslan 2009, p. 1-2).

Pour un très bon article journalistique sur les aspects identitaires des pratiques de prénomination dans la Turquie contemporaine, voir Ashdown 2021.

#### ❖ **Le contenu sémantique des prénoms**

En France et plus généralement en Europe depuis le Moyen Âge, la plupart des prénoms proviennent du Nouveau Testament. Comme ils sont d'origine hébraïque, grecque ou latine, ils n'ont aucune signification pour un locuteur standard : ils sont « sémantiquement opaques » (Saarelma 2009, p. 201), ce qui explique peut-être en partie pourquoi les prénoms sont si difficiles à se remémorer (MacKenzie 2018, p. 303). Mais ce n'est le cas presque nulle part ailleurs et notamment pas dans le monde arabe, où bien au contraire la plupart des prénoms sont des noms communs, des adjectifs ou des verbes issus du vocabulaire courant, si bien qu'ils sont sémantiquement transparents (c'est-à-dire, de signification évidente pour un arabophone) (Landman 2016, p. 131). Parmi les prénoms arabo-musulmans sémantiquement transparents, on trouve notamment des prénoms de personnages du Coran comme « Mohamed » lui-même (le très loué), aussi appelé « Ahmed » (particulièrement louable) et « Moustafa » (l'élus), des prénoms théophores comme « Abdallah » (serviteur d'Allah), et des prénoms profanes à la signification ou à l'esthétique plaisantes, comme « Bachir » (annonciateur de bonnes nouvelles), « Samir » / « Samira » (compagnon de discussion du soir) ou « Yasmine » (jasmin) (Bramwell 2011, p. 38-40).

#### ❖ **Une typologie des prénoms attribués en Europe et au Maghreb**

Pour se repérer dans la variété des paysages onomastiques d'Europe et du Maghreb, il est utile de distinguer (i) des prénoms religieux tirés du texte sacré, (ii) des prénoms religieux tirés de l'histoire religieuse et (iii) des prénoms profanes.



**Tableau 6.3. Typologie des prénoms attribués en Europe et au Maghreb**

Prénoms religieux	Tirés du texte sacré	Europe	Maghreb
	Tirés de l'histoire	Prénoms des personnages du Nouveau Testament	Prénoms des personnages du Coran
Prénoms profanes		Prénoms des saints et martyrs chrétiens	Prénoms des proches et successeurs de Mohamed
		D'origines grecque, latine, germanique, slave, celte, etc.	D'origine principalement arabe, mais aussi berbère, persane, etc.

• **Les prénoms tirés du texte sacré**

En Europe, on l'a vu, la plupart des prénoms attribués depuis les environs du 12<sup>e</sup> siècle sont issus de personnages du Nouveau Testament. Rédigé en grec, le Nouveau Testament contient divers prénoms d'origine grecque comme « André » (qui signifie masculin), « Luc » (de Lucanie), « Philippe » (ami des chevaux), « Pierre » (pierre), « Stéphane » (couronne) et « Timothée » (honorer Dieu). Le Nouveau Testament contient aussi des prénoms d'origine latine, comme « Marc » (dieu romain Mars) ou « Paul » (petit). D'autres prénoms du Nouveau Testament sont d'origine hébraïque, comme « Élisabeth » (Dieu est un serment), « Gabriel » (Dieu est ma force), « Jacques » (sens incertain), « Jean » (Dieu est bon), « Joseph » (il ajoutera), « Marie » (sens incertain), « Matthieu » (don de Dieu) ou « Simon » (il a entendu). D'autres encore sont d'origine araméenne, comme « Barthélemy » (fils de Talmî<sup>104</sup>) et « Thomas » (jumeau).

Au Maghreb, beaucoup des prénoms attribués depuis le 7<sup>e</sup> siècle sont issus de personnages du Coran. Rédigé en arabe, le Coran contient des prénoms d'origine principalement arabe, comme « Mohamed » (le très loué). Mais comme le Nouveau Testament, le Coran contient aussi des prénoms d'origine hébraïque ou araméenne, y compris des formes arabes de prénoms issus de la Bible hébraïque comme « Adam », « Ayoub » (Job), « Daoud » (David), « Haroun » (Aaron), « Ibrahim » (Abraham), « Ismail » (Ismaël),<sup>105</sup> « Moussa » (Moïse), « Sulaymân » (Solomon),<sup>106</sup> « Youssef » ou « Youssef » (Joseph), « Yunus » (Jonas), « Zakaria » (Zacharie), etc.<sup>107</sup> Certains de ces prénoms sont courants dans le Maghreb contemporain, comme « Ibrahim », « Moussa », « Youssef » et « Zakaria ». Le Coran contient aussi des formes arabes de prénoms d'origine hébraïque issus du Nouveau Testament, comme « Aïssa » (Jésus),<sup>108</sup> ou « Yahya » (Jean). Le Coran contient enfin des formes arabes de

<sup>104</sup> « Barthélemy » proviendrait de Bar Talmî, où Bar signifie « fils de » en araméen, et où Talmî pourrait avoir une étymologie araméenne (fils du sillon) ou grecque (fils de Ptolémée).

<sup>105</sup> D'après la tradition musulmane, les juifs descendraient d'Isaac, fils d'Abraham, tandis que les musulmans descendraient de son demi-frère Ismail.

<sup>106</sup> La forme « Sulaymân » ne provient pas directement de l'hébreu Shlomo, mais de l'araméen, qui est aussi à l'origine des formes grecque et latine elles-mêmes à l'origine des formes européennes.

<sup>107</sup> Dans le Maghreb contemporain, certains de ces noms d'origine hébraïque sont plus répandus comme noms de famille que comme prénoms ; c'est le cas, semble-t-il, de « Ayoub », « Daoud » ou « Yacoub ». Plus généralement, « La présence hébraïque dans les noms de personne qui remonterait à très loin dans l'histoire maghrébine est significative. Les noms juifs présents dans l'anthroponymie algérienne relèvent pour la plupart du domaine religieux et sont probablement passés au berbère par l'arabe, *Youcef* (Josef), *Aïssa* (Jésus), *Moussa* (Moïse), *Braka* « bénédiction », *Elass/Elias* (Elie), *Haddad/Addad* "dieu des Orages, dieu national de Damas", *Ishaq* (Isaac), *Nouh* (Noé), *Rarrbo/Rebb* "Dieu" avec une extension de sens en kabyle et en arabe algérien sous la forme *Rebbi* "mon Dieu", *Younes* (Jonas), *Youssef* (Joseph), *Yakoub* (Jacob), *Zakari* (Zacharie), *Yahia* (Jean), *Yakoub* (Job)... » (Yermèche 2013, p. 51-52).

<sup>108</sup> Si des Maghrébins et plus généralement des Arabes peuvent prénommer leur fils « Aïssa » (Jésus) alors que les Européens sans origine maghrébine n'ont jamais prénommé leurs fils « Mohamed », c'est que Jésus est un prophète pour la tradition musulmane, alors que Mohamed — postérieur aux textes sacrés chrétiens — n'a aucune place dans le corpus chrétien. Et si « Mohamed » est un prénom plus populaire chez les musulmans que le prénom

prénoms d'origine hébraïque issus à la fois de l'Ancien et du Nouveau Testament, comme « Djibril » (Gabriel) et « Maryam » ou « Mariam » (Marie). Dans le Maghreb contemporain, ces prénoms coraniques d'origine hébraïque semblent avoir des connotations assez variées. Certains, fréquents, sont associés à des personnages juifs ou chrétiens préislamiques, comme « Ibrahim », « Moussa » et « Maryam », mentionnés dans la sourate Maryam. D'autres, moins fréquents, sont plus spontanément associés à de grandes figures de l'islam, que ce soit des personnages centraux du Coran comme « Djibril » ou des personnages historiques comme « Haroun » (Haroun Al-Rachid, calife abbaside des 8<sup>e</sup>-9<sup>e</sup> siècles, mis en scène dans *Les mille et une nuits*) et « Suleyman » (Soliman le Magnifique, sultan ottoman du 16<sup>e</sup> siècle).

Toujours au Maghreb, d'autres prénoms — théophores — sont issus des attributs que la tradition musulmane reconnaît à Dieu (encadré 6.17), de structure suivante : « Abd al-X », où Abd signifie « serviteur » (ou esclave), et X est l'un des 99 attributs d'Allah mentionnés dans le Coran. Parmi les plus courants figurent « Abdallah » (serviteur de Dieu), « Abdelali » (serviteur du sublime), « Abdelaziz » (serviteur du puissant), « Abdelhakim » (serviteur du sage), « Abdelhamid » (serviteur du loué), « Abdelkarim » (serviteur du généreux), « Abdellatif » (serviteur du doux), « Abdelmajid » (serviteur du glorieux), « Abdelmalik » (serviteur du roi), « Abdenmour » (serviteur de la lumière), « Abdelwahid » (serviteur de l'unique), « Abdul Rahman » (serviteur du miséricordieux), etc., parfois abrégés en « Abdel ». Dans le Maghreb contemporain, ces prénoms théophores et d'autres comme « Bouabdallah » (père du serviteur de Dieu) ont une forte connotation religieuse. Mais tel n'est pas le cas des prénoms comprenant seulement l'un des attributs de Dieu (« Aziz », « Karim », etc.), puisque ces attributs sont aussi des adjectifs du langage courant. À ces prénoms on peut aussi associer les prénoms terminant en -eddine, comme « Nasreddine » (soutien de la foi) et « Nouredine » (lumière de la religion), ainsi que d'autres prénoms comme « Iman » (foi).

### • Les prénoms tirés de l'histoire religieuse

En Europe, la plupart des prénoms attribués depuis les environs du 12<sup>e</sup> siècle sont les prénoms des calendriers des saints (chrétiens), qui varient selon les coutumes de chaque pays ou région. Au Maghreb aussi, parmi les prénoms les plus fréquemment attribués figurent ceux des proches du prophète Mohamed, comme « Aïcha » (vivante, du nom de la 3<sup>e</sup> épouse du prophète), « Ali » (sublime, du nom d'un cousin et gendre du prophète et 4<sup>e</sup> calife), « Amine » (véridique, du nom du 6<sup>e</sup> calife abbaside), « Amir » (prince, du nom du commandeur [des croyants]), « Abu Bakr » (chamelon, du nom d'un compagnon du prophète et 1<sup>er</sup> calife), « Fatima » (sevrer un petit, du nom d'une fille du prophète et épouse du 4<sup>e</sup> calife), « Hamza » (sens incertain, du nom du fondateur du druzisme), « Khadija » (prématuré, du nom de la première épouse du prophète et première personne convertie à l'islam), « Omar » (florissant, du nom du 2<sup>e</sup> calife), « Rachid » / « Rachida » (bien guidé(e), du nom des 4 premiers califes), « Saïd » / « Saïda » (heureux, du nom du titre des descendants de Mahomet), « Uthman » (oisillon, du nom du 3<sup>e</sup> calife), « Zaynab » prononcé au Maghreb « Zineb » (sens incertain, du nom d'une fille et de deux épouses du prophète), etc. Dans le Maghreb contemporain ces prénoms traditionnels semblent souvent transmis non seulement pour leur connotation religieuse, mais aussi en hommage à un(e) aïeul(e) qui en était porteur. De même, chez les Français originaires du Maghreb ils sont souvent attribués en hommage à un parent ou grand-parent, sinon en premier du moins en deuxième prénom.

---

« Jésus » ne l'est chez les chrétiens, c'est peut-être parce que pour les chrétiens Jésus est Dieu lui-même, tandis que pour les musulmans Mohamed est un prophète.

#### Encadré 6.4. Les prénoms des Arabes chrétiens

Si le Maghreb actuel compte de nos jours très peu de chrétiens, tel n'est pas le cas de tout le monde arabe. Mais alors, quels prénoms portent les chrétiens d'Égypte, du Liban et du reste du monde arabe ? Il semble que de nos jours les Arabes chrétiens peuvent porter trois ensembles de prénoms :

- i. la forme arabe de prénoms de l'Ancien ou du Nouveau Testament ou d'autres figures chrétiennes : des prénoms d'origine hébraïque mentionnés dans le Coran (et donc aussi portés par des Arabes musulmans) comme « Adam », « Ibrahim » (Abraham), « Ilyâs » (Elias), prononcé au Maghreb « Ilyes » ou « Lyes », « Issa » (Jésus), « Maryam » (Marie), « Mikhail » (Michel), « Sara », « Yahya » (Jean) et « Yûsuf » ou « Youssef » (Joseph) ; d'autres formes arabisées de l'hébreu comme « Hannah » (abréviation de « Yuhannah », qui est la forme arabe de l'hébreu « Yohanan ») ; des formes arabisées de prénoms d'origine grecque comme « Boutros » (Pierre), « Boulos » (Paul) et « Girgis » (Georges) ; on note aussi des formes non arabes mais coptes, comme « Shenouda » (fils de Dieu, en copte) ;
- ii. la forme européenne (principalement anglaise ou française) de prénoms de l'Ancien ou du Nouveau Testament ou de saints ou martyrs chrétiens : « Charles », « Cyril », « Émile », « Gabrielle », « George » / « Georges », « Joelle » / « Joëlle », « Joseph », « Mary » / « Marie », « Mark » / « Marc », « Michael » / « Michel », « Paul », « Peter » / « Pierre », etc., ainsi que divers prénoms européens non bibliques ;
- iii. des prénoms arabes profanes, qui contrairement aux prénoms de la catégorie précédente ne sont pas, dans le monde arabe, spécifiques aux chrétiens, sauf dans une certaine mesure des prénoms comme « Fadi » (sauveur, en référence à Jésus) ; certains de ces prénoms portés au Levant peuvent l'être par des chrétiens comme par des musulmans, comme « As'ad » (très heureux), « Jamîl » (beau), « Kamar » (lune), « Rami » (archer), « Salîm » (sain), etc.

#### • Les prénoms profanes

En Europe, les prénoms profanes peuvent être d'origine grecque comme « Alexandre » (défendre un homme) mais aussi latine, slave, celte, etc., et souvent germanique. Parmi les prénoms européens profanes d'origine germanique se trouvent notamment « Albert » (noble et brillant), « Alfred » (conseil des elfes), « Édouard » (riche garde), « Edgar » (lance riche), « Éric » (toujours chef), « Frédéric » (chef pacifique), « Gérard » (lance courageuse), « Henri » (chef de la maison), « Hugues » (esprit), « Mathilde » (force dans la bataille), « Raymond » (conseil protecteur), « Richard » (chef courageux) et « Robert » (brillante célébrité).

Au Maghreb, les prénoms profanes sont extrêmement nombreux. Si certains de ces prénoms arabes sont d'origine persane comme « Chakib » (patience), « Yara » (ami) et « Yasmine » (jasmin), la très grande majorité sont d'origine arabe et sémantiquement transparents dans cette langue. C'est le cas, notamment, de « Adel » (juste), « Alyâ » (ciel), « Amir » (prince) / « Amira » (princesse), « Anis » (ami) / « Anissa » (amie), « Assia » (qui reconforte), « Bachir » (annonciateur de bonnes nouvelles), « Fadila » (vertueuse), « Jamel » (beau) / « Jamila » (belle), « Dunya » (monde), « Farès » (chevalier), « Farhat » (heureux), « Farid » / « Farida » (unique), « Haifa » (svelte), « Haitham » (aiglon), « Hamza » (fort), « Hania » (plaisante), « Hisham » (généreux), « Jalil » (sublime), « Kamel » (parfait) / « Kamila » (parfaite), « Kenza » (trésor), « Khaled » (éternel), « Khalil » (ami), « Lamia » (brillante), « Leila » (nuit), « Malika » (reine), « Mona » (souhaits), « Messahoud » (heureux), « Mouloud » (nouveau-né), « Najwa » (secret), « Nijma » (étoile), « Nour » (lumière), « Rim »

(antilope), « Sadek » (loyal), « Sahar » (aube), « Sami » / « Samia » (sublime), « Samir » / « Samira » (compagnon de discussion du soir), « Sana » (splendeur), « Sayf » (épée), « Selma » (en sûreté), « Souad » (heureuse), « Sumaya » (là-haut), « Tahir » / « Tahirah » (pureté), « Tayeb » (agréable), « Walid » (né), « Wissam » (beau de visage) ou « Zahra » (rose). Dans le Maghreb contemporain, ces prénoms profanes semblent être attribués aussi bien par des parents très religieux que peu religieux. À ces prénoms arabes profanes peuvent être rattachés des prénoms non-arabes mais attribués au Maghreb parce que leur finale en -i ou en -(i)a sonne « arabe », qu'ils soient d'origine possiblement hébraïque (« Samy ») ou encore grecque (« Lydia »), celte et latine (« Sabrina ») ou slave (« Katia » et « Sonia »).

### ❖ Les prénoms des juifs en Europe et au Maghreb

De la période médiévale jusqu'aux alentours du 19<sup>e</sup> siècle, en Europe comme au Maghreb (mais aussi au Proche-Orient), la plupart des hommes juifs portent deux prénoms : un prénom religieux (*shem ha-qodesh*) d'origine hébraïque ou araméenne, attribué le jour de la circoncision et utilisé lorsque l'homme est appelé à lire la Torah à la synagogue et lors des rites de passage (bar mitzvah, mariage et décès) ; et un prénom séculier (*kinnui*), originaire de la langue vernaculaire (arabe, berbère, grecque, romane, germanique, slave, etc.) ou d'une langue étrangère, et utilisé dans leurs interactions quotidiennes entre juifs ainsi que dans les relations avec les non-juifs et avec l'administration (Beider 2012, p. 44, 51). Quant aux femmes juives, n'étant pas appelées à lire la Torah, elles ne se voient attribuer qu'un seul prénom, séculier (Beider 2012, p. 44). Mais jusqu'aux alentours du 19<sup>e</sup> siècle, même ces prénoms séculiers, masculins comme féminins, sont le plus souvent spécifiques aux juifs.

En Europe occidentale à la fin du 18<sup>e</sup> siècle puis au 19<sup>e</sup> siècle, l'émancipation des juifs — c'est-à-dire le fait qu'ils obtiennent des droits pleinement égaux à ceux de leurs concitoyens non juifs — les conduit, pour la première fois, à attribuer à leurs enfants des prénoms chrétiens, c'est-à-dire des prénoms du Nouveau Testament (Landman 2016, p. 137-138). Par exemple dans la bourgeoisie juive parisienne au cours du 19<sup>e</sup> siècle, la part des premiers prénoms attribués qui sont des prénoms juifs ne cesse de baisser, de près de 40 % (pour les hommes) dans la décennie 1810 à moins de 20 % dans la décennie 1830, pour passer sous les 5 % dans la décennie 1890 (Grange 2016b, p. 78-79).<sup>109</sup> De même dans la population juive de Turin (Italie du nord), en même temps que l'émancipation et la sortie du ghetto accroissent la part de mariages avec des non-juifs à partir de la décennie 1870, elles font chuter la part des prénoms attribués qui sont des prénoms juifs (Allegra 1997, p. 80-86).<sup>110</sup> Entre le début et la fin du 19<sup>e</sup> siècle, les garçons juifs de Turin se voient ainsi relativement moins souvent prénommés « Abram », « David », « Isach », « Moïse », « Salomon » ou « Samuel », pour être relativement plus souvent prénommés « Giuseppe » (forme italienne de « Joseph », déjà présente début 19<sup>e</sup> siècle) mais aussi « Dario » ou « Guido » (Allegra 1997, p. 82-84). Dans les Marches (Italie centrale),

« il est intéressant de noter qu'au moment où les juifs commencent à puiser ailleurs que dans le stock habituel de noms hébraïques, les chrétiens commencent à adopter des prénoms habituellement portés par des juifs, qu'ils n'avaient que très peu voire jamais portés jusqu'au début du XIX<sup>e</sup> siècle, comme Davide, Samuele, Emmanuele, Ester, Sara, etc. Ce changement dans les comportements est immédiatement visible et significatif pour

<sup>109</sup> Parmi les prénoms qu'il catégorise comme « classiques judéo-hébreux », Grange (2016, p. 77) mentionne : pour les garçons, Abel, Abraham, Daniel, David, Elie, Emmanuel, Isaïe, Isaac, Isidore, Israël, Jacob, Jérémie, Job, Jonas, Josué, Judas, Lazare, Moïse, Salomon, Samuel, Saul, Simon ; et pour les filles, Abigaël, Deborah, Esther, Judith, Léa, Rachel, Rebecca, Salomé, Sarah. Dans la bourgeoisie juive parisienne comme dans la population générale française, au cours du 19<sup>e</sup> siècle de plus en plus de nouveau-nés se voient attribuer des prénoms doubles ou triples (Grange 2016b, p. 70-73).

<sup>110</sup> Sur les prénoms des habitants du ghetto juif de Rome en 1733, voir Gasperoni 2018, p. 146-149.

l'historien moderniste travaillant sur le début de la période contemporaine et qui s'accommode habituellement de cette situation très nette de différenciation anthroponymique, puisqu'elle lui permet d'identifier rapidement et facilement les juifs dans les sources » (Gasperoni 2018, p. 148).

En Europe centrale et orientale aussi, la bourgeoisie juive se met pour la première fois à attribuer des prénoms chrétiens, et ce dès avant l'émancipation. C'est le cas à partir de la seconde moitié du 18<sup>e</sup> siècle dans les États allemands, parmi les juifs qui abandonnent le yiddish en faveur de l'allemand. Puis c'est le cas aussi, à partir de la seconde moitié du 19<sup>e</sup> siècle, dans certaines des plus grandes villes de l'actuelle Pologne (Cracovie, Varsovie), de l'actuelle Ukraine (Lwów, Odessa), de l'actuelle Lettonie (Riga) et en Russie (Saint-Petersbourg, Moscou), parmi les juifs qui abandonnent le yiddish en faveur de la langue slave ou balte locale (Beider 2012, p. 45-50).

Au Maghreb à la même période, et plus précisément dans l'Algérie coloniale,

« Quelques éléments montrent qu'une certaine évolution se produit au sein des familles indigènes israélites au cours du XIX<sup>e</sup> siècle, en particulier au cours des dernières décennies, celles postérieures au décret Crémieux [décret de 1870 qui attribue collectivement la citoyenneté française aux indigènes algériens de confession juive, qu'ils soient berbérophones ou séfarades]. Hommes et femmes tendent à se marier plus tard, et cela est particulièrement sensible pour les femmes, proportionnellement moins nombreuses à se marier avant leur vingtième anniversaire. L'étude des prénoms montre également une influence de la culture européenne, avec l'adoption, minoritaire mais significative, de prénoms inconnus auparavant dans cette population. La pratique des prénoms doubles, israélites et européens, peut témoigner d'une volonté de double appartenance. À la fin du XIX<sup>e</sup> siècle, le processus transitionnel est indéniablement engagé, du moins pour une partie des familles, celles qui sont le plus en contact avec les Européens ou le plus sensibles à leur culture » (Kateb et Brunet 2018, p. 20).

Puis l'assimilation onomastique des juifs du Maghreb se poursuit au 20<sup>e</sup> siècle, particulièrement à partir de leur arrivée en métropole.

« Le prénom des juifs du Maghreb témoigne, comme le nom, de la diversité des langues utilisées par ces communautés dans leur histoire contemporaine. Nous pouvons ainsi recenser six stocks linguistiques de prénoms : les Esther, David, Rachel, Salomon, Deborah et Michaël relèvent de la syntaxe hébraïque très ancienne ; les Makhoulouf, Saada, H'biba, Messaouda rappellent l'usage du judéo-arabe, tandis que les Perla, Mercedes ou Nino, Umberto et Gilda indiquent que les juifs d'Afrique du Nord ont également parlé, pour certains d'entre eux, l'espagnol et l'italien. Le prénom traduit donc la volonté des juifs de s'inscrire dans leur contexte social local, et dans les formes de communication spécifiques du milieu ambiant. [...] On voit ainsi apparaître progressivement dans le courant du XX<sup>e</sup> siècle les prénoms français, au fur et à mesure que les familles s'occidentalisent et surtout, scolarisent leurs enfants dans le système éducatif français. [...] L'attribution d'un prénom français était la porte ouverte ou le premier pas vers l'insertion dans la société française » (Bahloul 1985).

On verra plus loin qu'en France métropolitaine en 2008, seuls 6 % des habitants d'origine juive portent un prénom juif, c'est-à-dire que la plupart des personnes d'origine juive portent des prénoms non juifs (tableau 6.15a). En outre, seuls 6 % des porteurs d'un prénom juif sont d'origine juive, c'est-à-dire que la plupart des porteurs d'un prénom juif ne sont pas d'origine juive (tableau 6.15b). Plus généralement, en France en 2008, l'origine religieuse des individus continue de déterminer fortement leur type de prénom : la plupart des habitants d'origine juive (59 %) ou chrétienne (82 %) ou sans origine religieuse (75 %) portent un prénom « français »,

tandis que la plupart des habitants d'origine musulmane portent un prénom « arabo-musulman » (73 %) plutôt que « français » (4 %) (tableau 6.15a).

### Encadré 6.5. Les prénoms juifs

Les prénoms juifs, c'est-à-dire les prénoms qui sont usuellement attribués par les juifs à leurs enfants (même s'ils ne sont pas tous exclusivement attribués par des juifs), sont composés de trois ensembles : les prénoms hébreux antiques, aussi appelés prénoms hébraïques ; les prénoms des diasporas juives ; et les prénoms hébreux modernes, aussi appelés prénoms israéliens.

#### i. Les prénoms hébraïques (et araméens)

Dans l'Israël antique jusqu'au 6<sup>e</sup> siècle avant l'ère commune, tant que les juifs parlent hébreu, la plupart d'entre eux portent un prénom unique d'origine hébraïque, c'est-à-dire un prénom dont la langue d'origine est l'hébreu classique (Beider 2012, p. 51). Puis dans le Levant antique des alentours du 6<sup>e</sup> siècle avant l'ère commune (date de l'exil d'une partie de l'élite juive à Babylone) jusqu'au 7<sup>e</sup> siècle de l'ère commune (date du début de l'arabisation de la région), alors que la plupart des juifs parlent non plus hébreu mais araméen, la plupart d'entre eux portent un prénom unique d'origine hébraïque *ou* araméenne, notamment des prénoms terminant en -ai comme « Haggai » (festif) ou « Shabbetai » (repos) (Beider 2012, p. 43, 51).

Parmi les prénoms hébraïques se trouvent la plupart<sup>111</sup> des 2 000 prénoms mentionnés dans la Bible hébraïque (l'Ancien Testament), qui sont des prénoms d'origine hébraïque et sémantiquement transparents pour un locuteur de l'hébreu classique : c'est le cas de « Léa » (vache), « Rebecca » (veau gras) ou « Rachel » (agneau) — des prénoms typiques d'une société d'éleveurs comme les hébreux anciens (Lawson 2016, p. 171) —, mais aussi « Deborah » (abeille)<sup>112</sup> et des prénoms masculins comme « Isaac » (il rira) ou « Joseph » (il augmentera) (Beider 2012, p. 43 ; Landman 2016, p. 131, 136). La plupart des 2 000 prénoms mentionnés dans la Bible hébraïque sont des prénoms masculins théophores, c'est-à-dire qu'ils incluent le nom d'un dieu : c'est le cas notamment des prénoms « Ariel » (lion de Dieu), « Daniel » (Dieu est mon juge), « Emmanuel » (Dieu avec nous), « Ézéchiel » (que Dieu donne la force), « Gabriel » (Dieu est ma force), « Israël » (Dieu combat), « Joël » (Yahvé est Dieu), « Jonathan » (Dieu a donné), « Michel » (qui est comme Dieu ?), « Nathanaël » (don de Dieu), « Raphaël » (Dieu a guéri) et « Samuel » (nom de Dieu), le « el » étant l'abréviation de « Elohim », l'un des noms du Dieu de la Bible hébraïque (Landman 2016, p. 133). Parmi les prénoms féminins théophores, on trouve « Elisheva » (Dieu est un serment, à l'origine du prénom « Élisabeth ») (Landman 2016, p. 133). Les prénoms hébraïques mentionnés dans le Nouveau Testament, quant à eux, reflètent pour la plupart l'état du stock des prénoms chez les juifs du Levant du temps de Jésus, y compris ceux dont sont issus en français les prénoms « Joseph », « Marie » et « Anna » (Lawson 2016, p. 180).

Dans la littérature française du 17<sup>e</sup> au 20<sup>e</sup> siècles (base de données Frantext), les prénoms de l'Ancien Testament les plus fréquemment mentionnés sont « Adam » (mentionné dans 655 textes sur 3 000), « Moïse » et « Jacob », devant « Raphaël », « Israël », « Job » et — première femme — « Ève » (266 textes sur 3 000) (Brunet 1988, p. 140). Le prénom « Ève » vient de l'hébreu classique (*vie*), comme « Zoé » vient du grec classique (*vie*).

<sup>111</sup> Tous les prénoms de la Bible hébraïque (l'Ancien Testament), écrite en hébreu classique, ne sont pas d'origine hébraïque : par exemple, « Esther » est d'origine persane.

<sup>112</sup> Le prénom « Deborah » vient du terme désignant en hébreu classique une « abeille », tout comme le prénom « Melissa » vient du terme grec classique pour « abeille ».

## **ii. Les prénoms diasporiques**

Dans la diaspora juive antique du pourtour méditerranéen des alentours du 4<sup>e</sup> siècle avant l'ère commune jusqu'au 7<sup>e</sup> siècle de l'ère commune, parmi les juifs qui parlent grec puis latin, certains hommes portent deux prénoms : un prénom d'origine hébraïque ou araméenne, d'usage entre juifs, et un autre prénom d'origine grecque ou latine, comme « Alexandros » ou « Titus », d'usage dans les contacts avec la majorité non juive ou avec l'administration (mais parfois aussi entre juifs) (Beider 2012, p. 43-44, 51).

Plus tard, dans la diaspora juive médiévale et moderne du 7<sup>e</sup> aux alentours du 18<sup>e</sup> siècles, alors que les juifs ne parlent plus l'araméen (remplacé au Moyen-Orient par l'arabe) mais les diverses langues locales de leurs lieux de résidence, la plupart des hommes portent deux prénoms : un prénom religieux (*shem ha-qodesh*) d'origine le plus souvent hébraïque (y compris des prénoms délaissés depuis plusieurs siècles comme « Abraham », « David », « Gabriel » ou « Salomon »), et un prénom séculier (*kinnui*), originaire de la langue vernaculaire locale (arabe, grecque, romane, germanique, slave, etc.) mais généralement spécifique aux juifs (Beider 2012, p. 44, 51).

Parmi les prénoms ashkénazes, c'est-à-dire les prénoms attribués par les juifs yiddishophones de l'Europe germanique et slave des alentours du 10<sup>e</sup> au 19<sup>e</sup> siècles, on peut mentionner « Ber » (ours) et « Volf » (loup) pour les garçons, et « Gute » (bonne) et « Freyde » (joie) pour les filles. Et parmi les prénoms séfarades, c'est-à-dire les prénoms attribués par les juifs ladinophones originaires de la Péninsule Ibérique et qui suite à leur expulsion d'Espagne (1492) se sont, jusqu'au 19<sup>e</sup> siècle, dispersés dans l'empire ottoman, on peut mentionner « Angel » (ange) et « Vidal » (vital) pour les garçons, et « Oro » (or) et « Sol » (soleil) pour les filles. Ces prénoms séfarades n'incluent pas les prénoms des juifs présents au Moyen-Orient (du Yémen à l'Irak) et au Maghreb dès l'antiquité, et devenus pour la plupart arabophones à partir du 7<sup>e</sup> siècle.

Dans l'histoire des diasporas juives, alors qu'à partir de la seconde moitié du premier millénaire la plupart des garçons se voyaient attribuer (au moins) un prénom hébraïque, tel n'était pas le cas des filles, qui portaient plutôt des prénoms tirés de la langue localement utilisée par les juifs, ou des prénoms issus de la société non juive environnante (Landman 2016, p. 134 ; voir aussi Beider 2014). En effet, même si la judéité se transmet par la mère, diverses pratiques religieuses se transmettent par les pères, ce qui peut contribuer à expliquer le choix, pour les garçons, de prénoms plus traditionnels et plus religieux.

## **iii. Les prénoms israéliens**

En Israël aujourd'hui, les juifs portent des prénoms hébraïques mais aussi des prénoms ashkénazes et séfarades, et surtout des prénoms israéliens modernes. Ces prénoms sont eux aussi sémantiquement transparents pour les locuteurs de l'hébreu moderne, comme le prénom masculin « Yuval » (crique) et les prénoms féminins « Avital » (Dieu est la rosée), « Merav » (le plus), « Ofrah » (chevreuil) et « Tovah » (bien) (Landman 2016, p. 136). Landman (2016) distingue ainsi trois périodes de nomination en Israël depuis 1948. Jusqu'aux années 1960, les prénoms attribués sont dans la lignée des prénoms de la diaspora, comme par exemple « Aharon », « David », « Pinchas » et « Yosef » pour les garçons, et « Channah », « Leah », « Rivkah » et « Ziporah » pour les filles (Landman 2016, p. 137-138). Puis au cours des années 1960, ces prénoms traditionnels refluent au profit de prénoms nouveaux qui reflètent le melting-pot spécifiquement israélien : certains sont tirés de la Bible hébraïque comme « Asaf » (rassemblé) et « Yuval » (crique) pour les garçons, et « Merav » (le plus) et « Ofrah » (chevreuil) pour les filles ; d'autre sont tirés de l'hébreu moderne, comme par exemple « Amit » (ami) et « Tal » (rosée) pour les garçons, et « Limor » (mon parfum) et « Talie » (ma rosée) pour les filles (Landman 2016, p. 138-139). Et depuis les années 1970, dominant des prénoms spécifiquement israéliens, tirés parfois de la Bible hébraïque comme « Ethan » (ferme) ou « Naomi » (jolie), mais surtout de l'hébreu moderne, qu'ils soient de

nature théophore ou qu'ils évoquent la nature ou des sentiments plaisants (Landman 2016, p. 139-143), ou qu'ils soient des prénoms anglo-américains avec un sens en hébreu moderne, comme « Dana » (elle juge), « Guy » (vallée) ou « Ron » (chanson) (Landman 2016, p. 144).

### ❖ Résumé

En Europe depuis les environs du 12<sup>e</sup> siècle, la plupart des prénoms attribués sont issus du Nouveau Testament ou font référence à un saint chrétien, quelle que soit l'origine étymologique de ces prénoms — hébraïque, grecque, latine, germanique, etc. Au Maghreb en revanche, depuis les environs du 7<sup>e</sup> siècle, la plupart des prénoms attribués sont des noms communs, des adjectifs ou des verbes de la langue arabe elle-même (parfois berbérisés), dont la signification est compréhensible par les arabophones. Pendant la période coloniale en Afrique du Nord et notamment en Algérie (1830-1962), les prénoms portés par les Arabes et les Berbères d'une part, par les Français et les Européens d'autre part, restent presque totalement disjoints. Quant aux juifs, c'est lors de leur émancipation en Europe aux 18<sup>e</sup> et 19<sup>e</sup> siècles que certains se sont mis, pour la première fois, à attribuer à leurs enfants des prénoms chrétiens. Dans le monde depuis les années 1970, les prénoms attribués par les musulmans non-arabes (indonésiens et malais, pakistanais, iraniens, africains, etc.) semblent être de plus en plus souvent des prénoms d'origine arabe, par souci de signaler leur piété islamique.



## **Annexe 6B : La nomenclature des prénoms par origine**

Cette Annexe 6B, associée au Chapitre 6, p. 25, comprend la méthodologie de la reproduction au plus près concernant la nomenclature des prénoms par origine.<sup>113</sup>

### **❖ La nomenclature des prénoms par origine de l'INED**

La nomenclature des prénoms que les auteurs et l'INED m'ont finalement communiquée est issue des travaux de Humpert et Schneiderheinze (2000 et 2013). Pour attribuer à chaque anthroponyme (binôme prénom-nom) de l'annuaire téléphonique allemand une origine géolinguistique probable, Humpert et Schneiderheinze (2000) se fondent sur des répertoires et encyclopédies contenant au total plusieurs centaines de milliers de noms de famille issus de plusieurs dizaines de langues ou dialectes, ainsi que des listes contenant au total plusieurs dizaines de milliers de prénoms, les plus fréquemment attribués dans certains pays. Humpert et Schneiderheinze (2013) ont ensuite enrichi leur base de données, qui comprend à ce jour plus de 26 millions d'anthroponymes (et leur origine géolinguistique probable), ces anthroponymes étant composés de plusieurs millions de noms de famille et de plusieurs centaines de milliers de prénoms, y compris des prénoms « arabo-musulmans » catégorisés comme tels car ils figurent dans des livres listant des prénoms arabes et/ou musulmans.

Le recodage de la variable CAT1 qui doit être effectué pour passer de 82 modalités (comme dans la documentation méthodologique communiquée par les auteurs et l'INED au réseau Quételet et au CASD) à 8 modalités (comme dans leur article) est explicité dans le message que Patrick Simon m'a finalement envoyé le 4 décembre 2020 (Annexe 4A, p. 94).<sup>114</sup>

---

<sup>113</sup> Les thèses soutenues dans le présent document n'engagent personne d'autre que moi et je suis seul responsable de toute erreur. Je tiens toutefois à remercier chaleureusement des proches et de nombreux chercheurs de m'avoir aidé dans l'analyse des résultats présentés dans cette Annexe 6B :

- sur les départements français d'outre-mer : Vincent Cousseau, historien spécialiste d'onomaistique martiniquaise ;
- sur l'Europe du Sud : Michaël Gasperoni, historien spécialiste d'onomaistique italienne (et juive) ;
- sur l'Europe centrale et orientale : Antonela Capelle-Pogacean, historienne et politiste spécialiste de l'Europe centrale et orientale ; Nadège Ragaru, historienne et politiste spécialiste de l'Europe du Sud-est ; Marie Vrinat-Nikolov, spécialiste de littérature bulgare ;
- sur le Maghreb et le monde arabe : Manel Khelifa, arabophone ; Mohamed Lahrouchi, linguiste spécialiste des langues berbères ; Pierre Larcher, linguiste spécialiste de l'arabe ; Catherine Mayeur-Jaouen, historienne spécialiste du monde arabo-musulman ;
- sur le monde turc : Olivier Bouquet, historien spécialiste d'onomaistique ottomane ; Élise Massicard, politologue spécialiste de la Turquie contemporaine ; Özgür Türesay, historien de l'empire ottoman ;
- sur l'Afrique subsaharienne et particulièrement l'Afrique de l'ouest : Laurent Barry, anthropologue spécialiste des Peuls ; Doris Bonnet, anthropologue spécialiste du Burkina Faso et du Cameroun ; Fred Eboko, politologue spécialiste du Cameroun ; Oumar Marega, géographe spécialiste de l'Afrique de l'ouest ; Sadio Soukouna, politiste spécialiste du Mali ; Henry Tourneux, linguiste spécialiste des langues tchadiques ; Valentin Vydrin, linguiste spécialiste des langues mandé ;
- sur l'Asie et particulièrement l'Asie du sud-est : Anne Guillou, anthropologue spécialiste du Cambodge ; Philippe Le Failler, historien du Vietnam ;
- sur le monde juif : Alexandre Beider, linguiste spécialiste d'onomaistique juive ;
- je remercie aussi Luca Marin, Directeur du Centre d'information et d'études sur les migrations internationales (Ciemi), de m'avoir communiqué une riche bibliographie sur les noms des personnes en lien avec les migrations internationales.

J'en profite pour dire l'importance majeure, pour la recherche en sciences sociales, du recrutement de chercheurs spécialisés dans l'une des diverses aires culturelles du monde (études aréales). Qui a eu la chance et l'honneur de pouvoir échanger avec de grands savants, comme Pierre Larcher ou Alexandre Beider, sait la valeur inestimable de telles compétences.

<sup>114</sup> Je ne tiens pas compte du recodage de la variable CAT1 de 82 à 8 modalités tel que réalisé dans les 3 programmes informatiques R que les auteurs et l'INED m'ont finalement communiqués (6 janvier 2021), et qui semble contenir des codages différents de ceux indiqués dans le message de Patrick Simon, comme le

Si le recodage de Coulmont et Simon apparaît globalement raisonnable, j'y apporte plusieurs corrections :

- ✓ je catégorise les prénoms « arméniens » (modalité 422 de la variable CAT1 en 82 modalités, comme « Sevan », « Anouch » et « Haig ») comme originaires d'« autre Europe » (comme les prénoms géorgiens) plutôt que comme « arabo-musulmans » ; en effet la langue arménienne est indo-européenne et nullement apparentée à l'arabe, et surtout l'Arménie, massivement chrétienne, compte une proportion de musulmans qui est négligeable ; et pour cause : l'Arménie est le premier État du monde à avoir adopté le christianisme comme religion d'État, dès le 4<sup>e</sup> siècle ; la décision qu'ont prise Baptiste Coulmont, Professeur des universités et spécialiste des prénoms, et Patrick Simon, Directeur de recherche à l'INED et spécialiste des migrations, de catégoriser les prénoms arméniens comme « arabo-musulmans » constitue une erreur, qui heureusement porte sur un effectif de prénoms trop faible pour risquer d'impacter leurs résultats ;
- ✓ je catégorise les prénoms « érythréens et éthiopiens (amhariques) » (modalité 224 de la variable CAT1 en 82 modalités, comme « Yohannes », <sup>115</sup> « Zemzem » et « Nesha ») comme « africains », plutôt que comme « arabo-musulmans » ; en effet, si une part non négligeable d'Érythréens et d'Éthiopiens sont musulmans, les prénoms amharas (par contraste avec les prénoms oromos ou somalis) sont, eux, assurément chrétiens (orthodoxes) et non pas musulmans ;
- ✓ je catégorise les prénoms « grecs » (modalité 134 de la variable CAT1 en 82 modalités, comme « Ilias », « Thalia » et « Ioannis » <sup>116</sup>) comme originaires d'« autre Europe », plutôt que comme « latins » ; et enfin
- ✓ je renomme « autre Europe » la catégorie que Coulmont et Simon appellent « autre groupe » ou « divers », puisqu'elle regroupe 25 modalités correspondant très principalement à des pays d'Europe (hors France, pays latins et pays anglophones), principalement des pays d'Europe septentrionale, centrale et orientale (élargis au Caucase chrétien, à Israël et à la Polynésie). <sup>117</sup>

---

regroupement des prénoms suédois, scandinaves, russes, géorgiens et israéliens avec les prénoms « internationaux ou anglophones » plutôt qu'avec les prénoms d'« autre Europe » (Annexe 4B, p. 124).

<sup>115</sup> « Yohannes » est la forme amharique du prénom « Jean ».

<sup>116</sup> « Ioannis » est la forme grecque moderne du prénom « Jean ».

<sup>117</sup> Cette nomenclature de l'INED comporte au final les 8 modalités suivantes (en gras), qui elles-mêmes regroupent les 82 catégories ci-dessous, exprimées en langue anglaise — mes corrections au recodage de Coulmont et Simon étant indiquées en barré (pays retiré de la modalité) et en italiques (pays ajouté à la modalité) :

- « **français** » : France ;
- « **latins** » : ~~Greece~~ ; Italy ; Portugal ; Spain ; Argentine ; Brazil ; "Spain, Portugal, Latin America, Philippines, Angola, Mozambique (Spanish, Portuguese, in Latin America also Italian first names)";
- « **autre Europe** » / « **autre groupe** » : Germany ; Albania ; Bosnia ; Bulgaria ; Danmark ; Finland ; ? ; Serbo-Croat ; Latvia ; Lithuania ; Republic of Macedonia ; Netherlands ; Norway ; Poland ; Romania ; Sweden ; Russia ; Czechia ; Hungary ; Ukraine ; White Russia ; Georgia ; Israel ; Scandinavia ; Polynesia (Polynesian Pacific) ; *Greece ; Armenia* ;
- « **internationaux ou anglophones** » : Ireland ; Great Britain ; USA ; Prénoms largement diffusés / First names common in many countries (not only European), some very widely used (Maria), some less (Alina) ; Prénoms internationaux (principalement anglo-américains) / International names mainly from the USA ;
- « **arabo-musulmans** », y compris **turcs** : Turkey ; Algeria ; ~~Eritrea / Ethiopia (Amharic names)~~ ; Morocco ; ~~Armenia~~ ; Azerbaijan ; Iran ; Kyrgyzstan ; Arabic speaking states ; Etats de langue arabe (hors Maghreb) / Libya, Egypt, Jordan, Syria, Iraq, Palestine, Arabian Peninsula ; Middle East (Iran, Afghanistan, Pakistan, Bangladesh, Indian Muslims) ; Muslim without any further classification (names like Ahmed, Mohammed) ;
- « **africains** » : Ivory Coast ; Nigeria ; Ghana ; Kenya ; Comores ; Democratic Republic of Congo ; Madagascar ; Mali ; Cameroon ; South Africa ; Senegal ; Togo ; Africa except North Africa (Maghreb,

C'est cette nomenclature ainsi recodée et corrigée que j'appelle, dans le présent document, la « nomenclature de l'INED », et que j'applique non seulement aux données de l'enquête TeO mais aussi au fichier INSEE des prénoms. Mais une fois expurgée des recodages erronés de Coulmont et Simon, cette nomenclature des prénoms par origine de l'INED est-elle fiable ? Pour le savoir, je la compare à une autre nomenclature, dite « synthétique ».

### ❖ La nomenclature synthétique des prénoms par origine

Dans son ouvrage de synthèse sur les prénoms, Coulmont (2014) fait référence à plusieurs articles scientifiques qui utilisent une nomenclature des prénoms par origine. Parmi eux, certains portent sur la France :

- d'une part, un article cherchant à expliquer le choix du type de prénom (prénom de saint catholique, prénom américain ou prénom arabe) en France des années 1950 aux années 2000 (Head et Mayer 2008, cité par Coulmont 2014, p. 48), prolongé par un autre article qui cherche à estimer le surcroît de bien-être associé — pour les immigrés du Maghreb et leurs descendants en France — au fait de transmettre à leurs enfants un prénom arabe (plutôt que non-arabe), dans les années 2000 (Algan et al. 2013, cité par Coulmont 2014, p. 44) ;
- d'autre part, un article qui estime l'effet de l'anonymisation des *curriculum vitae* — c'est-à-dire, de la dissimulation des prénoms des candidats — sur la probabilité, pour les porteurs d'un prénom non-français, d'obtenir un entretien d'embauche en France en 2010 (Behaghel et al. 2015, dont une version précédente est citée par Coulmont 2014, p. 82).

En outre, au même moment où paraissait l'article de Coulmont et Simon (avril 2019), un ouvrage utilisant lui aussi une nomenclature des prénoms par origine était publié :

- un ouvrage retraçant « l'archipelisation » de la société française, notamment à partir des prénoms attribués en France depuis le début du 20<sup>e</sup> siècle (Fourquet et Manternach 2019a).

J'ai demandé à ces auteurs de me communiquer leurs nomenclatures de prénoms par origine et, contrairement à Coulmont, Simon et l'INED, non seulement aucun n'a prétendu qu'il n'avait pas le droit de communiquer sa nomenclature,<sup>118</sup> mais en outre ils ont tous accepté.<sup>119</sup> J'utilise donc ces trois nomenclatures, ici appelées respectivement « nomenclature 1 », « nomenclature 2 » et « nomenclature 3 » (tableau 6.6). Pour mon travail, ces nomenclatures ont un double avantage : elles ont été conçues avant la parution de l'article de Coulmont et Simon (2019) et sont donc insoupçonnables d'avoir été conçues dans le but de démentir (ou de

---

Libya, Egypt); West Africa (Benin, Burkina Faso, Ivory Coast, Gambia, Ghana, Guinea, Guinea-Bissau, Liberia, Mali, Niger, Nigeria, Senegal, Sierra Leone, Togo) ; Eritrea / Ethiopia (Amharic names) ;

- « **asiatiques** » : Myanmar; Sri Lanka; Vietnam; India; Indonesia; Japan; Cambodia; Laos; Mongolia; Pakistan; Philippines; Korea; Thailand; China; Tibet; Far East / Indochina (China, Myanmar, Thailand, Vietnam, Cambodia, Laos, Korea);
- « **inclassable** » : ?; Impossible à classer / unclassified but not French (not enough information for a further classification).

<sup>118</sup> La « [Charte nationale de déontologie des métiers de la recherche](#) » du 26 janvier 2015, dont l'INED est signataire, indique : « La description détaillée du protocole de recherche, dans le cadre de cahiers de laboratoire ou de tout autre support, doit permettre la reproductibilité des travaux expérimentaux. Tous les résultats bruts (qui appartiennent à l'institution) ainsi que l'analyse des résultats doivent être conservés de façon à permettre leur vérification. » Sur les règles de l'intégrité scientifique, voir Chapitre 4, p. 15.

<sup>119</sup> Les diverses thèses soutenues dans le présent document n'engagent personne d'autre que moi et je suis seul responsable de toute erreur. Je tiens toutefois à remercier chaleureusement Clément Malgouyres (nomenclature 1), Thomas Le Barbanchon (nomenclature 2) et Sylvain Manternach (nomenclature 3), ainsi que tous leurs coauteurs, d'avoir accepté de me communiquer leurs nomenclatures de prénoms par origine. Je tiens à remercier particulièrement Sylvain Manternach, qui a accepté de me communiquer sa nomenclature alors même qu'il n'est pas membre d'une institution de recherche publique et n'était donc pas tenu de le faire.

confirmer) le caractère raisonnable de leur nomenclature ; et elles ont également donné lieu à des publications scientifiques de qualité, elles-mêmes favorablement citées par Coulmont (2014), si bien qu'elles sont peu susceptibles d'être contestées par lui et son coauteur.

La nomenclature 1, qui indique l'origine de 4 045 prénoms masculins et féminins (« arabe », « africain » et « arabe et africain »), permet de catégoriser le prénom de 7 % des individus de l'échantillon maximal pondéré (N = 83 045). La nomenclature 2, qui avant nettoyage<sup>120</sup> indiquait l'origine de 23 388 prénoms, classe quant à elle l'origine de 22 620 prénoms masculins et féminins (« français », « méditerranéen », « britannique », « germanique », « européen », « africain », « asiatique », « musulman » et « juif »), ce qui lui permet à elle seule de couvrir le prénom de 91 % des individus de l'échantillon maximal pondéré. Enfin la nomenclature 3, qui avant nettoyage indiquait l'origine de 15 480 prénoms masculins, catégorise 14 664 prénoms masculins (« arabo-musulman », « autre », « arabo-musulman et autre »), ce qui lui permet de couvrir le prénom de 48 % des individus de l'échantillon maximal pondéré (précisément, le prénom de 96 % des hommes mais aucun prénom féminin).

Je rassemble les informations de ces 3 nomenclatures pour n'en constituer qu'une seule, capable de répondre au double objectif suivant :

- d'une part, il s'agit de maximiser la proportion des individus de l'enquête TeO dont le prénom se voit attribuer une origine, plutôt que de voir son origine non renseignée<sup>121</sup> ; la nomenclature synthétique des prénoms permet ainsi de catégoriser l'origine de 94 % des individus du champ retenu, c'est-à-dire que seuls 6 % des individus du champ retenu voient l'origine de leur prénom non renseignée (tableau 6.7) ;
- d'autre part, il m'est apparu important de coller « au plus près » aux modalités d'origine de la nomenclature de Coulmont et Simon (2019), sans toutefois s'interdire de les raffiner<sup>122</sup> ; pour ce faire, je conserve la plupart des modalités

<sup>120</sup> Le « nettoyage » harmonisé des prénoms des 3 nomenclatures consiste à supprimer les doublons, les traits d'union, les apostrophes, les accents, etc. initialement présents dans certains prénoms des nomenclatures mais pas dans les prénoms des tables TeO, dans le but de pouvoir appairer les prénoms des nomenclatures à un maximum de prénoms de l'enquête TeO.

<sup>121</sup> Précisément, la logique de la construction de la variable ORIG\_PRENOM\_IND (origine du prénom de l'individu), à partir des 3 nomenclatures de prénoms par origine, est la suivante, en 7 étapes : (i) si un prénom simple est non renseigné ou ne voit son origine catégorisée par aucune des 3 nomenclatures, ou si un prénom composé de 2 à 4 « composants » (termes séparés par un trait d'union, un espace, une apostrophe, etc.) ne voit son origine catégorisée par aucune des 3 nomenclatures et qu'aucun de ses composants ne voit son origine catégorisée par l'une des nomenclatures, alors je le catégorise comme « inclassable » ; (ii) si un prénom simple ne voit son origine catégorisée que par une seule des 3 nomenclatures, alors je le catégorise de ladite origine ; (iii) si un prénom simple voit son origine catégorisée par au moins 2 des 3 nomenclatures et que ces origines sont les mêmes, alors je le catégorise de ladite origine ; (iv) si un prénom simple voit son origine catégorisée par au moins 2 des 3 nomenclatures mais que ces origines ne sont pas les mêmes, alors je le catégorise d'une origine « mixte » ou « ambiguë », constituée des 2 ou 3 origines différentes attribuées par les nomenclatures ; (v) si un prénom composé constitué de 2 « composants » ne voit son origine catégorisée par aucune des 3 nomenclatures, mais qu'au moins l'un de ses composants voit son origine catégorisée par au moins l'une des nomenclatures, alors je le catégorise selon l'origine du premier des composants dont l'origine est catégorisée ; (vi) si un prénom composé constitué de 3 « composants » ne voit son origine catégorisée par aucune des 3 nomenclatures, mais qu'au moins l'un de ses composants voit son origine catégorisée par au moins l'une des nomenclatures, alors je le catégorise selon l'origine du premier des composants dont l'origine est catégorisée ; (vii) si un prénom composé constitué de 4 « composants » ne voit son origine catégorisée par aucune des 3 nomenclatures, mais qu'au moins l'un de ses composants voit son origine catégorisée par au moins l'une des nomenclatures, alors je le catégorise selon l'origine du premier des composants dont l'origine est catégorisée. Pour plus de précisions, voir Annexe 5B, p. 183.

<sup>122</sup> Le programme ci-dessus aboutit à la création de 23 modalités : les 9 modalités qui sont finalement retenues (« français », « latin », « international ou anglophone », « autre Europe », « arabo-musulman » (y compris turc), « africain », « asiatique », « juif » et « inclassable »), mais aussi 14 autres modalités « mixtes » ou « ambiguës » qui au total caractérisent les prénoms de moins de 1 % des individus de l'échantillon maximal TeO et qui sont

d'origine de la nomenclature 2, qui sont les plus proches de celles de la nomenclature de Coulmont et Simon (2019), mais je garde aussi la modalité des prénoms « juifs » dont on a vu l'importance dans l'histoire des prénoms en Europe et au Maghreb (Annexe 6A, p. 267 et encadré 6.5) ;

- ainsi la nomenclature synthétique des prénoms par origine comprend-elle 9 modalités de nature ethnoreligieuse : « français », « latin », « international ou anglophone », « autre Europe », « arabo-musulman », « africain », « asiatique », « juif » et « inclassable ». <sup>123</sup> La principale source de cette nomenclature synthétique de prénoms par origine se trouve être la nomenclature 2, elle-même en partie fondée sur *L'indispensable manuel orthographique des prénoms français d'origine arabe et musulmane* (Chebira 2005).

C'est cette nomenclature ainsi recodée que j'appelle, dans la suite du présent document, la « nomenclature synthétique » et que j'applique non seulement aux données de l'enquête TeO mais aussi au fichier INSEE des prénoms.

---

donc trop pauvres en prénoms et en individus pour être utiles. Ces 14 modalités sont alors recatégorisées selon la façon dont le site [behindthename](#) et les portails [anglophone](#) et [francophone](#) d'anthroponymie de Wikipedia catégorisent leur(s) prénom(s) le(s) plus fréquent(s), comme suit : (i) la modalité « français et arabo-musulman », qui ne contient que 14 prénoms portés par 147 individus (« Sofiane », « Sophiane », « Omer », etc.), est recatégorisée en « arabo-musulman » ; (ii) la modalité « français et africain », qui ne contient que 4 prénoms portés par 26 individus (« Candice », « Marième », « Dédé » et « Dior »), est recatégorisée en « africain » ; (iii) la modalité « latin et arabo-musulman », qui ne contient que 16 prénoms portés par 178 individus (« Aminata », « Manel », « Orane », etc.), est recatégorisée en « arabo-musulman », même si « Aminata » aurait plutôt dû être catégorisé comme « africain » ; (iv) la modalité « latin, arabo-musulman et africain », qui ne contient que 1 prénom porté par 34 individus (« Adama »), est recatégorisée en « africain » ; (v) la modalité « latin et africain », qui ne contient que 3 prénoms portés par 47 individus (« Binta », « Fanta » et « Diogo »), est recatégorisée en « africain » ; (vi) la modalité « international ou anglophone, et arabo-musulman », qui ne contient que 5 prénoms portés par 18 individus (« Mody », « Sammy », « Duran », etc.), est recatégorisée en « international ou anglophone » ; (vii) la modalité « international ou anglophone, et africain », qui ne contient que 2 prénoms portés par 5 individus (« Candy » et « Tiana »), est recatégorisée en « international ou anglophone » ; (viii) la modalité « autre Europe et arabo-musulman », qui ne contient que 7 prénoms portés par 31 individus (« Wassila », « Nada », « Liès », etc.), est recatégorisée en « arabo-musulman » ; (ix) la modalité « autre Europe et africain », qui contient 1 prénom porté par 2 individus (« Lala »), est recatégorisée en « autre Europe » ; (x) la modalité « arabo-musulman et africain », qui ne contient que 275 prénoms portés par 1 416 individus (« Mamadou », « Hadda », « Tidiane », etc.), est recatégorisée en « africain » ; (xi) la modalité « arabo-musulman, africain et asiatique », qui ne contient que 1 prénom porté par 1 individu (« s »), est recatégorisée en « inclassable » ; (xii) la modalité « arabo-musulman et asiatique », qui ne contient que 16 prénoms portés par 47 individus (« May », « Keltoum », « Hana », etc.), est recatégorisée en « arabo-musulman » ; (xiii) la modalité « arabo-musulman et juif », qui ne contient que 6 prénoms portés par 15 individus (« Aaron », « Nissim », « Rivka », etc.), est recatégorisée en « juif » ; et (xiv) la modalité « africain et asiatique », qui ne contient que 2 prénoms portés par 2 individus (« s » et « s », etc.), est recatégorisée en « asiatique ».

<sup>123</sup> Pour une définition de ces catégories, voir Annexe 6A, p. 267.



**Tableau 6.6. Comparaison entre les nomenclatures de prénoms par origine**

	Nomenclature de l'INED	3 nomenclatures composant la nomenclature « synthétique »		
		Nomenclature 1	Nomenclature 2	Nomenclature 3
<b>Source</b>	Coulmont et Simon 2019, d'après <a href="#">Humpert et Schneiderheinze (2000)</a>	Head et Mayer 2008 et Algan et al. 2013, d'après Jouniaux L., <i>Les 20 000 plus beaux prénoms du monde</i> , Hachette Pratique (2001)	Behaghel et al. 2015, d'après Chebira A., <i>L'indispensable manuel orthographique des prénoms français d'origine arabe et musulmane</i> , APIC (2005) et d'autres sources étymologiques	Fourquet et Manternach 2019a, d'après les données de listes électorales de grandes villes françaises (lieu de naissance et nom de famille) et d'autres sources généalogiques et onomastiques
<b>Nombre de prénoms</b>	13 203 prénoms masculins et féminins	4 045 prénoms masculins et féminins	22 620 prénoms masculins et féminins	14 664 prénoms masculins
<b>Catégories d'origine des prénoms</b>	8 catégories : « français », « latin », « internationaux ou anglophones », « arabo-musulman », « africain », « asiatique », « autre groupe », « inclassable »	3 catégories : « arabe » (N=3 396), « africain » (N=577), « arabe et africain » (N=72)	9 catégories : « français » (N=6 725), « méditerranéen » (N=3 449), « britannique » (N=781), « germanique » (N=472), « européen » (N=987), « africain » (N=403), « asiatique » (N=2 095), « musulman » (N=7 581), « juif » (N=127)	3 catégories : « arabo-musulman » (N=4 375), « autre » (N=10 267) et « arabo-musulman et autre » (N=22)
<b>Taux de couverture des prénoms de l'échantillon maximal TeO pondéré (N=83 045)</b>	97 %	7 %	91 %	48 %

### ❖ Comparaison entre la nomenclature des prénoms par origine de l'INED et la nomenclature synthétique

La nomenclature de l'INED peut-elle être considérée comme fiable, en ce sens qu'elle concorde bien avec la nomenclature synthétique, conçue indépendamment ? Globalement, oui. En effet, quelle que soit la nomenclature utilisée, on note que 73 % de l'échantillon TeO retenu pondéré (représentatif des personnes de 18-50 / 18-60 ans en ménage ordinaire et de leurs enfants de 0-17 ans habitant en France métropolitaine en 2008) portent un prénom « français », 2 % portent un prénom originaire d'« autre Europe », 8 % portent un prénom « arabo-musulman », moins de 1 % portent un prénom « asiatique » et jusqu'à 1 % portent un prénom « africain » (tableau 6.7). Si l'on applique ces nomenclatures non plus à l'enquête TeO (2008) mais au fichier INSEE des prénoms (1900-2019), il apparaît là encore que la nomenclature de l'INED produit des résultats globalement similaires à ceux de la nomenclature synthétique : quelle que soit la nomenclature utilisée, 76 % à 77 % des nouveau-nés en France de 1900 à 2019 se sont vus attribuer un prénom « français », 1 % un prénom originaire d'« autre Europe », 4 % à 5 % un prénom « arabo-musulman », et moins de 1 % un prénom « africain » ou « asiatique » (tableau 6.7).

Globalement, les deux nomenclatures catégorisent de façon identique l'origine du prénom de 86 % de l'échantillon TeO retenu pondéré (tableau 6.8, résultat obtenu en additionnant les cases grises). Ces deux nomenclatures catégorisent aussi de façon identique l'origine des prénoms de 87 % des nouveau-nés en France de 1900 à 2019 (tableau non présenté). En outre, les divergences entre les deux nomenclatures ne présentent pas de biais manifeste (tableau 6.8). Au contraire, la nomenclature de l'INED catégorise comme « français » 95 % des prénoms ainsi catégorisés par la nomenclature synthétique, et réciproquement la nomenclature synthétique catégorise comme « français » 95 % des prénoms ainsi catégorisés par la nomenclature de l'INED. La nomenclature de l'INED catégorise aussi comme « arabo-musulmans » 88 % des prénoms ainsi catégorisés par la nomenclature synthétique, et réciproquement la nomenclature synthétique catégorise comme « arabo-musulmans » 88 % des prénoms ainsi catégorisés par la nomenclature de l'INED.

La principale différence entre les deux nomenclatures est que celle de l'INED catégorise relativement moins de prénoms comme « latins » (3 % contre 5 %) et relativement plus de prénoms comme « internationaux ou anglophones » (12 % contre 6 %) (tableau 6.7). En effet, la nomenclature de l'INED ne catégorise pas comme « latins » certains prénoms ainsi catégorisés par la nomenclature synthétique, comme « Patricia », « Laura », « Clara », « Hugo » et « Sandra ». Et la nomenclature de l'INED catégorise comme « internationaux ou anglophones » des prénoms que la nomenclature synthétique catégorise autrement, comme « Patrick », « Patricia », « Anthony », « Laura » et « Sarah ».

Au-delà de ces différences, la nomenclature de l'INED permet de catégoriser l'origine des prénoms de 98 % de l'échantillon TeO retenu pondéré, contre « seulement » 94 % pour la nomenclature synthétique (tableau 6.7).<sup>124</sup> Précisément, la nomenclature de l'INED assure non seulement un meilleur taux de couverture global des prénoms, mais aussi une proportion de prénoms « inclassables » plus faible parmi les habitants de *toutes* origines, qu'ils soient du groupe « majoritaire » (1 % contre 3 %) ou originaires des DOM (5 % contre 10 %), d'Europe du Sud (2 % contre 5 %), d'autre Europe (3 % contre 10 %), du Maghreb (4 % contre 9 %), d'autre Afrique (14 % contre 26 %), de Turquie (7 % contre 27 %), d'Asie (18 % contre 43 %) ou d'ailleurs (12 % contre 19 %) (tableau 6.9). Les prénoms « inclassables » étant rares au point

---

<sup>124</sup> La nomenclature synthétique permet toutefois de catégoriser autant de prénoms du fichier INSEE que la nomenclature de l'INED (94 %) (tableau 6.7), parce que même si elle ignore plus de prénoms TeO que la nomenclature de l'INED, les nomenclatures 1, 2 et 3 catégorisent de nombreux prénoms absents de TeO alors que la nomenclature de l'INED ne catégorise que des prénoms présents dans TeO.

de ne pas pouvoir être mentionnés sans risque de réidentification,<sup>125</sup> il est clair que la nomenclature de l'INED couvre bien le champ des prénoms portés en métropole en 2008 (tableau 6.9). Cela dit, la proportion d'individus portant un prénom « inclassable » reste plus élevée parmi les personnes originaires de l'étranger qu'au sein du groupe « majoritaire », si bien que la nomenclature de l'INED (comme a fortiori la nomenclature synthétique) tend à sous-estimer la prévalence réelle des prénoms étrangers, particulièrement les prénoms de personnes originaires d'Afrique (hors Maghreb) et d'Asie, qui pour 14 % et 18 % d'entre eux sont catégorisés comme « inclassables » (tableau 6.9).

**Tableau 6.7. Origine des prénoms en France, selon la nomenclature de l'INED<sup>126</sup> et selon la nomenclature synthétique**

		Prénoms portés en 2008 (enquête TeO)		Prénoms attribués depuis 1900 (fichier INSEE des prénoms)	
		Nomenclature de l'INED	Nomenclature synthétique	Nomenclature de l'INED	Nomenclature synthétique
Origine du prénom des individus	<b>Français</b>	73 %	73 %	76 %	77 %
	<b>Latin</b>	3 %	5 %	2 %	5 %
	<b>International ou anglophone</b>	12 %	6 %	10 %	5 %
	<b>Autre Europe</b>	2 %	2 %	1 %	1 %
	<b>Arabo- musulman</b>	8 %	8 %	4 %	5 %
	<b>Africain</b>	1 %	0 %	0 %	0 %
	<b>Asiatique</b>	0 %	0 %	0 %	0 %
	<b>Juif</b>	-	1 %	-	1 %
	<b>Inclassable</b>	2 %	6 %	6 %	6 %
	<b>Total</b>	100 %	100 %	100 %	100 %

Lecture : d'après la nomenclature de l'INED comme d'après la nomenclature synthétique, 73 % des personnes de 18-50 / 18-60 ans en ménage ordinaire et de leurs enfants de 0-17 ans habitant en France métropolitaine en 2008 portent un prénom « français », et 76 % à 77 % des nouveau-nés en France de 1900 à 2019 se sont vus attribuer un prénom « français ».

Champ TeO : personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008, données pondérées, N = 42 208. Champ Fichier INSEE des prénoms : naissances en France entière (hors Mayotte) de 1900 à 2019, prénoms « rares » exclus, N = 84 264 960.

Source : TeO (voir Annexe 5B, p. 183).

De ces investigations méthodologiques, je conclus que non seulement la nomenclature de l'INED est globalement fiable en ce sens qu'elle concorde bien avec la nomenclature synthétique, mais qu'en outre elle permet de catégoriser l'origine de plus de prénoms portés par les individus de l'échantillon TeO retenu pondéré. C'est pourquoi, dans la reproduction « au plus près » des analyses de Coulmont et Simon (2019), j'utilise à mon tour la nomenclature de

<sup>125</sup> Conformément aux règles du secret statistique du [CASD](#) et de l'[INSEE](#), et pour qu'aucune information ne permette d'identifier directement ou indirectement une personne physique (ni d'inférer de nouvelles informations sur elle), les résultats obtenus sur des effectifs faibles (<5) sont notés « s » (comme « secret »).

<sup>126</sup> La « nomenclature de l'INED » est la nomenclature CAT1 communiquée par l'INED au réseau Quêtelet et au CASD, telle que recodée selon les indications fournies dans le message de Patrick Simon du 4 décembre 2020, mises à part les corrections précisées ci-dessus concernant les prénoms arméniens, amhariques, grecs, etc. Il ne s'agit pas de la nomenclature CAT0, elle aussi communiquée par l'INED au réseau Quêtelet et au CASD, mais dont on a vu qu'elle est fallacieuse (Chapitre 4, p. 15).



l'INED, et je n'utilise la nomenclature synthétique que pour analyser les prénoms « juifs ». Ainsi le caractère surprenant de certains des résultats de Coulmont et Simon (2019) ne peut pas s'expliquer par le caractère étonnant de la nomenclature des prénoms qu'ils m'ont communiquée. Les motifs de doute seraient plutôt à chercher du côté de leur programme informatique et/ou dans la base de données.

**Tableau 6.8. Origine des prénoms portés en France en 2008 (enquête TeO)**  
**Comparaison entre la nomenclature de l'INED (colonnes) et la nomenclature synthétique (lignes)**

		ORIGINE DU PRÉNOM DES INDIVIDUS (NOMENCLATURE DE L'INED)								
		Français	Latin	International ou anglophone	Autre Europe	Arabo-musulman	Africain	Asiatique	Inclassable	Total
ORIGINE DU PRÉNOM DES INDIVIDUS (NOMENCLATURE SYNTHÉTIQUE)	Français	69 %	0 %	3 %	0 %	0 %	0 %	0 %	0 %	73 %
	Latin	0 %	2 %	2 %	0 %	0 %	0 %	0 %	0 %	5 %
	International ou anglophone	1 %	0 %	4 %	0 %	0 %	0 %	0 %	0 %	6 %
	Autre Europe	0 %	0 %	0 %	1 %	0 %	0 %	0 %	0 %	2 %
	Arabo-musulman	0 %	0 %	0 %	0 %	7 %	0 %	0 %	0 %	8 %
	Africain	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
	Asiatique	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
	Juif	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	1 %
	Inclassable	1 %	0 %	1 %	0 %	1 %	0 %	0 %	2 %	6 %
	Total	73 %	3 %	12 %	2 %	8 %	1 %	0 %	2 %	100 %

Lecture : 69 % des habitants en France métropolitaine en 2008 portent un prénom que les 2 nomenclatures catégorisent comme « français ».

Champ : personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008, données pondérées, N = 42 208.

Source : TeO (voir Annexe 5B, p. 183).

**Tableau 6.9. Top 3 des prénoms « inclassables » selon l'origine de leur porteur en France en 2008 (enquête TeO)**  
**Comparaison entre la nomenclature de l'INED et la nomenclature synthétique**

ORIGINE GÉOGRAPHIQUE DES INDIVIDUS	Groupe « majoritaire »	NOMENCLATURE DE L'INED					NOMENCLATURE SYNTHÉTIQUE				
		Prénoms inclassables	Sexe	Prénom n°1	Prénom n°2	Prénom n°3	Prénoms inclassables	Sexe	Prénom n°1	Prénom n°2	Prénom n°3
	1 %		♂	s	s	s	3 %	♂	Killian	s	Evan
			♀	s	s	s		♀	s	Lou	Mégane
	5 %		♂	s	s	s	10 %	♂	s	s	s
			♀	s	s	s		♀	s	s	s
	2 %		♂	s	s	s	5 %	♂	s	Tiago	s
			♀	s	s	s		♀	s	Cassandra	s
	3 %		♂	s	s	s	10 %	♂	s	s	s
			♀	s	s	s		♀	s	s	s
	4 %		♂	s	s	s	9 %	♂	Ilan	s	s
			♀	s	s	s		♀	Alyssa	s	s
	14 %		♂	s	s	s	26 %	♂	s	s	s
			♀	s	s	s		♀	s	Assa	s
	7 %		♂	s	s	s	27 %	♂	s	s	s
			♀	s	s	s		♀	Beyza	Dilek	Merve
	18 %		♂	s	s	s	43 %	♂	s	s	s
			♀	s	s	s		♀	s	s	s
	12 %		♂	s	s	s	19 %	♂	s	s	s
			♀	s	s	s		♀	s	s	s
	-		♂	-	-	-	-	♂	-	-	-
			♀	-	-	-		♀	-	-	-
	2 %		♂	s	s	s	6 %	♂	Killian	s	Evan
			♀	s	s	s		♀	Fiona	Lou	Mégane

Lecture : 1 % du groupe « majoritaire » portent un prénom d'origine « inclassable » d'après la nomenclature de l'INED, aucun d'entre eux n'étant d'effectif suffisant pour pouvoir être mentionné (cf. NB ci-dessous).

Champ : individus du champ retenu (personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008) qui portent un prénom « inclassable », d'après la nomenclature de l'INED, données pondérées, N = 989, et d'après la nomenclature synthétique, données pondérées, N = 2 392.

Source : TeO (voir Annexe 5B, p. 183).

NB : Conformément aux règles du secret statistique du [CASD](#) et de l'[INSEE](#), et pour qu'aucune information ne permette d'identifier directement ou indirectement une personne physique (ni d'inférer de nouvelles informations sur elle), les résultats obtenus sur des effectifs faibles (<5) sont notés « s » (comme « secret »).

### ❖ Analyses préliminaires sur les origines des prénoms en France

Si la nomenclature des prénoms de l'INED est globalement fiable, comment décrit-elle l'évolution de l'origine des prénoms en France ? Si l'on examine l'origine des prénoms portés en 2008 par l'échantillon TeO retenu pondéré, selon la décennie de naissance de leur porteur, il apparaît que de la décennie 1950 à la décennie 2000 de moins en moins de prénoms attribués en France sont « français » (de 94 % dans les années 1950 à 49 % dans les années 2000), tandis que de plus en plus de prénoms attribués sont « internationaux ou anglophones » (de 4 % à 23 %) ou encore « arabo-musulmans » (de 1 % à 11 %) (figure 6.10a). Ces résultats, issus de la nomenclature des prénoms de l'INED, sont proches de ceux obtenus avec la nomenclature synthétique (figure 6.10b).

Si l'on examine maintenant l'origine des prénoms attribués en France depuis 1900 (fichier INSEE des prénoms), il apparaît que de 1900 à 2019 de moins en moins de prénoms attribués en France sont « français », la chute étant relativement spectaculaire, de 93 % en 1900 à 29 % en 2019 (figure 6.11a).<sup>127</sup> Parallèlement, de plus en plus de prénoms attribués sont « internationaux ou anglophones » (de 3 % à 20 %) ou encore « arabo-musulmans » (de 0 % à 14 %), mais aussi « inclassables » (de 3 % à 26 %)<sup>128</sup> (figure 6.11a). Là encore ces résultats, obtenus avec la nomenclature des prénoms de l'INED, sont relativement proches de ceux obtenus avec la nomenclature synthétique (figure 6.11b).

Si l'on se concentre sur l'évolution de la proportion de prénoms « arabo-musulmans » en France, il apparaît que la nomenclature de l'INED et la nomenclature synthétique produisent, sur l'échantillon TeO retenu pondéré, des résultats presque identiques : quelle que soit la nomenclature utilisée, la part de prénoms attribués qui sont « arabo-musulmans » est passée de 1 % dans les années 1950 à 11 % ou 12 % dans les années 2000 (figure 6.12a). En revanche, les deux nomenclatures produisent des résultats sensiblement différents sur le fichier INSEE des prénoms depuis les années 1990 : si la nomenclature synthétique estime que 18 % des prénoms attribués en 2019 sont « arabo-musulmans », la nomenclature de l'INED n'en compte que 14 % (figure 6.12b). On peut expliquer cet écart par le fait que la nomenclature de l'INED, contrairement à la nomenclature synthétique, ne catégorise pas comme « arabo-musulmans » des prénoms qui sont relativement fréquemment attribués depuis les années 1990, les plus fréquents d'entre eux étant, par ordre décroissant, « Nadia », « Inès », « Linda », « Yannis » et « Sara ».<sup>129</sup>

La plupart des prénoms que seule la nomenclature synthétique catégorise comme « arabo-musulmans » sont des prénoms féminins, parce que tel est le cas de la plupart des prénoms d'origine mixte ou ambiguë :

« la liste des prénoms masculins arabo-musulmans compte beaucoup moins de prénoms pouvant renvoyer ou non à la culture arabo-musulmane que la liste des prénoms féminins. Des prénoms comme Sonia, Nadia ou Inès peuvent, en effet, être donnés à leur fille par des parents n'ayant aucune ascendance immigrée » (Fourquet et Manternach 2019a, p. 135).

<sup>127</sup> Pour plus de précisions sur les causes de la forte baisse de la part de prénoms attribués qui sont « français », voir figure 9.2.

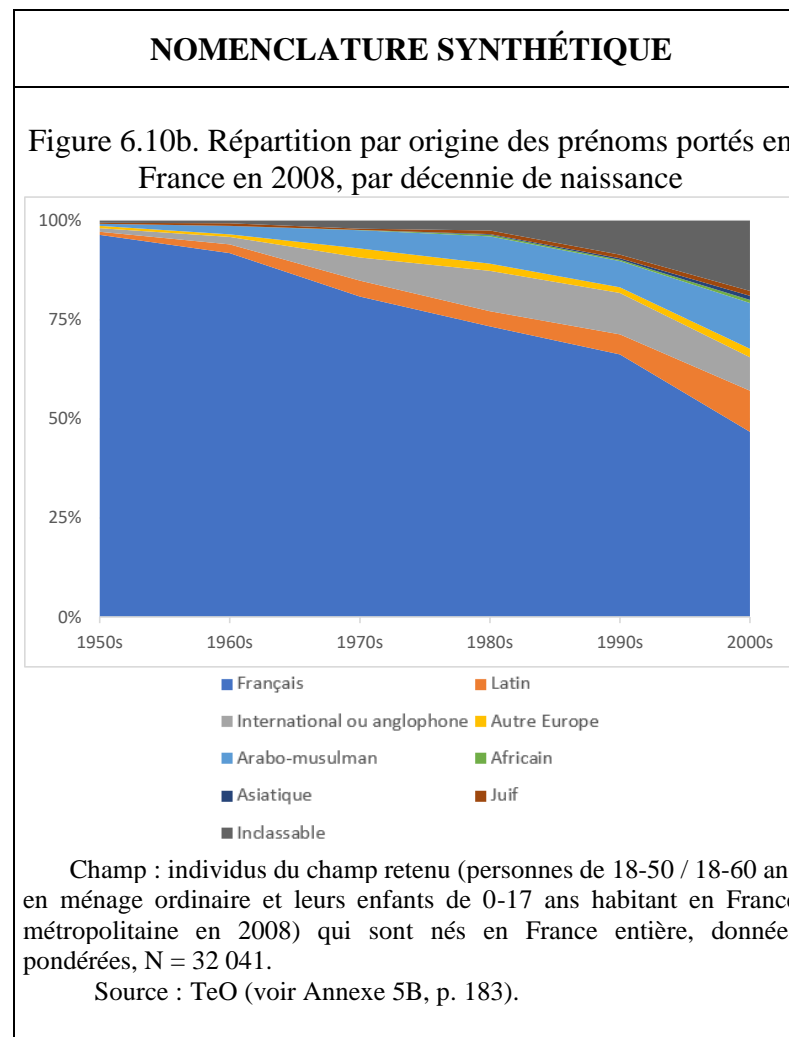
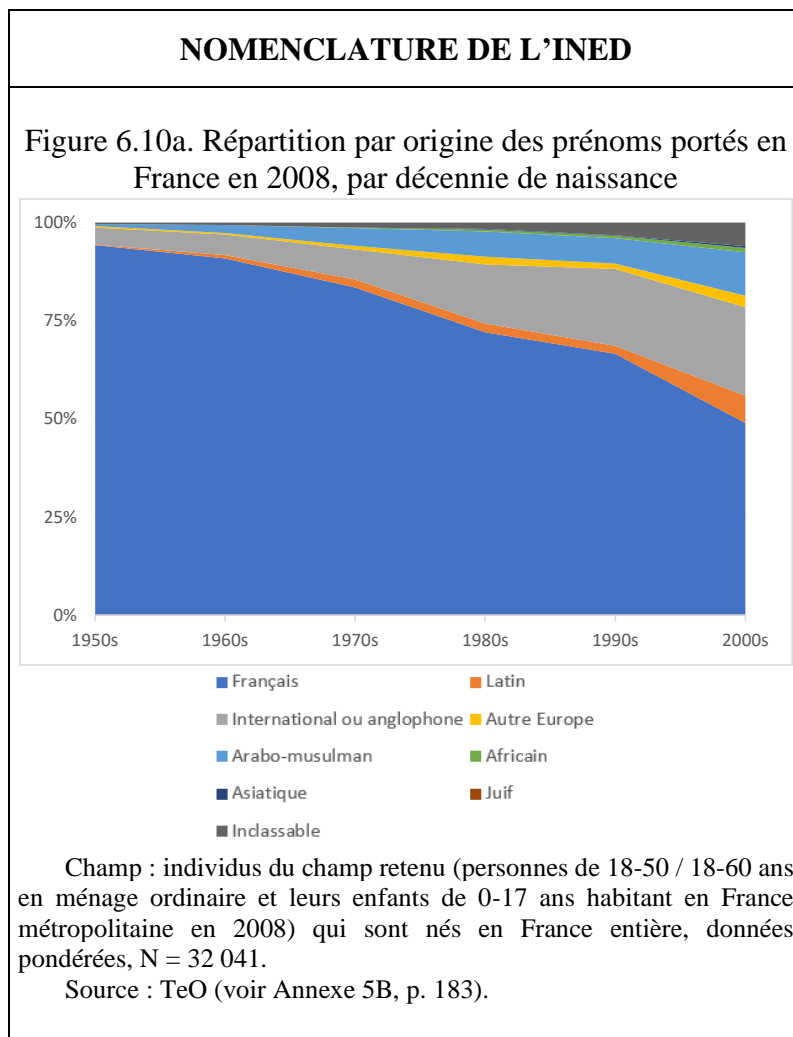
<sup>128</sup> Les prénoms « inclassables » les plus attribués en France à l'état civil entre 2000 et 2019 sont « Marceau », « Lilo », « Lyam », « Soan » et « Sasha » pour les prénoms masculins, et « Elea », « Lya », « Anae », « Lylou » et — encore — « Sasha » pour les prénoms féminins. Si la nomenclature de l'INED ne catégorise pas le prénom « Marceau » comme « français », c'est parce qu'elle ne catégorise pas ce prénom, absent de l'enquête TeO.

<sup>129</sup> Certes, la nomenclature de l'INED catégorise à l'inverse comme « arabo-musulmans » des prénoms que la nomenclature synthétique ne catégorise pas comme tels (comme « Lilou », « Monia », « Melyssa », « Gurvan » et « Lilia »), mais même depuis les années 1990 ces prénoms restent plus rarement attribués que les précédents.

Par conséquent, pour estimer la part de prénoms « arabo-musulmans » attribués en France — et plus généralement, pour estimer le poids de la population de culture arabo-musulmane — Fourquet et Manternach (2019a) proposent de n'étudier que les prénoms des nouveau-nés *masculins* (Fourquet et Manternach 2019a, p. 137). Il apparaît alors que, d'après leur nomenclature (la nomenclature 3 ci-dessus) appliquée au fichier INSEE des prénoms, la proportion de prénoms « arabo-musulmans » *parmi les naissances de garçons* est proche de 0 % jusqu'en 1950 puis dépasse 2 % dans les années 1960, 5 % dans les années 1970, 10 % dans les années 2000 et atteint 19 % en 2016 (Fourquet et Manternach 2019a, p. 136-140). De même, la proportion de prénoms « arabo-musulmans » *parmi les naissances de garçons* du fichier INSEE des prénoms atteint en 2019 15 % d'après la nomenclature de l'INED, et 20 % d'après la nomenclature synthétique.

## Origine des prénoms portés en France en 2008 (enquête TeO)

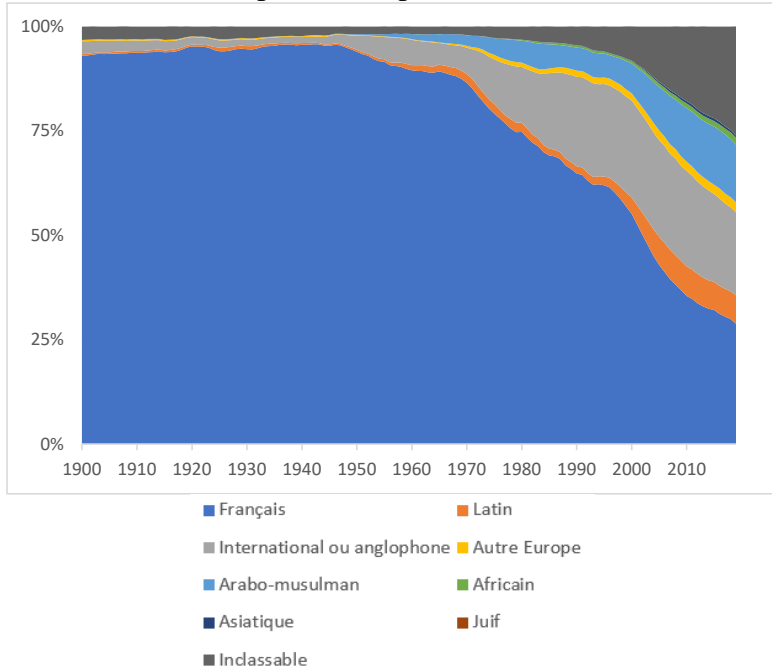
### Comparaison entre la nomenclature de l'INED et la nomenclature synthétique



## Origine des prénoms attribués en France depuis 1900 (fichier INSEE des prénoms) Comparaison entre la nomenclature de l'INED et la nomenclature synthétique

### NOMENCLATURE DE L'INED

Figure 6.11a. Répartition par origine des prénoms attribués en France depuis 1900, par année de naissance

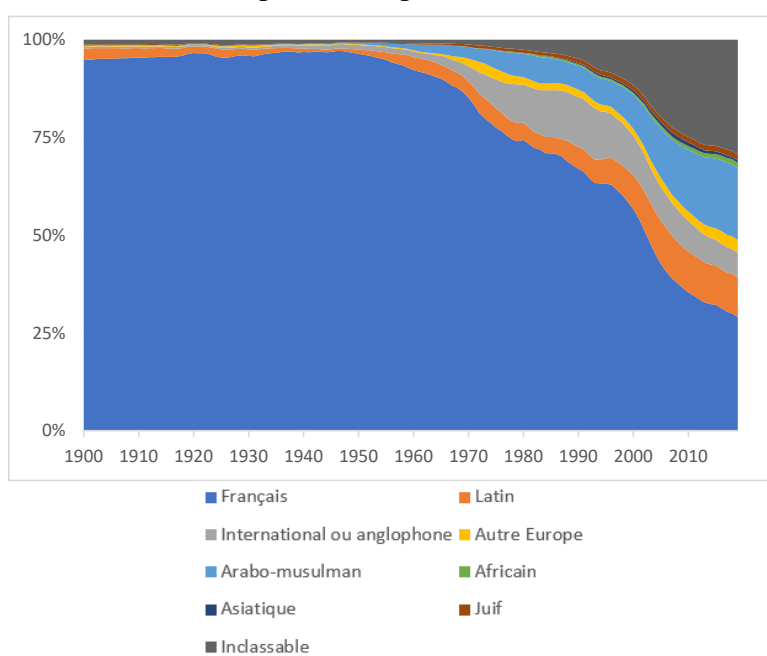


Champ : naissances en France entière (hors Mayotte) de 1900 à 2019, prénoms « rares » exclus, N = 84 264 960.

Source : [Fichier INSEE des prénoms](#).

### NOMENCLATURE SYNTHÉTIQUE

Figure 6.11b. Répartition par origine des prénoms attribués en France depuis 1900, par année de naissance



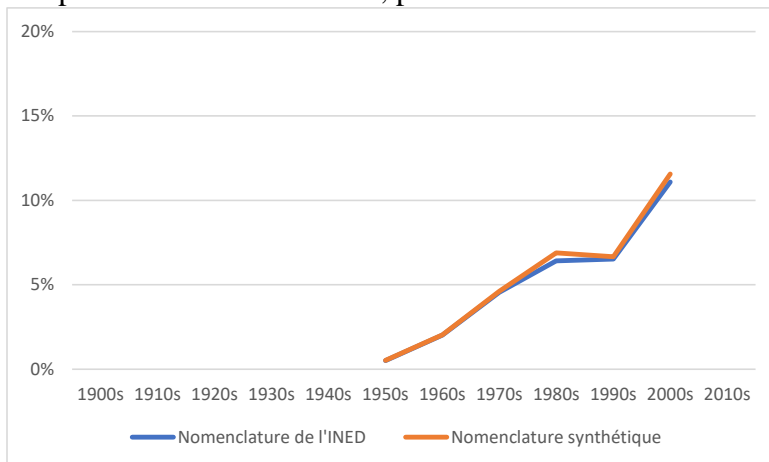
Champ : naissances en France entière (hors Mayotte) de 1900 à 2019, prénoms « rares » exclus, N = 84 264 960.

Source : [Fichier INSEE des prénoms](#).

## Proportion de prénoms en France qui sont « arabo-musulmans » Comparaison entre la nomenclature de l'INED et la nomenclature synthétique

### ENQUÊTE TeO

Figure 6.12a. Proportion de prénoms « arabo-musulmans » portés en France en 2008, par décennie de naissance

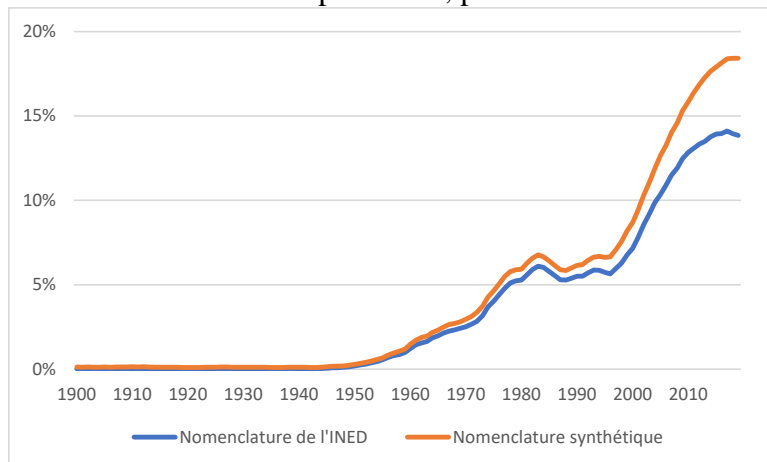


Champ : individus du champ retenu (personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008) qui sont nés en France entière, données pondérées, N = 32 041.

Source : TeO (voir Annexe 5B, p. 183).

### FICHIER INSEE DES PRÉNOMS

Figure 6.12b. Proportion de prénoms « arabo-musulmans » attribués en France depuis 1900, par année de naissance



Champ : naissances en France entière (hors Mayotte) de 1900 à 2019, N = 85 047 407.

Source : [Fichier INSEE des prénoms](#).



### ❖ Origine du prénom et origine de son porteur en France : une analyse

Qu'est-ce que disent les prénoms en France métropolitaine en 2008 des origines culturelles et de l'assimilation des immigrés et de leurs descendants ? Quelles sont les origines des prénoms des habitants, selon qu'ils sont de la population « majoritaire » ou originaires d'Europe, du Maghreb, d'Afrique, d'Asie, etc. (tableau 6.13a) ? Réciproquement, de quelles origines géographiques proviennent les habitants portant des prénoms « français », « latins », « arabo-musulmans », etc. (tableau 6.13b) ? Quels sont les prénoms les plus couramment portés en France en 2008, selon l'origine géographique de leur porteur (tableau 6.14a) et selon l'origine du prénom lui-même (tableau 6.14b) ? Enfin, quelles sont les origines des prénoms selon que leur porteur est d'origine chrétienne, musulmane, juive, etc. (tableau 6.15a), et réciproquement quelles sont les origines religieuses des habitants selon l'origine de leur prénom (tableau 6.15b) ?<sup>130</sup>

Les six tableaux suivants sont calculés sur l'échantillon TeO retenu pondéré (N = 42 208), représentatif des personnes de 18-50 / 18-60 ans en ménage ordinaire et de leurs enfants de 0-17 ans habitant en France métropolitaine en 2008. Les quatre premiers sont calculés avec la nomenclature de l'INED, tandis que les deux derniers — qui incluent la catégorie des prénoms « juifs » — ne peuvent être calculés qu'avec la nomenclature synthétique. Je commente ces tableaux à l'aide de plusieurs sources d'information : la revue de littérature sur l'histoire comparée des prénoms (Annexe 6A, p. 267) ; des collègues chercheurs, spécialistes d'onomastique ou de telle ou telle aire géographique ou linguistique ; et le site [behindthename](http://behindthename.com) ainsi que les portails [anglophone](http://anglophone.wiki) et [francophone](http://francophone.wiki) d'anthroponymie de Wikipedia. Je présente les résultats détaillés par origine géographique, avant de proposer une synthèse globale.

---

<sup>130</sup> L'enquête TeO renseigne la religion autodéclarée par l'enquêté, ainsi que celle de son conjoint cohabitant, de son éventuel premier conjoint et celle de ses parents (variable 21 du tableau 5.1). Le codage de la variable d'origine religieuse des individus (ORIG\_RELIG\_IND, variable 23 du tableau 5.1) est réalisé en 4 étapes : (i) si l'individu est descendant d'un seul parent affilié à une religion (soit christianisme, soit islam, soit judaïsme, soit bouddhisme ou hindouisme, etc.), c'est-à-dire si son autre parent n'est affilié à aucune religion, on lui associe cette origine religieuse ; (ii) si l'individu est descendant de deux parents affiliés à la même religion, on lui associe cette origine religieuse ; (iii) si l'individu est descendant de deux parents affiliés à des religions différentes, on lui associe — conformément à la convention adoptée dans l'enquête TeO pour la variable d'origine (géographique) des individus (ORIG\_IND, variable 11 du tableau 5.1) — la religion de son père ; (iv) si aucun des parents de l'individu n'est affilié à une religion (ils sont tous deux sans religion ou sans réponse), on ne lui associe aucune origine religieuse. Au final, sur l'échantillon maximal non pondéré, cette variable d'origine religieuse est renseignée pour 59 082 individus (sur 83 045).

**Tableau 6.13a. Origine du prénom (colonnes) selon l'origine géographique de son porteur (lignes), en France en 2008 (enquête TeO)**

		N	ORIGINE DU PRÉNOM DES INDIVIDUS (NOMENCLATURE DE L'INED)								
			Français	Latin	International ou anglophone	Autre Europe	Arabo- musulman	Africain	Asiatique	Inclassable	Total
ORIGINE GÉOGRAPHIQUE DES INDIVIDUS	Groupe « majoritaire »	31 775	84 %	2 %	11 %	1 %	1 %	0 %	0 %	1 %	100 %
	DOM	616	68 %	3 %	19 %	2 %	1 %	1 %	0 %	5 %	100 %
	Europe du Sud	2 704	58 %	20 %	17 %	2 %	1 %	0 %	0 %	2 %	100 %
	Autre Europe	1 538	59 %	3 %	20 %	13 %	2 %	0 %	1 %	3 %	100 %
	Maghreb	3 155	10 %	2 %	8 %	1 %	74 %	0 %	0 %	4 %	100 %
	Autre Afrique	1 025	32 %	5 %	13 %	2 %	12 %	21 %	1 %	14 %	100 %
	Turquie	415	11 %	0 %	3 %	1 %	78 %	0 %	0 %	7 %	100 %
	Asie	577	29 %	1 %	10 %	4 %	8 %	0 %	29 %	18 %	100 %
	Ailleurs	403	42 %	12 %	19 %	1 %	15 %	0 %	0 %	12 %	100 %
	Non réponse	0	-	-	-	-	-	-	-	-	-
	Total	42 208	73 %	3 %	12 %	2 %	8 %	1 %	1 %	2 %	100 %

Lecture : 84 % des individus du groupe « majoritaire » portent un prénom « français » (d'après la nomenclature de l'INED).

Champ : personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008, données pondérées, N = 42 208.

Source : TeO (voir Annexe 5B, p. 183).

NB : les données sous-jacentes à ce tableau contiennent une erreur qui, une fois corrigée, fait passer la part d'immigrés et de descendants d'immigrés du Maghreb qui portent un prénom « français » de 10 % à 5 %, et la part de ceux qui portent un prénom « arabo-musulman » de 74 % à 80 % (Chapitre 8, p. 46).

**Tableau 6.13b. Origine géographique de l'individu (lignes) selon l'origine de son prénom (colonnes), en France en 2008 (enquête TeO)**

		ORIGINE DU PRÉNOM DES INDIVIDUS (NOMENCLATURE DE L'INED)								
		Français	Latin	International ou anglophone	Autre Europe	Arabo-musulman	Africain	Asiatique	Inclassable	Total
<b>N</b>		30 713	1 271	4 899	662	3 217	248	209	989	42 208
<b>ORIGINE GÉOGRAPHIQUE DES INDIVIDUS</b>	<b>Groupe « majoritaire »</b>	87 %	38 %	71 %	52 %	8 %	3 %	6 %	40 %	75 %
	<b>DOM</b>	1 %	1 %	2 %	2 %	0 %	2 %	0 %	3 %	1 %
	<b>Europe du Sud</b>	5 %	44 %	10 %	6 %	1 %	0 %	2 %	5 %	6 %
	<b>Autre Europe</b>	3 %	3 %	6 %	29 %	1 %	0 %	6 %	5 %	4 %
	<b>Maghreb</b>	1 %	5 %	5 %	3 %	73 %	5 %	2 %	13 %	7 %
	<b>Autre Afrique</b>	1 %	4 %	3 %	3 %	4 %	90 %	4 %	15 %	2 %
	<b>Turquie</b>	0 %	0 %	0 %	1 %	10 %	0 %	0 %	3 %	1 %
	<b>Asie</b>	1 %	1 %	1 %	3 %	1 %	0 %	80 %	10 %	1 %
	<b>Ailleurs</b>	1 %	4 %	2 %	1 %	2 %	0 %	0 %	5 %	1 %
	<b>Non réponse</b>	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
	<b>Total</b>	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %

Lecture : 87 % des individus porteurs d'un prénom « français » (d'après la nomenclature de l'INED) sont originaires du groupe « majoritaire ».

Champ : personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008, données pondérées, N = 42 208.

Source : TeO (voir Annexe 5B, p. 183).

**Tableau 6.14a. Top 5 des prénoms selon l'origine géographique de leur porteur, en France en 2008 (enquête TeO)**

		N	Sexe	Prénom n°1	Prénom n°2	Prénom n°3	Prénom n°4	Prénom n°5
ORIGINE GÉOGRAPHIQUE DES INDIVIDUS	Groupe « majoritaire »	31 775	♂	Sébastien	Patrick	Michel	Éric	Philippe
			♀	Catherine	Martine	Nathalie	Isabelle	Françoise
	DOM	616	♂	David	Patrick	Olivier	Thierry	Alain
			♀	Nathalie	Stéphanie	Véronique	Aurélie	Karine
	Europe du Sud	2 704	♂	José	Daniel	Manuel	David	Antonio
			♀	Maria	Isabelle	Marie-Carmen	Céline	Élisabeth
	Autre Europe	1 538	♂	Jacques	Jean-Claude	Michel	Pierre	Nicolas
			♀	Catherine	Anne	s	Nadège	Sandra
	Maghreb	3 155	♂	Mohamed	Karim	Ahmed	Rachid	Farid
			♀	Fatima	Sarah	Nadia	Myriam	Samira
	Autre Afrique	1 025	♂	Marcellin	Jordan	Mamadou	Mohamed	Kevin
			♀	Aurélie	s	Aminata	Mariam	Fatoumata
	Turquie	415	♂	Mehmet	Ali	Mustafa	Ibrahim	Murat
			♀	Murielle	Fatma	Ayse	Hatice	Emine
	Asie	577	♂	David	Nicolas	Vincent	Michel	Kevin
			♀	Marie	Sophie	s	Nina	Zara
	Ailleurs	403	♂	Thomas	Mario	Paul	Georges	Jacques
			♀	Paulette	Emilie	Delphine	Sandra	Sylvie
	Non réponse	0	♂	-	-	-	-	-
			♀	-	-	-	-	-
	Total	42 208	♂	Sébastien	Patrick	Michel	Nicolas	Stéphane
			♀	Catherine	Nathalie	Martine	Isabelle	Françoise

Lecture : les prénoms les plus portés par le groupe « majoritaire » sont « Sébastien » et « Catherine ».

Champ : personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008, données pondérées, N = 42 208.

Source : TeO (voir Annexe 5B, p. 183).

NB : Conformément aux règles du secret statistique du [CASD](#) et de l'[INSEE](#), et pour qu'aucune information ne permette d'identifier directement ou indirectement une personne physique (ni d'inférer de nouvelles informations sur elle), les résultats obtenus sur des effectifs faibles (<5) sont notés « s » (comme « secret »).

**Tableau 6.14b. Top 5 des prénoms selon leur origine, en France en 2008 (enquête TeO)**

ORIGINE DU PRÉNOM DES INDIVIDUS (NOMENCLATURE DE L' INED)		N	Sexe	Prénom n°1	Prénom n°2	Prénom n°3	Prénom n°4	Prénom n°5
	Français	30 713	♂	Sébastien	Michel	Nicolas	Stéphane	Philippe
			♀	Catherine	Nathalie	Martine	Isabelle	Françoise
	Latin	1 271	♂	Enzo	José	Manuel	Antonio	Teo
			♀	Sabrina	Inès	Lola	Luna	Amelia
	International ou anglophone	4 899	♂	Patrick	Anthony	Kevin	Cédric	Jérémy
			♀	Patricia	Audrey	Laura	Sarah	Vanessa
	Autre Europe	662	♂	Dimitri	Axel	s	Ivan	Kilian
			♀	Ingrid	Johanna	Tatiana	s	Karin
	Arabo-musulman	3 217	♂	Mohamed	Karim	Yanis	Ahmed	Rachid
			♀	Fatima	Malika	Khadija	Samira	Leila
	Africain	248	♂	Mamadou	Abdoulaye	Ibrahima	Amadou	Gloire
			♀	Aminata	Fatoumata	Awa	Aya	Maina
	Asiatique	209	♂	s	Loan	s	s	s
			♀	Louane	s	s	s	s
	Inclassable	989	♂	s	s	s	s	s
			♀	s	s	s	s	s
	Total	42 208	♂	Sébastien	Patrick	Michel	Nicolas	Stéphane
			♀	Catherine	Nathalie	Martine	Isabelle	Françoise

Lecture : les prénoms « français » (d'après la nomenclature de l'INED) qui sont les plus portés sont « Sébastien » et « Catherine ».

Champ : personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008, données pondérées, N = 42 208.

Source : TeO (voir Annexe 5B, p. 183).

NB : Conformément aux règles du secret statistique du [CASD](#) et de l'[INSEE](#), et pour qu'aucune information ne permette d'identifier directement ou indirectement une personne physique (ni d'inférer de nouvelles informations sur elle), les résultats obtenus sur des effectifs faibles (<5) sont notés « s » (comme « secret »).

Tableau 6.15a. Origine du prénom (colonnes) selon l'origine religieuse de son porteur (lignes), en France en 2008 (enquête TeO)

		N	ORIGINE DU PRÉNOM DES INDIVIDUS (NOMENCLATURE SYNTHÉTIQUE)									
			Français	Latin	International ou anglophone	Autre Europe	Arabo- musulman	Africain	Asiatique	Juif	Inclassable	Total
ORIGINE RELIGIEUSE DES INDIVIDUS	Chrétienne	27 498	82 %	5 %	5 %	2 %	1 %	0 %	0 %	0 %	4 %	100 %
	Musulmane	3 675	4 %	3 %	1 %	0 %	73 %	4 %	0 %	1 %	13 %	100 %
	Juive	273	59 %	5 %	12 %	6 %	2 %	0 %	2 %	6 %	9 %	100 %
	Bouddhiste, hindouiste, etc.	489	34 %	6 %	9 %	1 %	5 %	0 %	8 %	1 %	36 %	100 %
	Sans origine religieuse	9 594	75 %	5 %	8 %	2 %	3 %	0 %	0 %	1 %	6 %	100 %
	Non réponse	680	74 %	4 %	7 %	2 %	8 %	0 %	0 %	1 %	4 %	100 %
	Total	42 208	73 %	5 %	6 %	2 %	8 %	0 %	0 %	1 %	6 %	100 %

Lecture : 82 % des individus d'origine chrétienne portent un prénom « français » (d'après la nomenclature synthétique).

Champ : personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008, données pondérées, N = 42 208.

Source : TeO (voir Annexe 5B, p. 183).

**Tableau 6.15b. Origine religieuse de l'individu (lignes) selon l'origine de son prénom (colonnes), en France en 2008 (enquête TeO)**

		ORIGINE DU PRÉNOM DES INDIVIDUS (NOMENCLATURE SYNTHÉTIQUE)									
		Français	Latin	International ou anglophone	Autre Europe	Arabo- musulman	Africain	Asiatique	Juif	Inclassable	Total
N		30 821	2 143	2 360	659	3 220	160	158	294	2 392	42 208
ORIGINE RELIGIEUSE DES INDIVIDUS	Chrétienne	73 %	70 %	61 %	66 %	6 %	8 %	39 %	45 %	46 %	65 %
	Musulmane	0 %	5 %	1 %	3 %	84 %	86 %	5 %	18 %	20 %	9 %
	Juive	1 %	1 %	1 %	3 %	0 %	0 %	3 %	6 %	1 %	1 %
	Bouddhiste, hindouiste, etc.	1 %	1 %	2 %	1 %	1 %	0 %	24 %	1 %	7 %	1 %
	Sans origine religieuse	23 %	22 %	33 %	27 %	8 %	5 %	29 %	27 %	24 %	23 %
	Non réponse	2 %	1 %	2 %	2 %	2 %	0 %	0 %	3 %	1 %	2 %
	Total	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %

Lecture : 75 % des porteurs d'un prénom « français » (d'après la nomenclature synthétique) sont d'origine chrétienne.

Champ : personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008, données pondérées, N = 42 208.

Source : TeO (voir Annexe 5B, p. 183).

- **Les prénoms du groupe « majoritaire » et les prénoms « français » en métropole en 2008**

**Les prénoms du groupe « majoritaire ».** En France métropolitaine en 2008, la plupart (75 %) des individus du champ retenu font partie du groupe « majoritaire », c'est-à-dire qu'ils ne sont ni immigrés ni nés dans un DOM et n'ont aucun parent ou grand-parent immigré ni né dans un DOM (tableau 6.13b). Parmi eux, la plupart portent un prénom « français » (84 %) et une minorité portent un prénom « international ou anglophone » (11 %) (tableau 6.13a). Les prénoms les plus portés dans ce groupe « majoritaire » sont, par ordre décroissant : « Sébastien », « Patrick », « Michel », « Éric » et « Philippe » pour les hommes, et « Catherine », « Martine », « Nathalie », « Isabelle » et « Françoise » pour les femmes (tableau 6.14a). Si presque tous ces prénoms sont « français », notons toutefois que l'un des prénoms les plus fréquents chez les hommes du groupe « majoritaire », « Patrick », n'est pas catégorisé par la nomenclature de l'INED comme « français » mais comme « international ou anglophone ». De fait, le prénom « Patrick » est la forme *anglaise* du prénom latin « Patricius », qui signifie « aristocrate » ; saint Patrick est l'évêque du 5<sup>e</sup> siècle qui a évangélisé l'Irlande, et le saint patron de l'île.

**Les prénoms « français ».** En France métropolitaine en 2008, 73 % des individus du champ retenu portent un prénom que la nomenclature de l'INED catégorise comme « français » (tableau 6.7), c'est-à-dire un prénom usuellement attribué par les locuteurs de langue française. La plupart (87 %) des porteurs d'un prénom « français » sont du groupe « majoritaire », c'est-à-dire qu'ils ne sont ni immigrés ni nés dans un DOM et n'ont aucun parent ou grand-parent immigré ni né dans un DOM (tableau 6.13b). La plupart des porteurs d'un prénom « français » sont aussi d'origine chrétienne (73 %) ou sans origine religieuse (23 %) (tableau 6.15b).

On a vu que depuis les environs du 12<sup>e</sup> siècle les prénoms « français » peuvent être issus du Nouveau Testament, du prénom d'un saint ou martyr chrétien ou d'un prénom profane, leur origine linguistique pouvant être très variée (Annexe 6A, p. 267 et tableau 6.3). En l'occurrence en métropole en 2008, les prénoms « français » les plus portés comprennent (tableau 6.14b) :

- ✓ des prénoms du Nouveau Testament, qu'ils soient d'origine hébraïque comme « Michel »<sup>131</sup> ou grecque comme « Stéphane »<sup>132</sup> et « Philippe »<sup>133</sup> ;
- ✓ des prénoms de saints et martyrs chrétiens, qu'ils soient d'origine grecque comme « Nicolas »<sup>134</sup> et « Catherine »,<sup>135</sup> ou d'origine latine comme « Sébastien »,<sup>136</sup> « Nathalie »<sup>137</sup> et « Martine »<sup>138</sup> ;

<sup>131</sup> « Michel » est la forme française du prénom hébraïque « Mikha'el », qui signifie « qui est comme Dieu ? » ; Michel est l'archange de la Bible hébraïque et du Nouveau Testament.

<sup>132</sup> « Stéphane » est la forme française du prénom grec « Stephanos », qui signifie « couronne » ; Stéphane est le premier martyr chrétien, d'après le Nouveau Testament.

<sup>133</sup> « Philippe » est la forme française du prénom grec « Philippos », qui signifie « ami des chevaux » ; Philippe est l'un des douze apôtres du Nouveau Testament.

<sup>134</sup> « Nicolas » est la forme française du prénom grec « Nikolaos », qui signifie « victoire du peuple » ; saint Nicolas de Myre était un évêque d'Asie Mineure du 4<sup>e</sup> siècle. Pour plus de précisions sur ce prénom et son histoire en Europe et en France, voir l'Annexe 7, p. 327.

<sup>135</sup> « Catherine » est la forme française du prénom grec « Aikaterine », de signification ambiguë ; sainte Catherine était une martyre d'Égypte du 4<sup>e</sup> siècle.

<sup>136</sup> « Sébastien » est la forme française du prénom latin « Sebastianus », qui signifie « de Sébaste », une ville d'Asie Mineure ; saint Sébastien était un soldat romain et martyr du 3<sup>e</sup> siècle.

<sup>137</sup> « Nathalie » est la forme française du prénom latin « Natalia », qui signifie « naissance (du seigneur) » ; ce prénom a été porté par l'épouse d'un saint et martyr d'Asie Mineure du 4<sup>e</sup> siècle.

<sup>138</sup> « Martine » est la forme française féminisée du prénom latin « Martinus », qui signifie « de Mars » ; saint Martin était un évêque du 4<sup>e</sup> siècle, devenu le saint patron de la France.



- ✓ des prénoms profanes d'origine latine, comme « Isabelle »<sup>139</sup> et « Françoise ».<sup>140</sup>

**Les prénoms des métropolitains originaires des DOM.** En métropole en 2008, 1 % des individus du champ retenu sont originaires des DOM (Guadeloupe, Martinique, Guyane et Réunion) (tableau 6.13b). Parmi eux, la plupart (68 %) portent un prénom « français », même si une minorité non négligeable (19 %) portent un prénom « international ou anglophone » (tableau 6.13a). Leurs prénoms les plus fréquents sont « David », « Patrick », « Olivier », « Thierry » et « Alain » pour les hommes, et « Nathalie », « Stéphanie », « Véronique », « Aurélie » et « Karine » pour les femmes (tableau 6.14a). Si l'un de ces prénoms est catégorisé comme « international ou anglophone » (« Patrick »<sup>141</sup>), la plupart sont « français », qu'ils soient originaires de l'Ancien Testament comme « David »,<sup>142</sup> issus de saints et martyrs chrétiens comme « Nathalie »,<sup>143</sup> « Véronique »<sup>144</sup> et « Karine »,<sup>145</sup> ou encore profanes — que ces prénoms profanes soient d'origine grecque comme « Stéphanie »,<sup>146</sup> latine comme « Aurélie »,<sup>147</sup> germanique comme « Olivier »<sup>148</sup> et « Thierry »<sup>149</sup> ou encore d'origines diverses comme « Alain ».<sup>150</sup>

- **Les prénoms des habitants originaires d'Europe du Sud et les prénoms « latins » en métropole en 2008**

**Les prénoms des personnes originaires d'Europe du Sud.** En France métropolitaine en 2008, 6 % des individus du champ retenu sont originaires d'Europe du Sud (Italie, Espagne et Portugal) (tableau 6.13b), et parmi eux la plupart portent un prénom « français » (58 %), plutôt que « latin » (20 %) ou « international ou anglophone » (17 %) (tableau 6.13a). Les prénoms les plus portés par les personnes originaires d'Europe du Sud sont « José », « Daniel », « Manuel », « David » et « Antonio » pour les hommes, et « Maria », « Isabelle », « Marie-Carmen », « Céline » et « Élisabeth » pour les femmes. Si certains de ces prénoms sont catégorisés comme « latins » comme « José »,<sup>151</sup> « Manuel »<sup>152</sup> et « Antonio »,<sup>153</sup> et si d'autres

<sup>139</sup> « Isabelle » est la forme française du prénom occitan médiéval qui est la forme occitane du prénom « Élisabeth ».

<sup>140</sup> « Françoise » est la forme française féminisée du terme germano-latin qui signifie « français ».

<sup>141</sup> « Patrick » est la forme *anglaise* du prénom latin « Patricius », qui signifie « aristocrate » ; saint Patrick est l'évêque du 5<sup>e</sup> siècle qui a évangélisé l'Irlande, et le saint patron de l'île.

<sup>142</sup> « David » est la forme (notamment) française du prénom hébraïque « Dawid », qui signifie « oncle » ou « bien-aimé », du nom d'un roi d'Israël du 10<sup>e</sup> siècle avant l'ère commune, dans l'Ancien Testament.

<sup>143</sup> « Nathalie » est la forme française du prénom latin « Natalia », qui signifie « naissance (du seigneur) » ; ce prénom a été porté par l'épouse d'un saint et martyr d'Asie Mineure du 4<sup>e</sup> siècle.

<sup>144</sup> « Véronique » est la forme française du prénom latin « Veronica », qui est la forme latinisée du prénom grec « Bérénice » ; sainte Véronique est une sainte qui, lors de la montée du Golgotha, aurait recueilli l'empreinte du visage de Jésus sur un linge.

<sup>145</sup> « Karine » est la forme française du prénom latin « Carina », qui signifie « bien-aimée » ; sainte Carine était une sainte d'Asie Mineure du 4<sup>e</sup> siècle.

<sup>146</sup> « Stéphanie » est la forme française féminine du prénom grec « Stephanos », qui signifie « couronne ».

<sup>147</sup> « Aurélie » est la forme française féminisée du nom de famille latin « Aurelius », qui signifie « doré », du nom d'un philosophe et empereur romain du 2<sup>e</sup> siècle.

<sup>148</sup> « Olivier » est la forme française d'un prénom germanique médiéval.

<sup>149</sup> « Thierry » est la forme française du prénom germanique « Théodoric », qui signifie « peuple puissant »

<sup>150</sup> « Alain » est la forme française d'un prénom d'origines diverses, notamment latine et celtique.

<sup>151</sup> « José » est la forme espagnole et portugaise du prénom d'origine hébraïque « Joseph ». À Mexico (Mexique), le prénom masculin le plus attribué est « Joseph » ou « Josep » ou « José » du 18<sup>e</sup> siècle jusqu'au moins la seconde moitié du 20<sup>e</sup> siècle (Boyd-Bowman 1970). Mais si le prénom « José » a été beaucoup attribué dans le monde hispanophone, de nos jours ce n'est plus le cas.

<sup>152</sup> « Manuel » est la forme diminutive italienne, espagnole et portugaise (mais aussi française) du prénom hébraïque « Emmanuel », qui signifie « Dieu avec nous ».

<sup>153</sup> « Antonio » est la forme italienne, espagnole et portugaise du nom de famille latin (d'origine étrusque) « Antonius », de signification inconnue, du nom d'un général qui a gouverné l'empire romain au 1<sup>er</sup> siècle avant

comme « Maria » ou « María »<sup>154</sup> sont — de façon contestable — catégorisés comme « internationaux ou anglophones », la plupart de ces prénoms sont effectivement « français » : c'est le cas de « Daniel »,<sup>155</sup> « David »,<sup>156</sup> « Isabelle »,<sup>157</sup> « Marie-Carmen »,<sup>158</sup> « Céline »<sup>159</sup> et « Élisabeth ».<sup>160</sup>

Le fait que les personnes originaires d'Europe du Sud portent majoritairement des prénoms « français » correspond en fait à deux phénomènes distincts. D'une part, un certain nombre d'immigrés d'Europe du Sud ont francisé leur prénom lorsqu'ils ont acquis la nationalité française.<sup>161</sup> En France en 1992, près d'un quart des immigrés d'Espagne et du Portugal devenus français avaient, lors de leur naturalisation, francisé leur prénom.<sup>162</sup> Notons qu'aux États-Unis au début du 20<sup>e</sup> siècle, les immigrés d'Europe du Sud (et d'Europe de l'Est) avaient encore plus massivement américanisé leurs prénoms.<sup>163</sup> D'autre part, et c'est le plus important, la grande majorité des immigrés d'Europe du Sud en métropole ont attribué à leurs enfants (et ces derniers, à leurs propres enfants) des prénoms « français ». Par exemple, les enfants d'immigrés du Portugal ont massivement reçu des prénoms « français » dès les années 1970 (Fourquet et Manternach 2019a, p. 179-180 ; pour plus de précisions, voir tableau 7.3).<sup>164</sup>

**Les prénoms « latins ».** En France métropolitaine en 2008, seuls 3 % des individus du champ retenu portent un prénom que la nomenclature de l'INED catégorise comme « latin » (tableau 6.7), c'est-à-dire un prénom couramment attribué par les locuteurs italophones,

---

l'ère commune.

<sup>154</sup> « Maria » (forme italienne et portugaise) ou « María » (forme espagnole) est la variante latinisée du prénom hébraïque de la sœur de Moïse dans la Bible hébraïque, et de la mère de Jésus dans le Nouveau Testament. Toutefois, la nomenclature de l'INED catégorise ce prénom latin comme « international ou anglophone ».

<sup>155</sup> « Daniel » est la forme française du prénom hébraïque qui signifie « Dieu est mon juge ».

<sup>156</sup> « David » est la forme à la fois française, espagnole et portugaise du prénom hébraïque « Dawid », qui signifie « oncle » ou « bien-aimé », du nom d'un roi d'Israël du 10<sup>e</sup> siècle avant l'ère commune, dans l'Ancien Testament. La nomenclature de l'INED catégorise ce prénom comme « français ».

<sup>157</sup> « Isabelle » est la forme française du prénom occitan médiéval qui est la forme occitane du prénom « Élisabeth ». À distinguer de « Isabel », qui est la forme espagnole et portugaise du même prénom occitan médiéval.

<sup>158</sup> Si « Marie-Carmen » est catégorisé comme français alors que « Carmen » est un prénom espagnol, c'est parce que le premier composant de ce prénom composé (« Marie ») est catégorisé comme « français ».

<sup>159</sup> « Céline » est la forme française d'un nom de famille latin qui signifie « céleste » ; sainte Céline était une sainte du 5<sup>e</sup> siècle, mère de saint Remi.

<sup>160</sup> « Élisabeth » est la forme française du prénom hébraïque qui signifie « Dieu est un serment ».

<sup>161</sup> Les immigrés d'Europe du Sud qui habitaient une colonie française ont pu voir leur prénom francisé à l'occasion d'une naturalisation (ou d'un mariage ou d'une naissance) dès avant leur migration en métropole. L'individu naturalisé (ou réintégré dans la nationalité française) peut franciser son prénom, soit en remplaçant son prénom étranger par un prénom français, soit en ajoutant un prénom français à son prénom étranger, soit en supprimant son prénom étranger, soit encore en plaçant en première position son deuxième (ou troisième) prénom français ; pour plus de précisions, voir la [notice explicative concernant la demande de francisation ou d'identification](#).

<sup>162</sup> Au sein de la population en France en 1992, 20 % des hommes et 25 % des femmes immigrés du Portugal ont francisé leur prénom, suite à l'acquisition de la nationalité française (Tribalat et al. 1996, p. 162). « Avec des proportions voisines, l'écart est inversé entre les deux sexes parmi les migrants espagnols, en raison de la plus grande part de femmes devenues françaises par mariage moins souvent suivi d'une francisation du prénom » (Tribalat et al. 1996, p. 162). Chez les immigrés espagnols en France en 1992, « Les francisations du nom mais surtout du prénom ont été relativement fréquentes (un quart environ de ceux qui sont devenus français) » (Tribalat et al. 1996, p. 183).

<sup>163</sup> Aux États-Unis de 1900 à 1930, 77 % des immigrés ont changé de prénom (américanisé leur prénom ou adopté un prénom américain) suivant la migration, surtout ceux d'entre eux originaires d'Europe plutôt que de Chine ou du Japon, et surtout ceux vivant dans des comtés où les immigrés de même origine sont peu concentrés (Carneiro et al. 2020). 'When continental Europeans immigrated to America, many adopted English forms of their names. For example, Italian Antonios became Anthonys. If an English equivalent was not well-known, names shifted to something with a similar sound. Many Polish Stanislaws became Stanleys' (Lawson 2016, p. 188).

<sup>164</sup> Aux États-Unis au début du 20<sup>e</sup> siècle, les immigrés — principalement d'Europe du Sud — avaient là aussi massivement attribué à leurs enfants des prénoms américains : de 1900 à 1930, 86 % des fils d'immigrés et 93 % des filles d'immigrés se voient attribuer un prénom américain (Carneiro et al. 2020).

hispanophones ou lusophones. Les porteurs de prénoms « latins » peuvent être originaires d'Europe du Sud (44 %) mais aussi du groupe « majoritaire » (38 %) (tableau 6.13b) et ils sont, plus largement, d'origine chrétienne (70 %) ou sans origine religieuse (22 %) (tableau 6.15b).

On a vu que depuis les environs du 12<sup>e</sup> siècle les prénoms « latins » en Europe peuvent être issus du prénom d'un personnage du Nouveau Testament, du prénom d'un saint ou martyr chrétien ou d'un prénom profane, leur étymologie pouvant être très variée (Annexe 6A, p. 267 et tableau 6.3). En l'occurrence en métropole en 2008, les prénoms « latins » les plus portés comprennent (tableau 6.14b) :

- ✓ des prénoms du Nouveau Testament comme « José », <sup>165</sup> mais aussi des prénoms dérivés de prénoms de l'Ancien Testament comme « Manuel » <sup>166</sup> ;
- ✓ des prénoms de saints et martyrs chrétiens, notamment d'origine grecque comme « Inès » ou encore « Inés » ou Inês » <sup>167</sup> ;
- ✓ des prénoms profanes, qu'ils soient d'origine grecque comme « Teo », <sup>168</sup> latine comme « Antonio », <sup>169</sup> « Sabrina », <sup>170</sup> « Lola » <sup>171</sup> et « Luna », <sup>172</sup> germanique comme « Enzo » <sup>173</sup> ou encore d'origines diverses comme « Amelia ». <sup>174</sup>

• **Les prénoms des habitants originaires d'autre Europe, et les prénoms « internationaux ou anglophones » et « d'autre Europe » en métropole en 2008**

**Les prénoms des personnes originaires d'autre Europe.** En France métropolitaine en 2008, 4 % des individus du champ retenu sont originaires d'autre Europe (c'est-à-dire d'Europe hors France, Italie, Espagne et Portugal) (tableau 6.13b), et parmi eux la plupart portent un prénom « français » (59 %) plutôt que « international ou anglophone » (20 %) ou d'« autre Europe » (13 %) (tableau 6.13a). Les prénoms les plus portés par les personnes originaires d'autre Europe sont « Jacques », « Jean-Claude », « Michel », « Pierre » et « Nicolas » pour les hommes, et « Catherine », « Anne », « s », <sup>175</sup> « Nadège » et « Sandra » pour les femmes (tableau 6.14a). De fait, les enfants d'immigrés de Pologne ont massivement reçu des prénoms « français » dès les années 1930 (Fourquet et Manternach 2019a, p. 178).

<sup>165</sup> « José » est la forme espagnole et portugaise du prénom d'origine hébraïque « Joseph ».

<sup>166</sup> « Manuel » est la forme diminutive italienne, espagnole et portugaise (mais aussi française) du prénom hébraïque « Emmanuel », qui signifie « Dieu avec nous ».

<sup>167</sup> « Inès » est la forme française du prénom espagnol « Inés », qui est la forme espagnole d'un prénom grec qui signifie « chaste » ; sainte Agnès (« Inés » en espagnol, « Inês » en portugais) était une martyre des 3<sup>e</sup> et 4<sup>e</sup> siècles.

<sup>168</sup> « Teo » est la forme diminutive italienne et espagnole du prénom d'origine grecque « Teodoro », qui signifie « don de Dieu ».

<sup>169</sup> « Antonio » est la forme italienne, espagnole et portugaise du nom de famille latin (d'origine étrusque) « Antonius », de signification inconnue, du nom d'un général qui a gouverné l'empire romain au 1<sup>er</sup> siècle avant l'ère commune.

<sup>170</sup> « Sabrina » est la forme italienne (mais aussi française) d'un prénom latin d'origine galloise.

<sup>171</sup> « Lola » est le diminutif du prénom espagnol « Dolores », qui signifie « douleurs » et qui est l'une des invocations mariales.

<sup>172</sup> « Luna » est la forme italienne et espagnole du prénom latin qui signifie « lune », du nom de la déesse romaine de la lune.

<sup>173</sup> « Enzo » est la forme italienne d'un prénom germanique de signification incertaine. Si le prénom « Enzo », qui ne figure pas dans le Top 5 des personnes originaires d'Europe du Sud, figure malgré tout en première place du Top 5 des prénoms « latins », c'est parce qu'il est relativement souvent porté, aussi, par les descendants du groupe « majoritaire » : c'est même le prénom masculin le plus attribué en France dans les années 2000, après « Lucas » (tableau 5.5).

<sup>174</sup> « Amelia » est un prénom italien et espagnol d'origines hébraïque, germanique et latine.

<sup>175</sup> Prénom « d'autre Europe » trop rare pour pouvoir être mentionné sans risque de réidentification.

**Les prénoms « internationaux ou anglophones ».** En France en 2008, 12 % des individus du champ retenu portent un prénom que la nomenclature de l'INED catégorise comme « international ou anglophone » (tableau 6.7), c'est-à-dire un prénom usuellement attribué par les locuteurs de la langue anglaise. La plupart des habitants porteurs de prénoms « internationaux ou anglophones » sont du groupe « majoritaire » (71 %) plutôt que d'autre Europe (6 %) (tableau 6.13b). La plupart des porteurs d'un prénom « international ou anglophone » sont aussi d'origine chrétienne (61 %) ou sans origine religieuse (33 %) (tableau 6.15b).

En Europe, depuis les environs du 12<sup>e</sup> siècle, les prénoms peuvent être issus du prénom d'un personnage du Nouveau Testament, du prénom d'un saint ou martyr chrétien ou d'un prénom profane, leur origine étymologique pouvant être très variée (Annexe 6A, p. 267 et tableau 6.3). En l'occurrence en métropole en 2008, les prénoms « internationaux ou anglophones » les plus portés comprennent (tableau 6.14b) :

- ✓ non pas des prénoms du Nouveau Testament mais des prénoms de l'Ancien Testament, comme « Jérémy » ou « Jeremy »<sup>176</sup> et « Sarah »<sup>177</sup> ;
- ✓ des prénoms de saints et martyrs chrétiens, qu'ils soient d'origine latine comme « Patrick »<sup>178</sup> et « Laura »,<sup>179</sup> d'origine germanique comme « Audrey »<sup>180</sup> ou d'origine celtique comme « Kévin » ou « Kevin »<sup>181</sup> ;
- ✓ des prénoms profanes, qu'ils soient d'origine latine comme « Anthony »<sup>182</sup> et « Patricia »,<sup>183</sup> ou inventés comme « Cédric » ou « Cedric »<sup>184</sup> ou encore « Vanessa ».<sup>185</sup>

La pénétration de ces prénoms « internationaux ou anglophones » dans la population « majoritaire » en France témoigne de l'internationalisation du stock de prénoms en France et, plus généralement, en Europe au 20<sup>e</sup> siècle (Leibring 2016, p. 210), au fur et à mesure que la télévision, les films et séries américains et Internet diffusent le lexique de prénoms nord-américains (Gerritzen 2006).

**Les prénoms d'« autre Europe ».** En France en 2008, seuls 1 % des individus du champ retenu portent un prénom que la nomenclature de l'INED catégorise comme d'« autre Europe » (tableau 6.7), c'est-à-dire un prénom usuellement attribué par les locuteurs d'une langue européenne hormis les langues française, italienne, espagnole, portugaise et anglaise. Les

<sup>176</sup> « Jeremy » est la forme anglaise du prénom hébraïque qui signifie « Yahvé exaltera », du nom de l'un des principaux prophètes de l'Ancien Testament.

<sup>177</sup> « Sarah » est la forme anglaise (mais aussi française) d'un prénom hébraïque qui signifie « princesse », du nom de l'épouse d'Abraham et mère d'Isaac, dans l'Ancien Testament.

<sup>178</sup> « Patrick » est la forme anglaise du prénom latin « Patricius », qui signifie « aristocrate » ; saint Patrick est l'évêque du 5<sup>e</sup> siècle qui a évangélisé l'Irlande, et le saint patron de l'île.

<sup>179</sup> « Laura » est la forme féminine (notamment) anglaise du prénom latin « Laurus », qui signifie « laurier », du nom d'une martyre espagnole du 9<sup>e</sup> siècle. Quoi qu'il en soit, la nomenclature de l'INED catégorise ce prénom comme « international ou anglophone ». De même « Daphné » vient du grec pour « laurier ».

<sup>180</sup> « Audrey » est la forme anglo-normande d'un prénom anglo-saxon qui signifie « force noble » ; sainte Audrey était une sainte anglaise du 7<sup>e</sup> siècle.

<sup>181</sup> « Kevin » est la forme anglaise d'un prénom gaélique qui signifie « belle naissance » ; saint Kevin était un abbé irlandais du 6<sup>e</sup> siècle. Ce prénom est absent de la littérature française du 17<sup>e</sup> au 20<sup>e</sup> siècles (base de données Frantext) (Brunet 1988, p. 127). En France, des analyses détaillées montrent que c'est là où les prénoms « Kevin » et « Dylan » sont les plus attribués dans les années 1990 (quart nord-est de la métropole) qu'on vote le plus Le Pen (plutôt que Macron) au second tour de la présidentielle de 2017 : c'est une France populaire affranchie des injonctions esthétiques et politiques des élites des grandes agglomérations (Fourquet et Manternach 2019, p. 126).

<sup>182</sup> « Anthony » est la forme anglaise du nom de famille latin (d'origine étrusque) « Antonius », de signification inconnue, du nom d'un général qui a gouverné l'empire romain au 1<sup>er</sup> siècle avant l'ère commune.

<sup>183</sup> « Patricia » est la forme féminine (notamment) anglaise du prénom latin « Patricius », qui signifie « aristocrate ».

<sup>184</sup> « Cedric » est un prénom anglais inventé en 1819 par le roman *Ivanhoé* de Walter Scott.

<sup>185</sup> « Vanessa » est un prénom anglais inventé en 1726 par le poème *Cadenus and Vanessa* de Jonathan Swift.

habitants porteurs d'un prénom d'« autre Europe » sont du groupe « majoritaire » (52 %) ou d'« autre Europe » (29 %) (tableau 6.13b). La plupart des porteurs d'un prénom d'« autre Europe » sont d'origine chrétienne (66 %) ou sans origine religieuse (27 %) (tableau 6.15b).

En métropole en 2008, les prénoms d'« autre Europe » les plus portés comprennent, outre deux prénoms trop rares pour pouvoir être mentionnés (tableau 6.14b) :

- ✓ des prénoms de l'Ancien et du Nouveau Testament, notamment germaniques comme « Axel »<sup>186</sup> et slaves comme « Ivan »<sup>187</sup>;
- ✓ des prénoms de saints et martyrs chrétiens, notamment germaniques comme « Kilian »<sup>188</sup> et « Karin »<sup>189</sup> et slaves comme « Tatiana »<sup>190</sup>;
- ✓ des prénoms profanes, qu'ils soient germaniques comme « Ingrid »<sup>191</sup>, slaves comme « Dimitri »<sup>192</sup> ou encore à la fois germaniques et slaves comme « Johanna ».<sup>193</sup>

#### • Les prénoms des habitants originaires du Maghreb et de Turquie et les prénoms « arabo-musulmans » en métropole en 2008

**Les prénoms des personnes originaires du Maghreb.** En France en 2008, 7 % des individus du champ retenu sont originaires du Maghreb (Algérie, Maroc et Tunisie) (tableau 6.13b), et parmi eux la plupart portent un prénom « arabo-musulman » (74 %) plutôt que « français » (10 %) (tableau 6.13a).<sup>194</sup> Les prénoms les plus portés par les personnes originaires du Maghreb sont « Mohamed », « Karim », « Ahmed », « Rachid » et « Farid » pour les hommes, « Fatima », « Sarah », « Nadia », « Myriam » et « Samira » pour les femmes (tableau 6.14a). Si certains de ces prénoms sont catégorisés comme « français » (« Myriam »<sup>195</sup>) ou « internationaux ou anglophones » (« Sarah »<sup>196</sup> et « Nadia »<sup>197</sup>), la plupart de ces prénoms sont

<sup>186</sup> « Axel » est la forme (notamment) danoise du prénom hébraïque « Absalom », qui signifie « mon père est paix » ; Axel de Lund était un archevêque danois du 12<sup>e</sup> siècle.

<sup>187</sup> « Ivan » est un cognat slave du prénom « Jean », lui-même d'origine hébraïque.

<sup>188</sup> « Kilian » est la forme allemande d'un prénom gaélique qui signifie « petite église », du nom d'un saint de Bavière (Allemagne) du 7<sup>e</sup> siècle.

<sup>189</sup> « Karin » est une forme germanique du prénom « Catherine » ; sainte Catherine était une martyre d'Égypte du 4<sup>e</sup> siècle.

<sup>190</sup> « Tatiana » est la forme féminine (notamment) russe du prénom latin « Tatianus », du nom d'une sainte de Rome au 3<sup>e</sup> siècle.

<sup>191</sup> « Ingrid » est la forme scandinave d'un prénom vieux-norrois qui signifie « la beauté de Ing ». Ce prénom est absent de la littérature française du 17<sup>e</sup> au 20<sup>e</sup> siècles (base de données Frantext) (Brunet 1988, p. 127).

<sup>192</sup> « Dimitri » est la forme russe du prénom grec « Demetrios », qui signifie « qui appartient à Déméter ».

<sup>193</sup> « Johanna » est, dans plusieurs langues d'Europe septentrionale, centrale et orientale, la forme du prénom grec « Ioanna », forme féminine du prénom hébraïque qui signifie « Dieu est bon ».

<sup>194</sup> Les données sous-jacentes aux tableaux 6.13a et 6.13b contiennent une erreur qui, une fois corrigée, fait passer la part d'immigrés et de descendants d'immigrés du Maghreb qui portent un prénom « arabo-musulman » de 74 % à 80 %, et la part de ceux qui portent un prénom « français » de 10 % à 5 % Chapitre 8, p. 46).

<sup>195</sup> « Myriam », prénom d'origine hébraïque de l'Ancien Testament, du nom de la sœur de Moïse, est catégorisé comme « français » par la nomenclature de l'INED mais comme « juif » par la nomenclature synthétique. Le succès de ce prénom chez les personnes originaires du Maghreb semble tenir au fait que, si c'est un prénom coranique (qui désigne la mère de Jésus, le prophète 'Isa, à laquelle est consacrée la sourate Maryam), il n'est pas exclusivement arabe ou musulman, a fortiori sous la forme « Myriam » plutôt que sous les formes « Maryam », « Mariam » ou « Meriem ».

<sup>196</sup> « Sarah » est un prénom d'origine hébraïque qui signifie « princesse », du nom de l'épouse d'Abraham et mère d'Isaac, dans l'Ancien Testament. Sarah n'est pas nommée dans le Coran (il est seulement fait allusion à la femme d'Ibrahim), mais elle est bien connue de la tradition islamique : le nom est prononcé Sāra(h) en arabe, le -h (qui est écrit) ayant pu être éventuellement prononcé à date ancienne. Le succès de ce prénom chez les personnes originaires du Maghreb semble tenir au fait que, comme « Myriam », il n'est pas exclusivement arabe ou musulman. Ce prénom est catégorisé comme « international ou anglophone » par la nomenclature de l'INED, mais comme « juif » par la nomenclature synthétique.

<sup>197</sup> « Nadia », prénom européen d'origine slave qui signifie « espoir », est aussi l'orthographe courante du prénom arabe « Nādiya » (qui est le participe actif, au féminin singulier, du verbe *nadā* [« appeler »]), qui signifie « qui

effectivement « arabo-musulmans » : c'est le cas de tous les prénoms masculins — « Mohamed », <sup>198</sup> « Karim », <sup>199</sup> « Ahmed », <sup>200</sup> « Rachid » <sup>201</sup> et « Farid » <sup>202</sup> — et de « Fatima » <sup>203</sup> et « Samira ». <sup>204</sup> Si ces prénoms sont courants au Maghreb, « Sarah » et « Nadia » y sont portés par des générations plus récentes que « Fatima » et « Samira », tandis que « Mohamed », « Ahmed » et « Karim » semblent se maintenir au fil des générations. Au Maghreb, ces prénoms ne signalent aucune position sociale ni aucun degré de religiosité particulier, même si « Mohamed » et « Ahmed » sont deux des noms du prophète de l'islam.

Le fait que les personnes originaires du Maghreb ne portent que rarement des prénoms « français » s'explique bien. D'une part, presque aucun immigré du Maghreb ne s'est vu attribuer à la naissance un prénom « français » (Annexe 6A, p. 267), et rares sont ceux qui ont francisé leur prénom lors de l'acquisition de la nationalité française. En France en 1992,

« les francisations [du prénom] sont très rares parmi les immigrés anciennement algériens, marocains ou africains. Les immigrés d'Algérie gardent des réticences à acquérir la nationalité et, lorsque le pas est franchi, ils ne vont pas jusqu'à renier leur propre origine en transformant ne serait-ce que le prénom. Cela vaut pour les Arabes et les Berbères : parmi les premiers dont l'effectif est important, la proportion de francisations n'est que de 4 %. On trouve une proportion voisine chez les Arabes du Maroc. Très peu de Berbères sont devenus français et aucun n'a francisé son prénom » (Tribalat et al. 1996, p. 162).

D'autre part, rares sont les immigrés du Maghreb qui attribuent à leurs enfants (et ces derniers, à leurs propres enfants) des prénoms « français ». Ainsi en France en 2003, les

---

appelle ». L'orthographe française reflète la prononciation dialectale qui est Nādyā (la présence d'une voyelle longue, portant l'accent, dans la première syllabe, entraîne la syncope de la voyelle brève *i* de la seconde syllabe, le *i* du français représentant le *y* de l'arabe. Le succès de ce prénom chez les personnes originaires du Maghreb semble relever du jeu de mots ou de l'homonymie. C'est ainsi que le prénom « Nadia » peut être catégorisé comme « international ou anglophone » par la nomenclature de l'INED mais comme « arabo-musulman » par la nomenclature synthétique.

<sup>198</sup> « Mohamed » est le prénom arabe du prophète de l'islam, qui signifie « le très loué », *muḥammad* étant en arabe le participe passif du verbe *ḥammada*, qui est l'itératif (« louer et louer ») du verbe *ḥamida* (« louer »). Au Maghreb ce prénom est souvent attribué à l'aîné pour remercier la providence d'avoir béni la famille avec la naissance d'un garçon. « Le fait que ce prénom se maintienne aussi fortement en France en dépit du risque de visibilité ethnique qu'il fait courir à son porteur témoigne de l'intensité du sentiment d'appartenance nationale, culturelle et religieuse de ce groupe » (Streiff-Fénart 1990, p. 6). En France le prénom « Mohamed », qui est entré dans le Top 100 des prénoms masculins les plus attribués dans les années 1960, figure dans le Top 20 depuis 2018 ([Fichier INSEE des prénoms](#)), peut-être en lien avec l'essor de la religiosité des musulmans en âge de devenir parents, y compris dans sa variante salafiste. Sur la fréquence à laquelle « Mohamed » est encore attribué en France, voir aussi Puzenat 2008, p. 122. En Angleterre-Galles le prénom « Muhammad », qui était entré dans le Top 100 des prénoms les plus attribués dès les années 1920, figure dans le Top 10 depuis 2016 (Stripe 2019), pour plusieurs raisons emboîtées : non seulement la proportion de musulmans a fortement augmenté et cette population continue d'attribuer à ses descendants des prénoms spécifiquement arabo-musulmans, mais en outre les prénoms que les musulmans attribuent à leurs garçons en Angleterre-Galles sont particulièrement concentrés sur « Muhammad » ; reste à savoir si, en outre, cette concentration sur le prénom « Muhammad » augmente elle aussi au fil des ans.

<sup>199</sup> « Karim » est un prénom arabe profane qui signifie « généreux ».

<sup>200</sup> « Ahmed » est un autre nom (arabe) de Mohamed (il appartient à la même famille lexicale), qui signifie « plus loué [que tout autre] » ou « particulièrement loué ».

<sup>201</sup> « Rachid » est un prénom arabe qui signifie « bien guidé » et qui renvoie aux quatre premiers califes, ce qui en fait un prénom typiquement sunnite.

<sup>202</sup> « Farid » est un prénom arabe profane qui signifie « unique ».

<sup>203</sup> « Fatima » est un prénom arabe issu (du participe actif féminin) d'un verbe qui signifie « sevrer (un petit) ». C'est un prénom préislamique, typique d'une société d'éleveurs, qui est devenu religieux, parce que porté par la fille préférée du prophète Mohamed et épouse du 4<sup>e</sup> calife (et mère de Hasan et Husayn). Notons que le prénom catholique portugais « Maria de Fátima », qui doit sa popularité aux apparitions mariales qui ont eu lieu dans le village du centre du Portugal appelé Fátima (en 1917), est bien catégorisé comme « latin », même si le nom de ce village provient lui-même de l'arabe.

<sup>204</sup> « Samira » est un prénom arabe profane qui signifie « compagnon de discussion du soir ».



immigrés du Maghreb attribuent des prénoms sensiblement plus éloignés des prénoms de la population « majoritaire » que ce n'est le cas des immigrants d'Europe (Arai et al. 2015, p. 145-146). De même, « L'analyse onomastique effectuée à partir des listes électorales de plusieurs villes fait ainsi apparaître que près de 95 % des personnes ayant un patronyme les rattachant aux mondes arabo-musulmans portent également un prénom renvoyant à cette aire culturelle » (Fourquet et Manternach 2019a, p. 175). La plupart des descendants d'immigrés du Maghreb en France reçoivent donc des prénoms spécifiques, ce qui n'est plus le cas des juifs depuis le 19<sup>e</sup> siècle (voir Grange 2016b, p. 78-79), ni des descendants d'immigrés polonais du Nord-Pas-de-Calais depuis les années 1930 (Fourquet et Manternach 2019a, p. 177-179), ni des descendants d'immigrés portugais depuis les années 1970 (Fourquet et Manternach 2019a, p. 179-180), ni des descendants d'immigrés asiatiques (voir ci-dessous).

**Les prénoms des personnes originaires de Turquie.**<sup>205</sup> En France en 2008, 1 % des individus du champ retenu sont originaires de Turquie (tableau 6.13b), et parmi eux la plupart portent des prénoms catégorisés comme « arabo-musulmans » (78 %), y compris turcs, plutôt que « français » (11 %) (tableau 6.13a). Autrement dit, les prénoms des personnes originaires de Turquie sont aussi fréquemment « arabo-musulmans » et aussi rarement « français » que ceux des personnes originaires du Maghreb. Les prénoms les plus portés par les personnes originaires de Turquie sont : « Mehmet », « Ali », « Mustafa », « Ibrahim » et « Murat » pour les hommes, et « Murielle », « Fatma », « Ayse », « Hatice », et « Emine » pour les femmes (tableau 6.14a). Si l'un de ces prénoms est « français » (« Murielle »<sup>206</sup>), tous les autres sont effectivement « arabo-musulmans » et, plus précisément, d'origine arabe : « Mehmet »,<sup>207</sup> « Ali »,<sup>208</sup> « Mustafa »,<sup>209</sup> « Ibrahim »,<sup>210</sup> « Murat »,<sup>211</sup> « Fatma »,<sup>212</sup> « Ayse »,<sup>213</sup> « Hatice »<sup>214</sup>

<sup>205</sup> L'analyse des prénoms des personnes originaires de Turquie est rendue difficile du fait de l'usage de plusieurs prénoms en Turquie — le premier prénom, celui d'un grand-parent, n'étant souvent pas utilisé dans la pratique — et d'un usage multiple accentué en migration (prénoms composés ou prénom d'usage fluctuant selon les interlocuteurs français et turcs).

<sup>206</sup> « Murielle » est la forme française d'un prénom médiéval breton qui signifie « mer radieuse ». Je reviens ci-dessous sur la place de « Murielle » parmi les prénoms des personnes originaires de Turquie.

<sup>207</sup> « Mehmet » est la forme turque du prénom arabe « Mohamed », classique en Turquie non seulement de 1950 à 2010 (Sabuncu et Coulmont 2011) mais dès les périodes kémaliste (1923-1938) et, surtout, ottomane (1299-1922). Étant donné la longue histoire du prénom « Mehmet » chez les Turcs, le fait qu'il soit d'origine arabe n'est plus saillant et le fait de l'attribuer ne semble pas envoyer un signal particulier de religiosité. En revanche le prénom « Muhammed », qui n'est attribué par les Turcs que depuis bien plus récemment, signale bien la religiosité : « le succès récent de Muhammed [...] semble remplacer Mehmet. Si l'on fait l'hypothèse que Mehmet, forme turquisée de "Mohamed", pouvait être lié au nationalisme des parents (préférant des prénoms "turcs" pour leurs enfants), alors on peut supposer que Muhammed est donné par des parents plus islamistes que nationalistes (ou trouvant désuet le recours à une forme éloignée de l'arabe) » (Sabuncu et Coulmont 2011). En Turquie en 2019, « Muhammed » est le 7<sup>e</sup> prénom le plus fréquemment porté par les garçons de 0 à 17 ans, derrière « Mehmet » (3<sup>e</sup> prénom le plus porté par les garçons) (Turkstat 2019).

<sup>208</sup> « Ali » est un prénom arabe qui signifie « sublime », du nom d'un cousin et gendre du prophète et 4<sup>e</sup> calife, et père de toute la descendance du prophète par son mariage avec Fatima et leurs deux fils Hasan et Husayn. Ce prénom est classique en Turquie de 1950 à 2010 (Sabuncu et Coulmont 2011).

<sup>209</sup> « Mustafa » est la forme turque du prénom arabe « Mustapha », qui signifie « l'élus » (en référence au prophète Mohamed). C'est le prénom de plusieurs sultans ottomans.

<sup>210</sup> « Ibrahim » est la forme turque et arabe du prénom coranique d'origine hébraïque « Abraham ».

<sup>211</sup> « Murat » est la forme turque du prénom arabe profane « Mourad », qui signifie « vœu ».

<sup>212</sup> « Fatma » est la forme turque du prénom arabe « Fatima », du prénom de la fille préférée du prophète Mohamed et épouse du 4<sup>e</sup> calife. Ce prénom est classique en Turquie de 1950 à 2010 (Sabuncu et Coulmont 2011) et dès la période ottomane.

<sup>213</sup> « Ayse » est la forme turque du prénom arabe « Aïcha », qui signifie « vivante », du nom de la 3<sup>e</sup> épouse du prophète Mohamed. Ce prénom est classique en Turquie dès la période ottomane, mais il est peu porté par les aînés.

<sup>214</sup> « Hatice » est la forme turque du prénom arabe « Khadija », qui signifie « prématuré », du verbe khadaja, qui signifie « mettre bas un petit prématuré ». C'est un prénom préislamique, typique d'une société d'éleveurs, porté par la première épouse du prophète Mohamed et première personne convertie à l'islam.

et « Emine ».<sup>215</sup> En Turquie, ces prénoms ne signalent aucune appartenance ethnique (turque ou kurde) ni religieuse (sunnite ou alévie) particulière (encadré 6.2). Ces prénoms sont globalement conformes au fait que, parmi les prénoms (féminins) attribués en Turquie de 1950 à 2010, environ 50 % sont d'origine arabe et seulement environ 25 % d'origine turque, les autres étant le plus souvent d'origine indéterminée ou persane (Sabuncu et Coulmont 2011).<sup>216</sup> Le fait que les personnes originaires de Turquie ne portent que rarement des prénoms « français » s'explique par les raisons évoquées ci-dessus à propos des personnes originaires du Maghreb. En effet le stock de prénoms turcs, comme le stock de prénoms arabes, est presque entièrement disjoint du stock de prénoms européens, qu'ils soient français ou encore allemands (Gerhards et Hans 2009, p. 1125). Et en France en 1992, « Parmi les immigrés de Turquie, on doit distinguer les anciens Arméniens dont la moitié ont francisé leur prénom. Chez les immigrés turcs de Turquie devenus français, un sur cinq environ l'a fait » (Tribalat et al. 1996, p. 162).<sup>217</sup>

Ces résultats établis à partir de l'enquête TeO sont globalement conformes à ce que laissaient attendre les connaissances établies en onomastique à propos des prénoms portés par les personnes originaires de Turquie, sauf le résultat selon lequel le prénom le plus fréquemment porté par les femmes originaires de Turquie serait « Murielle ». En effet plusieurs spécialistes de la Turquie m'ont indiqué qu'il est clair que le prénom « Murielle » n'est aucunement connu pour être utilisé par les personnes originaires de Turquie, ni en Turquie ni en France, ni à l'état civil ni comme prénom d'usage dans les interactions avec les francophones. De fait, « Murielle » n'est ni un prénom turc ni un prénom turc francisé, et ce prénom se prononcerait différemment en turc et en français. Vérification faite, la base de données TeO contient un nombre de « Murielle » originaire de Turquie qui est très faible (<5), et c'est leur forte pondération qui leur donne une telle importance dans les résultats.<sup>218</sup> Or, aucune de ces « Murielle » n'est de nationalité turque ou de religion musulmane, ni née d'un seul parent né turc ou d'un seul parent musulman. J'en conclus que ce résultat est un artefact de la base de données TeO, dû aux fortes variations de pondération d'une personne originaire de Turquie à l'autre.

**Les prénoms « arabo-musulmans ».** En France en 2008, 8 % des individus du champ retenu portent un prénom que la nomenclature de l'INED catégorise comme « arabo-musulman » (tableau 6.7), c'est-à-dire un prénom usuellement attribué par les locuteurs de langue arabe ou par des populations non arabophones mais de religion musulmane (berbérophones, turcophones, kurdophones, persanophones, de langues austronésiennes, etc.). La plupart des porteurs de prénom « arabo-musulman » sont originaires du Maghreb (73 %) ou de Turquie (10 %) (tableau 6.13b) et, plus généralement, d'origine musulmane (84 %) (tableau 6.15b).<sup>219</sup>

<sup>215</sup> « Emine » est la forme turque du prénom arabe « Amina », qui signifie « se sentir en sécurité », du nom de la mère du prophète Mohamed. Ce prénom est classique en Turquie dès la période ottomane.

<sup>216</sup> Ces observations semblent globalement fiables, même si dans le détail le grand répertoire et la grande plasticité des prénoms turcs rendent leur analyse difficile. En effet certains prénoms turcs d'origine supposément arabe (mais non utilisés dans le monde arabophone) sont inventés en Turquie ; et certains prénoms turcs d'origine arabe (mais dont l'étymologie a été sinon oubliée, du moins banalisée) se sont sécularisés, comme « Mehmet ».

<sup>217</sup> « Les Turcs devenus français se déclarent dans près d'un tiers des cas doubles nationaux et ont, dans une proportion voisine, francisé leur nom ou prénom. Mais la francisation ne touche que 20 % de ceux d'ethnie turque » (Tribalat et al. 1996, p. 186). Pour un témoignage personnel d'une jeune femme française d'origine turque alévie sur la francisation de son prénom (de « Çigdem » à « Claire »), voir Koç 2021.

<sup>218</sup> Conformément aux règles du secret statistique du [CASD](#) et de l'[INSEE](#), et pour qu'aucune information ne permette d'identifier directement ou indirectement une personne physique (ni d'inférer de nouvelles informations sur elle), je ne peux publier plus de précisions sur les résultats obtenus sur des effectifs aussi faibles (<5).

<sup>219</sup> Cela fait écho à l'observation des auteurs selon laquelle « Un prénom arabo-musulman désigne effectivement dans 82 % des cas une personne originaire du Maghreb (immigré ou descendant d'immigrés) » (Coulmont et Simon 2019, p. 2).



On a vu que depuis les environs du 7<sup>e</sup> siècle les prénoms « arabo-musulmans » peuvent être issus du prénom d'un personnage du Coran, du prénom d'un proche ou successeur de Mohamed ou d'un prénom profane d'origine principalement arabe mais aussi berbère, persane, etc. (Annexe 6A, p. 267 et tableau 6.3). En l'occurrence en métropole en 2008, les prénoms « arabo-musulmans » les plus portés, qui pour la plupart sont ceux qui sont les plus portés par les personnes originaires du Maghreb, comprennent (tableau 6.14b) :

- ✓ des prénoms du Coran comme « Mohamed », <sup>220</sup> « Karim », <sup>221</sup> et « Ahmed » <sup>222</sup> ;
- ✓ des prénoms de proches ou successeurs de Mohamed, comme « Fatima » <sup>223</sup> et « Khadija », <sup>224</sup> mais aussi « Rachid » <sup>225</sup> ;
- ✓ des prénoms profanes d'origine arabe comme « Malika », <sup>226</sup> « Samira » <sup>227</sup> et « Leila » <sup>228</sup> ;
- ✓ un prénom non arabe mais qui pourrait passer pour coranique : « Yanis » <sup>229</sup>.

Ces prénoms, courants au Maghreb, ne signalent pas une position sociale ou un degré de religiosité particulier.

• **Les prénoms des habitants originaires d'Afrique (hors Maghreb) et les prénoms « africains » en métropole en 2008**

**Les prénoms des personnes originaires d'Afrique.** En France en 2008, 2 % des individus du champ retenu sont originaires d'autre Afrique, c'est-à-dire d'Afrique hors Maghreb (tableau 6.13b). Ils portent des prénoms « français » (32 %) plus souvent qu'« africains » (21 %), « internationaux ou anglophones » (13 %) ou « arabo-musulmans » (12 %), la part de prénoms « africains » étant toutefois vraisemblablement sous-estimée puisque relativement beaucoup d'habitants originaires d'autre Afrique portent des prénoms « inclassables » (14 %) (tableau 6.13a). Une telle diversité d'origines ne saurait surprendre, puisque l'Afrique subsaharienne a été colonisée par des anglophones aussi bien que des francophones, et puisqu'elle comporte des musulmans (majoritaires au Sahel) et des chrétiens (majoritaires plus au sud) mais aussi des animistes, tous dotés de stocks de prénoms distincts (encadré 6.16). Alternativement, si à la suite des spécialistes de l'Afrique (comme Moreau 2001) on distinguait en Afrique des prénoms « chrétiens » et « musulmans », on observerait

<sup>220</sup> « Mohamed » est le prénom arabe du prophète de l'islam, qui signifie « le très loué », *muḥammad* étant en arabe le participe passif du verbe *ḥammada*, qui est l'itératif (« louer et louer ») du verbe *ḥamida* (« louer »).

<sup>221</sup> « Karim » est un prénom arabe profane qui signifie « généreux ».

<sup>222</sup> « Ahmed » est un autre nom (arabe) de Mohamed (il appartient à la même famille lexicale), qui signifie « plus loué [que tout autre] » ou « particulièrement loué ».

<sup>223</sup> « Fatima » est un prénom arabe issu (du participe actif féminin) d'un verbe qui signifie « sevrer (un petit) ». C'est un prénom préislamique, typique d'une société d'éleveurs, qui est devenu religieux, parce que porté par la fille préférée du prophète Mohamed et épouse du 4<sup>e</sup> calife (et mère de Hasan et Husayn).

<sup>224</sup> « Khadija » est un prénom arabe issu d'un nom de sens passif qui signifie « prématuré », du verbe *khadaja*, qui signifie « mettre bas un petit prématuré ». C'est un prénom préislamique, typique d'une société d'éleveurs, porté par la première épouse du prophète Mohamed et première personne convertie à l'islam.

<sup>225</sup> « Rachid » est un prénom arabe qui signifie « bien guidé » et qui renvoie aux quatre premiers califes, ce qui en fait un prénom typiquement sunnite.

<sup>226</sup> « Malika » est un prénom arabe profane, qui peut être le féminin de *malik* (« roi ») ou de *mālik* (« possesseur », éventuellement en référence à l'imam fondateur de la jurisprudence malékite, qui prédomine au Maghreb). Si le prénom « Malika » est plutôt compris comme signifiant « reine », c'est sans doute parce que la prononciation dialectale de *mālika* donnerait « Malka ».

<sup>227</sup> « Samira » est un prénom arabe profane qui signifie « compagnon de discussion du soir ».

<sup>228</sup> « Leila » est un prénom arabe profane qui signifie « nuit ».

<sup>229</sup> « Yanis » est un prénom grec d'origine hébraïque, qui pourrait désigner le personnage coranique de Jean le Baptiste.

que parmi les prénoms des personnes originaires d'Afrique environ 52 % sont « chrétiens » et environ 33 % « musulmans ».<sup>230</sup>

Les prénoms les plus portés par les personnes originaires d'Afrique sont : « Marcellin », « Jordan », « Mamadou », « Mohamed » et « Kevin » pour les hommes, et « Aurélie », « s »,<sup>231</sup> « Aminata », « Mariam » et « Fatoumata » pour les femmes (tableau 6.14a). Si certains de ces prénoms sont catégorisés par la nomenclature de l'INED comme « français » (« Marcellin »<sup>232</sup> et « Aurélie »<sup>233</sup>), « internationaux ou anglophones » (« Jordan »,<sup>234</sup> « Kevin »<sup>235</sup> et « Mariam »<sup>236</sup>) ou « arabo-musulmans » (« Mohamed »<sup>237</sup>), d'autres sont catégorisés comme « africains » : c'est le cas de « Mamadou »,<sup>238</sup> « Fatoumata »<sup>239</sup> et « Aminata ».<sup>240</sup> En France en 1992,

« Chez les Africains noirs, le recours à la nationalité française est plus fréquent et la double appartenance plus souvent invoquée, mais l'on change rarement de prénom, car beaucoup ont déjà un prénom français. C'est particulièrement vrai de certaines ethnies : environ 80 % des personnes de langue[s]<sup>241</sup> kwa [d'Afrique de l'ouest] ou [...] bantou [d'Afrique centrale et australe] n'ayant pas francisé leur prénom sont dans cette situation. C'est encore le cas d'un tiers des populations à langue unique. Par contre, presque tous les Mandés [du Mali] ou Wolof [du Sénégal] devenus français ont un prénom africain. Le faible recours des immigrés d'Afrique noire à la francisation des prénoms est donc tout relatif : il est inutile pour la plupart mais peu utilisé par les ethnies qui pourraient y recourir » (Tribalat et al. 1996, p. 162).

**Les prénoms « africains ».** En France en 2008, seuls 1 % des individus du champ retenu portent un prénom que la nomenclature de l'INED catégorise comme « africain » (tableau 6.7),

---

<sup>230</sup> Cet ordre de grandeur est calculé, à titre indicatif, en regroupant dans les prénoms « chrétiens » les prénoms catégorisés par la nomenclature de l'INED comme « français », « latins », « internationaux ou anglophones » et d'« autre Europe », et en regroupant dans les prénoms « musulmans » les prénoms catégorisés par la nomenclature de l'INED comme « arabo-musulmans » et « africains ».

<sup>231</sup> Prénom « inclassable » trop rare pour pouvoir être mentionné sans risque de réidentification.

<sup>232</sup> « Marcellin » est la forme française d'un nom de famille latin ; saint Marcellin était un pape et martyr du 4<sup>e</sup> siècle. Ce prénom est relativement courant dans l'Afrique noire francophone chrétienne.

<sup>233</sup> « Aurélie » est la forme française féminisée du nom de famille latin « Aurelius », qui signifie « doré », du nom d'un philosophe et empereur romain du 2<sup>e</sup> siècle. Le fait que ce prénom soit notamment attribué à des descendantes d'immigré d'Afrique est attesté par le fait que, parmi les « Aurélie » listées sur la page Wikipedia dédiée à ce prénom, se trouvent notamment Aurélie Konaté, actrice française d'origine guinéenne et Aurélie Kamga, sprinteuse française d'origine camerounaise.

<sup>234</sup> « Jordan » est la forme anglaise (mais aussi française) du prénom d'origine hébraïque qui désigne le fleuve Jourdain.

<sup>235</sup> « Kevin » est la forme anglaise d'un prénom gaélique qui signifie « belle naissance » ; saint Kevin était un abbé irlandais du 6<sup>e</sup> siècle.

<sup>236</sup> « Mariam » est une forme grecque mais aussi arabe d'un prénom d'origine hébraïque de signification incertaine, du nom de la sœur aînée de Moïse, dans la Bible hébraïque.

<sup>237</sup> Voir ci-dessus.

<sup>238</sup> « Mamadou » est l'une des nombreuses formes ouest-africaines du prénom arabe « Mohamed » (voir ci-dessus). « Mohamed » peut se décliner en « Mamadou » ou « Madou » chez les Bambaras, « Mahamadou » ou « Mamoudou » ou encore « Mahmoud » chez les Soninkés, « Modou » chez les Wolofs, etc.

<sup>239</sup> « Fatoumata » est la forme ouest-africaine du prénom arabe « Fatima », du prénom de la fille préférée du prophète Mohamed et épouse du 4<sup>e</sup> calife. Ce prénom est catégorisé comme « africain » par la nomenclature de l'INED mais comme « arabo-musulman » par la nomenclature synthétique. Comme le prénom « Mohamed », « Fatima » peut se décliner en « Hatouma » ou « Hatoumata » chez les Soninkés, etc.

<sup>240</sup> « Aminata » est la forme ouest-africaine du prénom arabe « Amina », qui signifie « se sentir en sécurité », du nom de la mère du prophète Mohamed. Ce prénom est catégorisé comme « africain » par la nomenclature de l'INED mais comme « arabo-musulman » par la nomenclature synthétique.

<sup>241</sup> C'est moi qui ajoute le pluriel à « langue[s] » sachant qu'il existe des dizaines de langues kwa et des centaines de langues bantou.

c'est-à-dire un prénom qui n'est ni « français » ni « international ou anglophone » ni « arabo-musulman », mais spécifique aux locuteurs des langues d'Afrique subsaharienne. La plupart des habitants porteurs de ces prénoms dits « africains » sont originaires d'autre Afrique (90 %) (tableau 6.13b)<sup>242</sup> et sont d'origine musulmane (86 %) (tableau 6.15b).

Les prénoms spécifiquement « africains » les plus portés en France en 2008 ne sont que rarement des prénoms chrétiens (qui sont plutôt catégorisés comme « français » ou « internationaux ou anglophones »), mais surtout des prénoms musulmans d'origine arabe (encadré 6.16 ; sur les variantes ouest-africaines des prénoms arabes, voir Thurston 2013). En métropole en 2008, les prénoms catégorisés comme « africains » les plus portés sont (tableau 6.14b) :

- ✓ des prénoms ouest-africains issus des prénoms du Coran (y compris théophores), comme « Mamadou »,<sup>243</sup> « Abdoulaye »,<sup>244</sup> « Ibrahima »,<sup>245</sup> « Amadou »,<sup>246</sup> « Awa »<sup>247</sup> et « Aya »<sup>248</sup> ;
- ✓ des prénoms ouest-africains issus des prénoms des proches ou successeurs du prophète Mohamed, comme « Aminata »<sup>249</sup> et « Fatoumata »<sup>250</sup> ;
- ✓ des prénoms africains non musulmans, comme « Gloire » et « Maina ».

En Afrique de l'ouest, les prénoms musulmans susmentionnés ne signalent pas d'origine ethnolinguistique particulière, qu'elle soit wolof, sérère, peul, bambara, malinké, dioula, soninké, etc. Au contraire, le fait que ces prénoms soient communs à de nombreuses origines ethniques peut contribuer à expliquer qu'ils sont les prénoms « africains » les plus courants en France. Par contraste, l'absence de prénoms africains « coutumiers », c'est-à-dire spécifiques à une origine ethnique donnée, peut s'expliquer par le fait qu'aucun d'entre eux n'est suffisamment fréquent pour figurer parmi les prénoms « africains » les plus fréquemment portés en France.

#### **Encadré 6.16. Les prénoms en Afrique**

Les prénoms qui sont usuellement attribués aux nouveau-nés par les habitants d'Afrique subsaharienne constituent un champ si vaste et si varié et, à ce jour, si peu étudié — relativement — qu'il est difficile d'en dresser un panorama global (Mensah et Rowan 2019). Le système onomastique le mieux connu est celui des Zoulous d'Afrique australe (Herbert 1995). D'après les recherches des spécialistes, on peut toutefois considérer que les prénoms en Afrique sont composés de trois stocks distincts : (i) les prénoms « coutumiers » (aussi appelés « traditionnels » ou « ethniques » ou « animistes », qui sont eux-mêmes extraordinairement variés) ; (ii) les prénoms « musulmans » ; et (iii) les prénoms

<sup>242</sup> Cela fait écho à l'observation des auteurs selon laquelle « un prénom "africain" est porté dans 95 % des cas par un ou une originaire d'Afrique subsaharienne » (Coulmont et Simon 2019, p. 2).

<sup>243</sup> Voir ci-dessus.

<sup>244</sup> « Abdoulaye » est la forme ouest-africaine du prénom arabe « Abdallah », qui signifie « serviteur d'Allah ».

<sup>245</sup> « Ibrahima » est une forme ouest-africaine de « Ibrahim », qui est la forme arabe (coranique) du prénom d'origine hébraïque « Abraham ».

<sup>246</sup> « Amadou » est la forme ouest-africaine de « Ahmed », qui est un autre nom (arabe) de Mohamed et qui signifie « plus loué [que tout autre] » ou « particulièrement loué ».

<sup>247</sup> « Awa » est l'une des formes ouest-africaines du prénom arabe « Hawa », cognat du prénom français d'origine hébraïque « Ève », du nom de la première femme, dans la Bible hébraïque. En Afrique de l'ouest et notamment chez les Bambaras, Malinkés et Dioulas, c'est un prénom souvent attribué à la fille issue d'une paire de jumeaux des deux sexes (« Adama » pour le garçon et « Hawa » pour la fille) ainsi qu'à l'une des filles d'une paire de jumelles (« Adam » et « Hawa »). Quant aux garçons d'une paire de jumeaux, ils sont souvent prénommés « Alassane » et « Fousseyni ».

<sup>248</sup> « Aya » est un prénom arabe qui signifie « verset » (du Coran).

<sup>249</sup> Voir ci-dessus.

<sup>250</sup> Voir ci-dessus.

« chrétiens ». Notons que ces trois grandes catégories de prénoms, qui peuvent être diversement combinées pour constituer le nom personnel d'un nouveau-né, diffèrent des catégories distinguées par la nomenclature de l'INED utilisée ci-dessus.

#### **i. Les prénoms coutumiers en Afrique**

Les prénoms d'Afrique subsaharienne préislamique et préchrétienne avaient en général un contenu sémantique : ils décrivaient la période (année, saison, jour de la semaine), le lieu ou les modalités de la naissance, la position dans la fratrie, l'apparence physique du nouveau-né, les émotions des parents, une attitude face à la vie, un souhait, etc. (Saarelma 2009, p. 194-195). Les dialectes peuls, par exemple, ont des prénoms coutumiers qui désignent le jour et le mois de naissance, le rang dans la fratrie, le nombre et le sexe des germains déjà nés et/ou décédés, la gémellité, le fait d'être né après la mort de son père, le statut libre ou servile, etc. (Parietti et Tourneux 2018, p. 581-583 ; voir aussi Barry 1996). De même, chez les Bambaras, certains parents prénomment leurs enfants selon leur sexe et leur rang dans la fratrie, même si ces prénoms coutumiers ont longtemps été dénigrés suite à la diffusion de l'islam : le premier fils est « N'Tji », le deuxième « Zan » et le troisième « N'Golo » ; la première fille est « Niélé », la deuxième « Nia » et la troisième « Gnènè ». En Afrique de l'ouest, il semble qu'il était fréquent qu'un individu porte un prénom secret, qui est le vrai prénom et dont la dissimulation est censée conjurer le mauvais sort et protéger contre la sorcellerie (Saarelma 2009, p. 194). Ces pratiques perdurent, par exemple chez les Mossi du Burkina Faso dans les années 1970 et 1980, chez qui certains prénoms

« accomplissent une action rituelle (réparation d'une faute à l'égard des règles sociales, résolution des conflits familiaux, lutte contre la mortalité infantile) : mais tous n'ont pas une intentionnalité si active. Les prénoms allusifs sont probablement les plus nombreux » (Bonnet 1996).

Dans l'ancien Congo belge, quand au début des années 1970 le Président Mobutu décide de la « zaïrianisation » des toponymes (le pays est renommé « Zaïre », la capitale Léopoldville devient « Kinshasa », etc.), il lance aussi une campagne de « zaïrianisation » des noms des citoyens, invités à remplacer leurs prénoms chrétiens par des prénoms (ou post-noms) censés témoigner d'une « authenticité » africaine. Dans un passionnant [entretien accordé à une chaîne télévisée belge en 1972](#), le Président Mobutu Sese Seko, dont le prénom de naissance était « Joseph-Désiré », affirme au journaliste belge qui l'interroge :

« Moi, voyez-vous, je suis zaïrois, africain, bantou, comment voulez-vous que je puisse porter un nom juif, moi qui ne suis pas juif ? Connaissez-vous un seul juif au monde qui porte un nom zaïrois ? Et pourquoi moi ? Croyez-vous vraiment que ce serait intelligent, pensable de la part d'une jeune fille de mon pays de porter le prénom de "Chantal", de "Marie-France", de " Marie-Flore" : ça veut dire quoi ? »

Chez les Joola de Casamance (Sénégal) à la fin du 20<sup>e</sup> siècle, « beaucoup ont à la fois un prénom joola plus un prénom chrétien ou musulman, officialisé ou non (sans compter les surnoms), et qu'on appelle soit exclusivement par l'un, soit exclusivement par l'autre, soit encore tantôt par l'un, tantôt par l'autre, sans que l'acte officiel de déclaration paraisse avoir vraiment de pertinence sur les pratiques » (Moreau 2001, p. 543).

#### **ii. Les prénoms musulmans et chrétiens en Afrique**

Les conversions à l'islam à partir des environs du 10<sup>e</sup> siècle, puis les conversions au christianisme à partir du 19<sup>e</sup> siècle, de pair avec la présence d'administrations coloniales et d'un système d'enseignement primaire à partir là aussi du 19<sup>e</sup> siècle, ont parfois conduit à

l'abandon partiel des systèmes de noms personnels africains, y compris les prénoms coutumiers appelés « mécréants », « païens », « non civilisés », etc. (Saarelma 2009, p. 198-199). Dans l'Afrique coloniale comme dans l'Europe médiévale (Annexe 6A, p. 267), les populations tendent à abandonner leurs prénoms locaux et peuvent adopter des (formes locales des) prénoms de saints chrétiens pour signaler leur religion, même si ces prénoms d'origine hébraïque, grecque ou latine sont sémantiquement opaques pour eux ; et la réduction du nombre de prénoms différents conduit à l'adoption de surnoms qui, devenus héréditaires avec l'essor de l'appareil d'État, deviennent des noms de famille (Saarelma 2009, p. 207-208). Ainsi des populations qui portaient des prénoms sémantiquement transparents, où parfois chaque individu était le seul à porter son prénom, ont-elles adopté des prénoms islamiques (d'origine principalement arabe) ou des prénoms chrétiens, souvent des noms de saints (d'origine principalement francophone mais aussi lusophone) pour les catholiques, et des noms de l'Ancien Testament (d'origine anglophone) pour les protestants, voire aussi des prénoms théophores issus des langues locales (Herbert 1995 ; Saarelma 2009). Par exemple en Casamance (Sénégal) de 1969 à 1999, certains groupes ethniques attribuent à l'état civil des premiers prénoms largement musulmans (Peul, Manding, Wolof), d'autres groupes attribuent des prénoms largement chrétiens (Joola, Manjak, Mankagne), et d'autres encore à peu près autant de prénoms chrétiens que musulmans (Sereer), mais aucun groupe n'attribue des prénoms coutumiers à plus de 20 % des nouveau-nés (Moreau 2001, p. 548).

- **Les prénoms des habitants originaires d'Asie et les prénoms « asiatiques » en métropole en 2008**

**Les prénoms des personnes originaires d'Asie.** En France en 2008, 1 % des individus du champ retenu sont originaires d'Asie (tableau 6.13b). Ils portent des prénoms « français » (29 %) aussi souvent que « asiatiques » (29 %), plus souvent que des prénoms « internationaux ou anglophones » (10 %), mais la part de prénoms « asiatiques » est vraisemblablement sous-estimée puisque relativement beaucoup d'habitants originaires d'Asie portent des prénoms « inclassables » (18 %) (tableau 6.13a).

Les prénoms les plus portés par les personnes originaires d'Asie sont « David », « Nicolas », « Vincent », « Michel » et « Kevin » pour les hommes, et « Marie », « Sophie », « s », <sup>251</sup> « Nina » et « Zara » pour les femmes (tableau 6.14a). Si deux de ces prénoms sont catégorisés comme « internationaux ou anglophones » (« Kevin » et « Nina »), l'un comme « arabo-musulman » (« Zara ») et un autre, trop rare pour pouvoir être mentionné sans risque de réidentification, comme « asiatique », la plupart sont « français » : « David », « Nicolas », « Vincent », « Michel », « Marie » et « Sophie ». D'une part, les immigrés d'Asie ont assez souvent francisé leur prénom lors de l'acquisition de la nationalité française. Ainsi en France en 1992,

« La francisation est la plus répandue parmi les immigrés du sud-est asiatique : environ 35 % pour les hommes comme pour les femmes. Le caractère particulier de leur migration introduit une coupure avec le pays d'origine qui facilite à la fois l'acquisition de la nationalité française et la francisation du prénom » (Tribalat et al. 1996, p. 162).

Les immigrés chinois eux aussi adaptent généralement leur prénom à leur lieu d'installation, pour faciliter la communication avec les locaux (Lawson 2016, p. 172). D'autre part, les immigrés d'Asie tendent à attribuer à leurs enfants des prénoms « français ». Ainsi dès les années 1990, « Quand on commence à s'intéresser aux prénoms [attribués aux descendants d'immigrés hmong du Laos], on s'entend souvent dire que la question n'a pas, ou n'a plus d'intérêt : "Ah vous étudiez les prénoms... il n'y a plus rien aujourd'hui, nous donnons des

<sup>251</sup> Prénom « asiatique » trop rare pour pouvoir être mentionné sans risque de réidentification.

prénoms français" » (Hassoun 1995, p. 252). Plus généralement, seulement 48 % des personnes (du 13<sup>e</sup> arrondissement de Paris) ayant un nom de famille asiatique portent un prénom renvoyant à cette aire culturelle (Fourquet et Manternach 2019a, p. 175).

**Les prénoms « asiatiques ».** En France en 2008, moins de 1 % des individus du champ retenu portent un prénom que la nomenclature de l'INED catégorise comme « asiatique » (tableau 6.7), c'est-à-dire un prénom usuellement attribué par les locuteurs des langues d'Asie.<sup>252</sup> La plupart des habitants porteurs de prénoms « asiatiques » sont originaires d'Asie (80 %) (tableau 6.13b). Ils peuvent être d'origine chrétienne (39 %), sans origine religieuse (29 %) ou encore d'origine bouddhiste, hindouiste, etc. (24 %) (tableau 6.15b). Comme les prénoms attribués en Afrique (encadré 6.16), les prénoms attribués en Asie constituent un champ si vaste et varié et, à ce jour, si peu étudié — relativement — qu'il est impossible d'en dresser un panorama global (Choo et al. 2016). Si les prénoms « asiatiques » les plus portés en France en 2008 comprennent des prénoms vietnamiens comme « Loan » et « Louane »<sup>253</sup> (tableau 6.14b), les autres prénoms « asiatiques » les plus fréquents dans l'enquête TeO sont trop rares pour pouvoir être mentionnés sans risque de réidentification — ou pour aboutir à des résultats robustes.

- **Les prénoms des habitants d'origine juive et les prénoms « juifs » en métropole en 2008**

**Les prénoms des personnes d'origine juive.** En France en 2008, 1 % des individus du champ retenu sont d'origine juive, c'est-à-dire qu'ils ont deux parents juifs, ou un parent juif et l'autre sans religion, ou encore un père juif et une mère d'une autre religion (tableau 6.15b). Par contraste, 66 % des individus du champ retenu sont d'origine chrétienne, 23 % sans origine religieuse, 9 % d'origine musulmane et 1 % d'origine bouddhiste ou hindouiste. Dans l'enquête TeO, la plupart des individus d'origine juive portent un prénom « français » (59 %) plutôt que « juif » (6 %) (tableau 6.15a). De fait, les prénoms les plus portés par les enquêtés d'origine juive sont, par ordre décroissant, « Alain », « David », « Samuel », « Raphaël » et « Sébastien » pour les hommes, et « Catherine », « Monique », « Johanna », « Véronique » et « Florence » pour les femmes. En France depuis le 19<sup>e</sup> siècle, les juifs tendent effectivement à attribuer à leurs enfants des prénoms qui recoupent de plus en plus ceux qui sont attribués par les non-juifs (Grange 2016b, p. 75-78). Cela dit, l'origine religieuse des individus continue de déterminer fortement le type de prénom qu'ils portent : en métropole en 2008, la plupart des habitants d'origine juive (59 %) ou chrétienne (82 %) ou sans origine religieuse (75 %) portent un prénom « français », tandis que la plupart des habitants d'origine musulmane portent un prénom « arabo-musulman » (73 %) plutôt que « français » (4 %) (tableau 6.15a).

**Les prénoms « juifs ».** En France en 2008, seulement 1 % des individus du champ retenu portent un prénom que la nomenclature synthétique catégorise comme « juif » (tableau 6.7), c'est-à-dire un prénom usuellement attribué par des populations de religion juive. Ces prénoms peuvent être d'origines étymologiques variées, allant de l'hébreu classique à l'hébreu moderne en passant par les langues des diasporas juives (encadré 6.5). D'après les 50 prénoms présents dans TeO que la nomenclature 2 et la nomenclature synthétique catégorisent comme « juifs », les prénoms « juifs » les plus portés en France en 2008 comprennent :

---

<sup>252</sup> Le tableau 6.7 compte moins de 1 % de prénoms « asiatiques » alors que le tableau 6.13a, calculé sur le même champ, en compte 1 %, parce que les arrondis diffèrent selon que les pourcentages sont calculés à partir de tableaux d'effectifs ou de tableaux de pourcentages.

<sup>253</sup> « Loan » est un prénom vietnamien d'origine sino-vietnamienne, qui désigne un oiseau mythologique. Ce prénom ne fait pas partie des prénoms les plus usuels au Vietnam, mais il est prononçable en français et peut se transcrire sans signe diacritique ni voyelle inconnue de l'alphabet français.

- ✓ principalement des prénoms hébraïques : « Samuel », <sup>254</sup> « Aaron », <sup>255</sup> « Élie », <sup>256</sup> « Joachim » <sup>257</sup> et « Nathanaël » <sup>258</sup> pour les hommes, et « Sarah », <sup>259</sup> « Myriam » <sup>260</sup> et « Judith » <sup>261</sup> pour les femmes ;
- ✓ aucun prénom diasporique ;
- ✓ deux prénoms israéliens : « Yaël » et « Yaëlle » <sup>262</sup>

Il s'agit pour la plupart de prénoms hébraïques de personnages de la Bible hébraïque, mentionnés aussi dans les textes sacrés chrétiens et musulmans. Et de fait, en France en 2008 la plupart des individus porteurs de prénoms « juifs » sont en fait d'origine chrétienne (45 %) ou sans origine religieuse (27 %) ou d'origine musulmane (18 %), et ils ne sont que marginalement d'origine juive (6 %) (tableau 6.15b). On a même vu ci-dessus que « Sarah » et « Myriam » font partie des prénoms les plus fréquemment portés par les femmes originaires du Maghreb (tableau 6.14a).

### **Encadré 6.17. Les prénoms théophores**

Les prénoms théophores sont des prénoms qui incluent le nom d'une divinité. On a déjà examiné les prénoms de la Bible hébraïque qui incluent le Dieu des juifs : « Daniel » (Dieu est mon juge), « Emmanuel » (Dieu avec nous), « Gabriel » (Dieu est ma force), « Michel » (qui est comme Dieu ?), « Raphaël » (Dieu a guéri), etc. (encadré 6.5). De même avec les prénoms qui incluent le Dieu des chrétiens : « Christian » et « Christine » (issus des prénoms latins qui signifient « chrétien » et « chrétienne »), et « Christophe » (issu du prénom grec qui signifie « celui qui porte le Christ »). Et de même avec les prénoms qui incluent le Dieu des musulmans, qui a donné lieu aux 99 noms de Dieu du Coran, comme « Abdallah » (serviteur de Dieu) ou « Abdelaziz » (serviteur du puissant) (Annexe 6A, p. 267). Mais les prénoms théophores s'observent dans bien d'autres contextes, avec une divinité de l'Égypte antique comme « Akhenaton » (serviteur d'Aton) ou « Isidore » (don

<sup>254</sup> « Samuel » est un prénom d'origine hébraïque, qui signifie « nom de Dieu » ou « Dieu a entendu », du nom d'un juge dans la Bible hébraïque, mentionné aussi dans le Nouveau Testament et dans le Coran.

<sup>255</sup> « Aaron » est un prénom hébraïque de signification incertaine, du nom du frère aîné de Moïse, dans la Bible hébraïque mais aussi dans le Nouveau Testament et dans le Coran.

<sup>256</sup> « Élie » est la forme française d'un prénom hébraïque qui signifie « mon Dieu est Yahvé », du nom d'un prophète de la Bible hébraïque mentionné aussi dans le Nouveau Testament et dans le Coran.

<sup>257</sup> « Joachim » est la forme française d'un prénom hébraïque qui signifie « Yahvé établit », du nom — selon la tradition chrétienne — du grand-père maternel de Jésus. Ce prénom hébraïque n'étant que très rarement attribué par les juifs, on peut considérer qu'il aurait été préférable que la nomenclature synthétique ne le catégorise pas comme « juif ».

<sup>258</sup> « Nathanaël » est la forme française d'un prénom hébraïque qui signifie « Dieu a donné », du nom de plusieurs personnages de la Bible hébraïque et du Nouveau Testament.

<sup>259</sup> « Sarah » est la forme (notamment) française d'un prénom hébraïque qui signifie « princesse », du nom de l'épouse d'Abraham et mère d'Isaac, dans la Bible hébraïque mais aussi dans le Nouveau Testament et dans le Coran. Dans la littérature française du 17<sup>e</sup> au 20<sup>e</sup> siècles (base de données Frantext), ce prénom de l'Ancien Testament est plus fréquent sous sa graphie « Sara » jusqu'au 18<sup>e</sup> siècle, et ce n'est que depuis le 19<sup>e</sup> siècle que la graphie « Sarah » est la plus fréquente (Brunet 1988, p. 144). Rappelons qu'en 1938, quand le Troisième Reich contrainst les juifs portant un prénom non-juif à ajouter à leur nom personnel, à partir du 1<sup>er</sup> janvier 1939, un prénom permettant de faciliter leur identification puis leur persécution, les prénoms additionnels imposés sont « Israël » pour les hommes et « Sara » pour les femmes (Gerhards et Hans 2009, p. 1103-1104).

<sup>260</sup> « Myriam » est une forme française d'un prénom hébraïque de signification incertaine, du nom de la sœur aînée de Moïse, dans la Bible hébraïque.

<sup>261</sup> « Judith » est une forme (notamment) française d'un prénom hébraïque qui signifie « femme juive », du nom d'un personnage de la Bible hébraïque.

<sup>262</sup> « Yaël » ou « Yaëlle » est un prénom d'origine hébraïque, qui signifie « bouquetin ». Si ce prénom est celui d'un personnage de la Bible hébraïque, il n'a jamais été porté par les juifs de l'antiquité à la première moitié du 20<sup>e</sup> siècle, et son succès en France est dû à l'influence culturelle d'Israël sur les juifs de la diaspora. C'est pourquoi je le qualifie d'« israélien ».

d'Isis) ; avec une divinité des religions grecque et romaine comme « Théodore » et « Dorothee » (don de Dieu), « Dimitri » (qui appartient à Déméter), « Denis » (Dionysos) ou « Marc » (consacré à Mars) ; ou avec une divinité nordique comme « Ingrid » (la beauté de Ing). D'autres prénoms européens sont théophores sans faire spécifiquement référence à une divinité plutôt qu'une autre, comme « Bogdan » (don de Dieu), « Dieudonné » et « Gustave » (soutien des dieux).

- **Des prénoms masculins plus souvent spécifiques aux origines de leur porteur**

Si en métropole en 2008 les immigrés et leurs enfants tendent à attribuer plus souvent des prénoms spécifiques à leurs origines *à leurs fils qu'à leurs filles*, il convient de noter que cela n'est pas particulier à la France contemporaine. C'est ce qu'on observe aussi dans la longue histoire de la diaspora juive — chez qui la culture religieuse passe par les garçons — jusqu'au 20<sup>e</sup> siècle,<sup>263</sup> dans la bourgeoisie juive parisienne au cours du 19<sup>e</sup> siècle,<sup>264</sup> chez les immigrés aux États-Unis à la fin comme au début du 20<sup>e</sup> siècle (Sue et Telles 2007, p. 1384 ; Carneiro et al. 2020), chez les immigrés et descendants d'immigrés hispaniques dans le comté de Los Angeles en 1995,<sup>265</sup> chez les immigrés (d'Europe du Sud, d'ex-Yougoslavie et de Turquie) en Allemagne des années 1980 aux années 2000,<sup>266</sup> chez les Palestiniens en Israël et en Cisjordanie au début du 21<sup>e</sup> siècle,<sup>267</sup> ou encore chez la minorité russophone de Finlande dans les années 2000 et 2010.<sup>268</sup>

<sup>263</sup> 'In general, there was a higher percentage of women bearing non-Hebrew names in the [Jewish] Diaspora up until the modern age (Spanish: Gracia; Greek: Zoe; Arabic: Hamama 'pigeon'; Yiddish: Shprinke from Esperanza)' (Lawson 2016, p. 180).

<sup>264</sup> Dans la bourgeoisie juive parisienne au cours du 19<sup>e</sup> siècle, la part des premiers prénoms qui sont des prénoms juifs ne cesse de baisser jusqu'à passer sous les 5 % à la fin du siècle, mais « la présence d'un ou de plusieurs prénoms juifs est toujours supérieure chez les hommes » (Grange 2016b, p. 78-79).

<sup>265</sup> Dans le comté de Los Angeles (le plus peuplé des États-Unis) en 1995, les immigrés et descendants d'immigrés hispaniques attribuent plus souvent des prénoms américains (et des prénoms américains intraduisibles en espagnol) à leurs filles qu'à leurs fils (Sue et Telles 2007, p. 1395-1408), peut-être en partie parce beaucoup de prénoms féminins terminent en -a en anglais comme en espagnol, alors que les terminaisons des prénoms masculins concordent moins entre l'anglais et l'espagnol (Sue et Telles 2007, p. 1389).

<sup>266</sup> En Allemagne des années 1980 aux années 2000, les immigrés (d'Europe du Sud, d'ex-Yougoslavie et de Turquie) donnent plus souvent à leurs filles qu'à leurs fils des prénoms allemands, ne serait-ce que parce qu'il y a plus de prénoms féminins que de prénoms masculins qui sont partagés par plusieurs cultures (les frontières culturelles entre prénoms sont moins élevées pour les prénoms féminins que masculins) (Gerhards et Hans 2009, p. 1118).

<sup>267</sup> Parmi des collégiens palestiniens en Israël et en Cisjordanie dans les années 2000 ou 2010, 60 % des garçons portent un prénom à connotation religieuse (y compris les prénoms les plus portés : « Muhammad », « Ahmad » et « Mahmoud »), contre 40 % des filles (Amara 2014, p. 220-222). En effet 30 % des filles (contre 20 % des garçons) portent des prénoms évoquant la nature (animaux, plantes, métaux) ou des vertus (beauté, délicatesse ou chasteté pour les filles, gloire ou bravoure pour les garçons), et 10 % des filles (contre 0 % des garçons) portent des prénoms persans ou européens (Amara 2014, p. 220).

<sup>268</sup> Dans les années 2000 et 2010, les premiers prénoms attribués par la minorité russophone en Finlande (1 % des naissances) sont plus souvent des prénoms russes qu'internationaux pour les garçons, alors que ce sont plus souvent des prénoms internationaux que russes que pour les filles (Eskola et Hämäläinen 2019, p. 207).



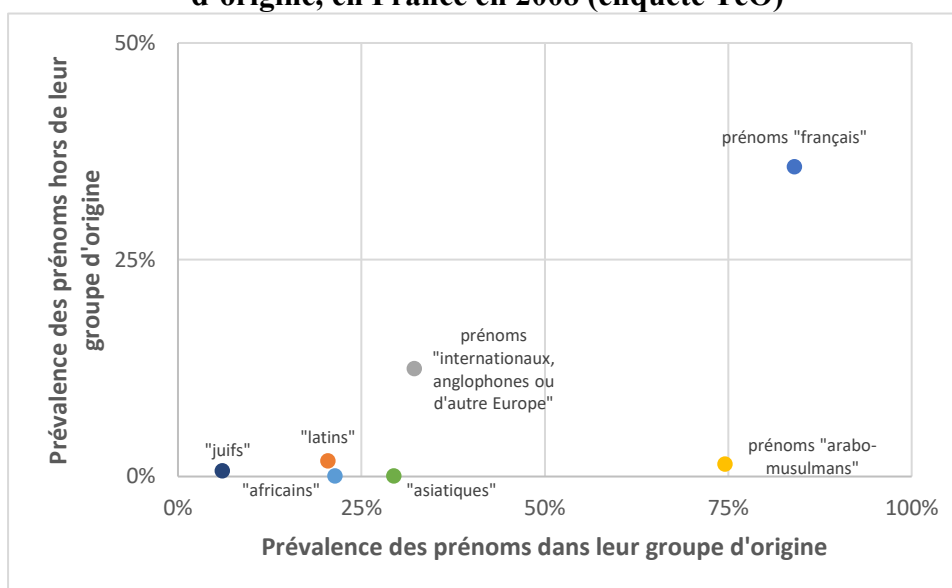
### ❖ Origine du prénom et origine de son porteur en France : une synthèse

Pour synthétiser les résultats présentés ci-dessus, je propose de caractériser les prénoms d'une origine donnée selon 2 dimensions :

- d'une part, selon qu'ils sont courants ou rares (plus ou moins prévalents) *dans le groupe dont ils sont originaires* ; et
- d'autre part, selon qu'ils sont rares ou courants *hors du groupe dont ils sont originaires* — ou autrement dit selon qu'ils sont « spécifiques » ou « non spécifiques » au groupe dont ils sont originaires.

Je rassemble les informations déjà analysées (tableaux 6.13a et 6.15a) dans la figure 6.18.

**Figure 6.18. Prévalence des prénoms dans leur groupe d'origine et hors de leur groupe d'origine, en France en 2008 (enquête TeO)**



Lecture : d'après la nomenclature de l'INED, les prénoms « français » sont portés par 84 % des membres du groupe « majoritaire » (y compris les personnes originaires des DOM) et par 36 % des membres du groupe « minoritaire » (personnes originaires d'Europe du Sud, d'autre Europe, du Maghreb, d'autre Afrique, de Turquie, d'Asie et d'ailleurs).

Champ : personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008, données pondérées, N = 42 208 (voir tableaux 6.13a et 6.15a).

NB : les données sous-jacentes à ce graphique 6.18 contiennent une erreur qui, une fois corrigée (Chapitre 7, p. 26), accroît encore la prévalence des prénoms « arabo-musulmans » dans leur groupe d'origine, de 74 % à 80 %.

Source : TeO (voir Annexe 5B, p. 183).

La Figure 6.18 appelle plusieurs observations. D'une part, on note que les prénoms « français » se situent « plus haut » que les autres, c'est-à-dire qu'ils sont les seuls prénoms à être aussi fréquemment portés hors de leur groupe d'origine. En effet, les prénoms « français » ont pour caractéristique d'être non seulement très fréquents chez les personnes du groupe « majoritaire » et chez les personnes originaires des DOM (84 %), mais aussi d'être relativement courants dans les autres groupes d'origine (36 %).

D'autre part, les prénoms « arabo-musulmans » se situent « plus à droite » que les autres prénoms non « français », c'est-à-dire qu'ils sont les seuls prénoms non « français » à être aussi fréquemment portés par leur groupe d'origine. En effet, les prénoms « arabo-musulmans » ont pour caractéristique d'être particulièrement courants dans leur groupe d'origine, c'est-à-dire chez les personnes originaires du Maghreb et de Turquie (74 %).

Ensuite, sur la Figure 6.18, les prénoms « latins », « africains », « asiatiques » et « juifs » se situent « plus à gauche » que les autres prénoms non « français », c'est-à-dire qu'ils sont les

moins fréquemment portés par leur groupe d'origine. En effet, les prénoms « latins » sont portés par seulement 20 % des personnes originaires d'Europe du Sud, les prénoms « africains » par 21 % des personnes originaires d'Afrique hors Maghreb, les prénoms « asiatiques » par 29 % des personnes originaires d'Asie, et les prénoms « juifs » sont portés par seulement 6 % des personnes d'origine juive.

Enfin, les prénoms « internationaux, anglophones ou d'autre Europe » se situent, assez logiquement, dans une position intermédiaire : ils sont portés par 32 % des personnes originaires d'autre Europe (Europe hors France, Italie, Espagne et Portugal), mais aussi par 12 % des personnes originaires d'autres groupes.

### ❖ Résumé

Parmi les naissances en France entière de 1900 à 2019, de moins en moins de prénoms attribués sont « français », de 93 % en 1900 et en 1950 à 29 % en 2019. De plus en plus de nouveau-nés se voient attribuer un prénom « international ou anglophone », de 3 % en 1900 et en 1950 à 20 % en 2019. De même, de plus en plus de nouveau-nés reçoivent un prénom « arabo-musulman », de 0 % en 1900 et en 1950 à 14 % en 2019. Les autres prénoms sont « latins », d'« autre Europe », « africains », « asiatiques » ou « inclassables ».

Parmi les personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008, 73 % portent un prénom « français » comme Sébastien et Catherine : il s'agit le plus souvent du prénom d'un personnage du Nouveau Testament ou d'un saint ou martyr chrétien. En outre, 3 % portent un prénom « latin » comme Enzo et Sabrina ; 12 % un prénom « international ou anglophone » comme Patrick et Patricia ; 2 % un prénom d'« autre Europe » comme Dimitri et Ingrid ; 8 % un prénom « arabo-musulman » comme Mohamed et Fatima ; 1 % un prénom « africain » comme Mamadou et Aminata ; moins de 1 % un prénom « asiatique » comme Louane ; et 2 % un prénom « inclassable » (trop rare pour pouvoir être mentionné sans risque de réidentification). Moins d'un tiers des personnes originaires d'Europe du Sud, d'Afrique hors Maghreb ou d'Asie portent un prénom spécifique à leurs origines, et à l'extrême seulement 6 % des individus d'origine juive portent un prénom « juif », mais 79 % des immigrés et descendants d'immigrés du Maghreb et de Turquie portent un prénom « arabo-musulman ». Non seulement les immigrés du Maghreb et de Turquie ne francisent que rarement leur prénom, mais en outre leurs descendants continuent largement de se voir attribuer des prénoms spécifiquement « arabo-musulmans », qu'il s'agisse du prénom d'un personnage du Coran, du prénom d'un proche ou successeur du prophète Mohamed ou encore d'un prénom profane d'étymologie arabe voire berbère, turque, persane, etc. L'origine religieuse des habitants continue de déterminer fortement le type de prénom qu'ils portent : en métropole en 2008, la plupart des habitants d'origine chrétienne (82 %) ou juive (59 %) ou sans origine religieuse (75 %) portent un prénom « français », tandis que la plupart des habitants d'origine musulmane portent un prénom « arabo-musulman » (73 %), plutôt que « français » (4 %).

## Annexe 7 : Le « Nicolas » du tableau page 3

Cette Annexe 7 comprend les analyses du début du Chapitre 7, p. 26.

### ❖ Analyses préliminaires sur le prénom « Nicolas »

« Nicolas » est la forme française du prénom grec « Νικόλαος » ou « Nikolaos », qui signifie « conquérant du peuple » (Lawson 2016, p. 177). La popularité de ce prénom en Europe est due principalement à saint Nicolas, Nicolas de Myre, un évêque grec d'Asie Mineure du 4<sup>e</sup> siècle. Saint Nicolas est un objet de culte dans l'empire romain d'Orient (empire byzantin), c'est-à-dire en Europe orientale, puis aussi à partir du 10<sup>e</sup> siècle dans la chrétienté latine, en Europe occidentale (Gazeau et al. 2015). Suite à la conquête musulmane de Myre (11<sup>e</sup> siècle), les reliques de saint Nicolas sont translatées de Myre jusqu'à Bari en Italie (1087), où elles se trouvent toujours. Dans l'Europe médiévale, saint Nicolas est associé aux croisades : « En Irlande, il est devenu le patron des chevaliers partis en croisade » (Gazeau 2015, p. 494 ; pour plus de précisions, voir Olivieri 2015). Mais saint Nicolas est aussi et surtout associé à la Reconquista, c'est-à-dire à la reconquête, par les royaumes chrétiens, des territoires espagnols et portugais conquis par les musulmans (8<sup>e</sup>-15<sup>e</sup> siècles). En Espagne,

« Saint Nicolas donne son nom à de très nombreuses églises de villes reprises aux musulmans dans le cadre de la Reconquista, peut-être dès les VIII<sup>e</sup>-IX<sup>e</sup> siècles, preuve s'il en est que l'évêque de Myre n'était pas un inconnu. [...] les prisonniers chrétiens des musulmans sont libérés grâce à Saint Nicolas en Espagne » (Gazeau 2015, p. 491-494).

« Le culte de saint Nicolas en Espagne naît de façon exactement parallèle à celui de saint Jacques : Nicolas est le saint qui réconforte avec tendresse le pèlerin, mais qui défend également avec force *son* pèlerin, notamment contre les Sarrasins, et donc face aux musulmans » (Cioffari 2015, p. 170 ; sur les liens divers entre saint Nicolas et la Reconquista à partir du 11<sup>e</sup> siècle, voir Cioffari 2015, p. 183-191).

Dans ce contexte, l'attribution relativement massive, en France, du prénom « Nicolas » à des petits-fils d'immigrés du Maghreb constituerait un signal fort de désir d'assimilation culturelle.

En France, saint Nicolas devient populaire à partir du 13<sup>e</sup> siècle, bien avant de devenir *santa Klaus*, le père Noël. Dans la littérature française du 17<sup>e</sup> au 20<sup>e</sup> siècles (base de données Frantext), le prénom « Nicolas » apparaît dans 452 textes sur 3 000, ce qui en fait le 34<sup>e</sup> prénom le plus souvent mentionné (Brunet 1988, p. 131). En France au 19<sup>e</sup> siècle, « Nicolas » est le 28<sup>e</sup> premier prénom (et le 33<sup>e</sup> deuxième prénom) le plus fréquemment attribué aux garçons (Dupâquier et al. 1987, p. 41). Ce prénom décline tout au long du 19<sup>e</sup> siècle même s'il résiste en Lorraine (dont saint Nicolas est le patron) et plus généralement dans le quart nord-est (Dupâquier et al. 1987, p. 83-84, 86, 114, 116). Dans un article intitulé « Les enfants de Michel et Martine Dupont s'appellent Nicolas et Céline », devenu un « classique », Desplanques (1986) prolonge l'histoire de la fréquence d'attribution du prénom « Nicolas » au 20<sup>e</sup> siècle :

« le temps de montée des prénoms est variable d'un prénom à un autre. Nicolas, qui débute son ascension vers 1960, avant Sébastien, parvient à son sommet [en 1980] quatre ans après ce dernier. Comme Philippe, Nicolas fait en quelque sorte partie du vieux fonds de prénoms. Par exemple, dans le Vexin, c'était un prénom très fréquent au XVIII<sup>e</sup> siècle » (Desplanques 1986, p. 72 ; voir aussi Besnard et Grange 1993, p. 274).

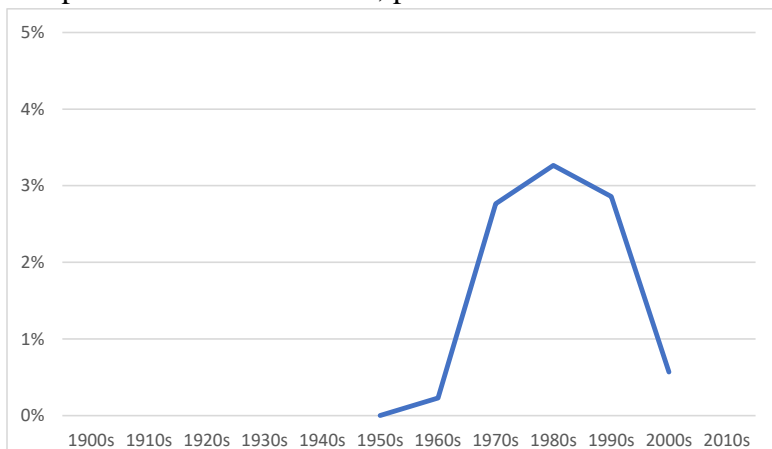
Comme le montre le [classement des prénoms en France depuis 1900](#), « Nicolas » devient de plus en plus fréquent à partir des années 1960 : « Nicolas est l'un des deux enfants qui, à la fin des années soixante sur le petit écran, s'endorment chaque soir bercés par Nounours » (Desplanques 1986, p. 72). En outre, « l'émergence de la plupart des prénoms qui atteignent à la notoriété commence toujours dans les catégories aisées : c'était le cas pour Nathalie, cela le reste pour Sébastien, Céline et Nicolas. [...] Entre 1965 et 1969, la fréquence des prénoms Sébastien et Nicolas, encore faible, décroît depuis les cadres jusqu'aux agriculteurs » (Desplanques 1986, p. 75). « Nicolas » est le prénom masculin le plus donné en France de 1980 à 1982 (puis de nouveau en 1995) : en 1980 on compte près de 8 « Nicolas » pour 100 naissances masculines. Mais depuis le début des années 1980 le prénom « Nicolas » est de moins en moins attribué : on compte actuellement 0,2 « Nicolas » pour 100 naissances masculines, soit autant que dans les années 1950, au tout début du décollage de ce prénom au 20<sup>e</sup> siècle.

L'enquête TeO, dont l'échantillon maximal non pondéré comprend 377 « Nicolas, reflète-t-elle bien l'évolution de la fréquence d'attribution de ce prénom depuis les années 1950 ? Pour le savoir, je compare l'incidence décennale de ce prénom dans l'enquête TeO (figure 7.1a) et dans le fichier INSEE des prénoms (figure 7.1b). Il en ressort que l'échantillon TeO retenu pondéré (représentatif des personnes de 18-50 / 18-60 ans en ménage ordinaire et de leurs enfants de 0-17 ans habitant en France métropolitaine en 2008) reflète bien le mouvement général de hausse — des années 1950 aux années 1980 — puis de baisse — depuis les années 1980 — de la fréquence d'attribution de ce prénom, ce qui, après les multiples tests déjà réalisés (Annexe 5A, p. 147), achève de me convaincre que **l'échantillonnage de cette enquête et sa variable « prénom » sont de (très) bonne qualité**. Je peux donc avancer vers la reproduction « au plus près » du tableau publié par Coulmont et Simon (2019, p. 3).

## Fréquence d'attribution du prénom « Nicolas » en France, selon l'enquête TeO et selon le Fichier INSEE des prénoms

### ENQUÊTE TeO

Figure 7.1a. Part des habitants masculins en 2008 qui se prénomment « Nicolas », par décennie de naissance

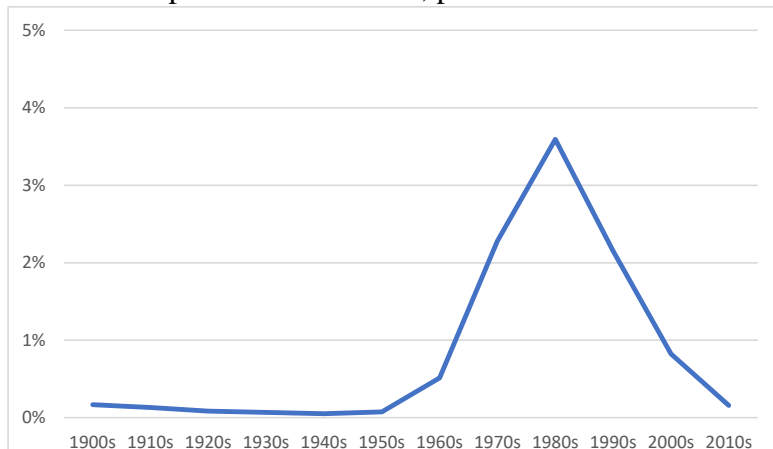


Champ : hommes du champ retenu (personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008) qui sont nés en France entière, données pondérées, N = 16 006.

Source : TeO (voir Annexe 5B, p. 183).

### FICHIER INSEE DES PRÉNOMS

Figure 7.1b. Part des nouveau-nés masculins qui se voient attribuer le prénom « Nicolas », par décennie de naissance



Champ : naissances masculines en France entière (hors Mayotte) de 1900 à 2019, N = 42 887 293.

Source : [Fichier INSEE des prénoms](#).

### ❖ Les prénoms portés en France en 2008, par origine et par « génération » de leur porteur

On a présenté ci-dessus le Top 5 des prénoms les plus portés en France en 2008 selon l'origine géographique de leur porteur (tableau 6.14a), et on a commenté ces Top 5 dans le détail (Annexe 6B, p. 285). Mais qu'observe-t-on si on présente non plus le Top 5 mais le Top 20 des prénoms, et ce par « génération », c'est-à-dire selon que les individus sont immigrés (G1) ou enfants d'immigrés (G2) ou petits-enfants d'immigrés (G3) ? Savoir cela permet de connaître le contexte immédiat dans lequel s'inscrit le tableau de Coulmont et Simon (2019, p. 3), et notamment de savoir si « Nicolas » est le seul résultat surprenant.

#### • Les prénoms des habitants du groupe « majoritaire », par « génération »

Si l'on examine tout d'abord le Top 20 des prénoms des habitants du groupe « majoritaire » (c'est-à-dire des personnes qui ne sont ni immigrées ni nées dans un DOM et dont aucun parent ou grand-parent n'est immigré ni né dans un DOM), selon qu'ils sont enquêtés ou enfants des enquêtés, aucun résultat n'apparaît particulièrement surprenant (tableau 7.2).

La grande majorité des prénoms les plus portés sont, d'après la nomenclature de l'INED, « français » (en blanc dans le tableau 7.2), sauf quelques-uns (en grisé) qui sont catégorisés comme « internationaux ou anglophones », comme « Patrick » et « Patricia » chez les enquêtés du groupe « majoritaire », et comme « Tom », « Anthony », « Léo », « Théo », « Léa », « Laura », « Clara » et « Emma » chez les enfants d'enquêtés qui sont du groupe « majoritaire ». La plupart des prénoms « français » sont issus du prénom d'un personnage du Nouveau Testament ou du prénom d'un saint ou martyr chrétien, leur origine étymologique pouvant être très variée, notamment hébraïque, grecque, latine, germanique ou celtique (Annexe 6A, p. 267 et Annexe 6B, p. 285). En outre, les prénoms les plus portés passent de « Sébastien » et « Catherine » (enquêtés du groupe « majoritaire ») à « Maxime » et « Manon » (enfants d'enquête qui sont du groupe « majoritaire »), ce qui correspond globalement bien à l'évolution temporelle de la mode des prénoms, telle que décrite non seulement par l'enquête TeO mais aussi par le fichier INSEE des prénoms (tableau 5.5). Enfin, les 20 prénoms masculins les plus portés suffisent à prénommer 35 % des hommes enquêtés et 33 % des fils des personnes enquêtées. Et les 20 prénoms féminins les plus portés permettent de prénommer 31 % des femmes enquêtées et 28 % des filles des personnes enquêtées. Autrement dit, quand on passe des enquêtés à leurs enfants (tous du groupe « majoritaire »), plus d'individus portent des prénoms relativement originaux, un phénomène qu'on a déjà observé et analysé (Annexe 5A, p. 147). Le prénom des fils étant parfois encore tiré de celui d'un aïeul, alors que celui des filles est plus souvent choisi pour être à la mode ou distingué, les femmes sont globalement plus nombreuses à porter des prénoms relativement originaux.

Au total, le Top 20 des prénoms des habitants du groupe « majoritaire » selon la « génération » (tableau 7.2) est conforme à ce que l'intuition, les données d'état civil, les nomenclatures de prénoms et les connaissances déjà établies permettaient d'anticiper.

**Tableau 7.2. Top 20 des prénoms des habitants du groupe « majoritaire » selon leur « génération », en France en 2008 (enquête TeO)**

		Enquêtés du groupe « majoritaire » (N = 24 250)			Enfants d'enquête du groupe « majoritaire » (N = 7 525)		
		Prénom	Effectif non pondéré	% pondéré	Prénom	Effectif non pondéré	% pondéré
TOP 20 DES PRÉNOMS MASCULINS	1	Sébastien	37	2,7 %	Maxime	29	2,4 %
	2	Patrick	36	2,6 %	Thomas	30	2,2 %
	3	Michel	40	2,4 %	Julien	24	2,1 %
	4	Philippe	41	2,2 %	Nicolas	23	2,0 %
	5	Éric	37	2,2 %	Lucas	22	1,8 %
	6	Stéphane	33	2,1 %	Alexis	19	1,8 %
	7	Thierry	39	1,9 %	Pierre	22	1,7 %
	8	Frédéric	37	1,8 %	Clément	20	1,6 %
	9	Alain	23	1,8 %	Tom	21	1,6 %
	10	Laurent	28	1,6 %	Adrien	17	1,5 %
	11	Didier	24	1,6 %	Florian	22	1,5 %
	12	Bruno	28	1,6 %	Antoine	17	1,5 %
	13	Christophe	32	1,5 %	Romain	19	1,5 %
	14	Nicolas	29	1,5 %	Arthur	19	1,4 %
	15	Christian	19	1,5 %	Guillaume	16	1,4 %
	16	David	30	1,4 %	Quentin	17	1,4 %
	17	Dominique	21	1,3 %	Anthony	15	1,4 %
	18	Daniel	16	1,3 %	Léo	18	1,3 %
	19	Pascal	25	1,2 %	Théo	20	1,2 %
	20	Julien	27	1,2 %	Vincent	13	1,2 %
TOP 20 DES PRÉNOMS FÉMININS	1	Catherine	57	2,8 %	Manon	31	2,7 %
	2	Martine	36	2,5 %	Pauline	25	1,8 %
	3	Nathalie	50	2,3 %	Anaïs	21	1,8 %
	4	Isabelle	40	2,0 %	Marie	20	1,8 %
	5	Françoise	33	2,0 %	Océane	18	1,7 %
	6	Sandrine	36	1,7 %	Camille	22	1,5 %
	7	Sylvie	34	1,7 %	Marine	19	1,5 %
	8	Valérie	32	1,6 %	Léa	21	1,3 %
	9	Patricia	30	1,5 %	Laura	16	1,3 %
	10	Laurence	32	1,5 %	Mathilde	16	1,3 %
	11	Christine	27	1,4 %	Justine	17	1,3 %
	12	Véronique	25	1,4 %	Julie	17	1,2 %
	13	Michèle	19	1,3 %	Clara	16	1,2 %
	14	Élodie	20	1,2 %	Chloé	18	1,2 %
	15	Céline	21	1,1 %	Emma	14	1,2 %
	16	Sophie	24	1,1 %	Elodie	14	1,1 %
	17	Corinne	28	1,1 %	Marion	11	1,0 %
	18	Elisabeth	13	1,0 %	Claire	11	1,0 %
	19	Dominique	15	1,0 %	Lucie	15	1,0 %
	20	Delphine	16	0,9 %	Clémence	12	0,9 %

Lecture : le prénom masculin le plus porté par les enquêtés du groupe « majoritaire » est « Sébastien », porté par 2,7 % d'entre eux. Sont grisés les prénoms qui, d'après la nomenclature de l'INED, ne sont pas « français ».

Champ : individus du champ retenu (personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008) qui sont du groupe « majoritaire », données pondérées, N = 31 775.

Source : TeO (voir Annexe 5B, p. 183).

- **Les prénoms des habitants originaires d'Europe du Sud, par « génération »**

Si l'on examine maintenant le Top 20 des prénoms des habitants originaires d'Europe du Sud, selon qu'ils sont immigrés (G1) ou enfants d'immigrés (G2) ou petits-enfants d'immigrés (G3), là non plus aucun résultat n'apparaît particulièrement surprenant (tableau 7.3).

Beaucoup des prénoms les plus portés par les immigrés (G1) d'Europe du Sud sont des prénoms que la nomenclature de l'INED catégorise comme « latins » (en blanc dans le tableau 7.3), comme « José », « Antonio » et « Manuel » ou encore « Isabel », « Lucia » et « Paula ». La plupart de ces prénoms, comme plus généralement les prénoms « latins » et européens, sont issus du prénom d'un personnage du Nouveau (ou de l'Ancien) Testament ou du prénom d'un saint ou martyr chrétien (Annexe 6A, p. 267 et Annexe 6B, p. 285). Toutefois, tous les immigrés (G1) d'Europe du Sud ne portent pas des prénoms « latins », d'abord parce que la nomenclature de l'INED catégorise — de façon contestable — certains prénoms « latins » (comme « Maria ») comme « internationaux ou anglophones », et surtout parce que certains de ces immigrés ont francisé leur prénom à l'acquisition de la nationalité française (Annexe 6B, p. 285), comme probablement ceux qui se prénomment « Antoine », « Fernand », « Marie » ou « Isabelle ».

Parmi les enfants d'immigrés (G2) d'Europe du Sud, presque plus aucun des prénoms les plus portés n'est « latin » (en blanc dans le tableau 7.3) et la plupart de ces prénoms sont « français », comme « Daniel », « David », « Marie-Carmen » ou « Élisabeth ». Il s'agit là d'un phénomène majeur : les immigrés (G1) d'Europe du Sud n'ont que très rarement attribué à leurs fils ou à leurs filles (G2) des prénoms « latins », leur préférant des prénoms largement « français » (Annexe 6B, p. 285). Les enfants d'immigrés (G2) d'Europe du Sud (tableau 7.3) portent ainsi des prénoms semblables à ceux que portent les enfants d'enquêtés du groupe « majoritaire » (tableau 7.2). Il reste toutefois possible que les enfants d'immigrés d'Europe du Sud portent, plus souvent que les enfants d'enquêtés qui sont du groupe « majoritaire », des prénoms « qui ont fait leurs preuves » et dont le pic d'attribution est plus ou moins largement dépassé, plutôt que des prénoms à la mode qui exposeraient au risque de commettre une faute de goût.<sup>269</sup>

Parmi les petits-enfants d'immigrés (G3) d'Europe du Sud, la plupart des prénoms les plus portés sont là encore « français » (ou dans une certaine mesure « internationaux ou anglophones »), comme « Jean-Baptiste », « Samuel » et « Thomas » pour les garçons ou « Anne-Laure », « Émilie » et « Carinne » pour les filles. Là encore, les prénoms les plus portés par les descendants d'immigrés d'Europe du Sud appartiennent globalement au même stock que ceux des enfants d'enquêtés qui sont du groupe « majoritaire ». Parmi les 40 prénoms les plus courants chez les enfants d'enquêtés qui sont du groupe « majoritaire » (tableau 7.2), 15 se retrouvent chez les enfants d'enquêtés qui sont des petits-enfants d'immigrés d'Europe du Sud (tableau 7.3) : « Adrien », « Alexis », « Anthony », « Julien », « Lucas », « Maxime », « Nicolas », « Pierre » et « Thomas » pour les garçons, et « Camille », « Clara », « Julie », « Laura », « Léa » et « Marie » pour les filles.

Au total, le Top 20 des prénoms des habitants originaires d'Europe du Sud selon la « génération » (tableau 7.3) est lui aussi conforme à ce que l'intuition, les données d'état civil, les nomenclatures de prénoms et les connaissances établies permettaient d'anticiper.

---

<sup>269</sup> En métropole en 2008, parmi les descendants d'immigrés portant un prénom français, 'Male children consistently receive "older" names whereas female children receive names that are "younger"' (Coulmont 2017c, p. 11). De même, aux États-Unis, 'Immigrants who want children to fit in tend to give them American names. However, since they are not fully assimilated, they use the names of adult native-born acquaintances. In 2013 most newborn American girls named Jennifer or Ashley, generally popular a few decades ago, have Hispanic or Asian immigrant parents' (Lawson 2016, p. 188).



**Tableau 7.3. Top 20 des prénoms des habitants originaires d'Europe du Sud selon leur « génération », en France en 2008 (enquête TeO)**

		G1 Europe du Sud (N = 566)			G2 Europe du Sud (N = 1 497)			G3 Europe du Sud (N = 641)		
		Prénom	Effectif non pondéré	% pondéré	Prénom	Effectif non pondéré	% pondéré	Prénom	Effectif non pondéré	% pondéré
TOP 20 DES PRÉNOMS MASCULINS	1	José	70	10,3 %	Daniel	25	3,1 %	Jean-Baptiste	5	3,7 %
	2	Antonio	46	6,9 %	David	58	3,1 %	Samuel	7	2,7 %
	3	Manuel	41	6,2 %	Philippe	41	2,2 %	Thomas	35	2,4 %
	4	Carlos	16	2,8 %	Michel	17	2,2 %	Richard	s	2,1 %
	5	Antoine	13	2,1 %	Stéphane	39	2,2 %	Enzo	30	2,0 %
	6	Mario	11	2,1 %	Gilbert	s	2,1 %	Pierre	13	2,0 %
	7	Fernando	11	1,7 %	Frédéric	33	2,0 %	Emmanuel	8	1,9 %
	8	Alberto	10	1,7 %	Pierre	15	1,7 %	François	s	1,9 %
	9	Joaquim	11	1,6 %	Christophe	35	1,7 %	Lucas	27	1,8 %
	10	Joao	11	1,5 %	Dominique	11	1,6 %	Nicolas	30	1,7 %
	11	Jorge	8	1,4 %	Nicolas	38	1,5 %	s	s	1,6 %
	12	Luis	10	1,2 %	Bruno	29	1,5 %	Julien	23	1,6 %
	13	Francisco	7	1,0 %	Alexandre	46	1,5 %	Anthony	23	1,4 %
	14	Rafael	s	0,9 %	Sébastien	25	1,4 %	Adrien	19	1,4 %
	15	Fernand	s	0,9 %	Aimé	s	1,4 %	Kevin	20	1,4 %
	16	Miguel	6	0,9 %	Patrick	20	1,4 %	Alexandre	22	1,3 %
	17	José Manuel	s	0,8 %	Albert	s	1,3 %	Damien	17	1,3 %
	18	Juan	s	0,8 %	José	23	1,2 %	Maxime	18	1,3 %
	19	Daniel	6	0,8 %	Christian	7	1,2 %	Hugo	19	1,3 %
	20	s	s	0,8 %	Joseph	s	1,2 %	Alexis	23	1,3 %
TOP 20 DES PRÉNOMS FÉMININS	1	Maria	80	11,9 %	Marie-Carmen	s	2,5 %	Anne-Laure	s	2,6 %
	2	Marie	16	2,2 %	Élisabeth	19	2,3 %	Émilie	7	2,6 %
	3	Isabelle	12	2,0 %	Évelyne	5	2,2 %	Carinne	s	2,5 %
	4	Isabel	9	1,5 %	Sylvie	36	2,2 %	Virginie	10	2,5 %
	5	Fatima	10	1,4 %	Isabelle	39	2,1 %	Céline	9	2,3 %
	6	Rosa	7	1,2 %	Annie	s	2,1 %	Johanna	s	2,3 %
	7	Lucia	5	1,1 %	Maryse	s	2,0 %	Maeva	10	2,2 %
	8	Paula	6	1,0 %	s	s	1,9 %	Magali	s	2,2 %
	9	Marie-Hélène	7	1,0 %	Marie-Jeanne	s	1,8 %	Séverine	s	1,9 %
	10	Deolinda	6	0,9 %	Sandrine	32	1,6 %	Charlotte	s	1,6 %
	11	s	s	0,9 %	Nathalie	35	1,6 %	Maude	s	1,4 %
	12	Gloria	5	0,9 %	Stéphanie	30	1,5 %	Laura	27	1,4 %
	13	Marie-José	6	0,9 %	Christine	24	1,5 %	Sophie	s	1,4 %
	14	Natalia	5	0,9 %	Céline	26	1,3 %	Léa	22	1,3 %
	15	Maria-Isabel	6	0,8 %	Sandra	26	1,3 %	Caroline	5	1,3 %
	16	Joséphine	6	0,8 %	Patricia	21	1,1 %	Julie	21	1,3 %
	17	Carmen	6	0,8 %	Mélanie	25	1,1 %	Marie	17	1,2 %
	18	Adélaïde	5	0,8 %	Sophie	16	1,0 %	Camille	19	1,2 %
	19	Ana	7	0,8 %	Françoise	12	1,0 %	s	s	1,1 %
	20	s	s	0,8 %	Béatrice	17	0,9 %	Clara	20	1,0 %

Lecture : le prénom masculin le plus porté par les immigrés (G1) d'Europe du Sud est « José », porté par 10,3 % d'entre eux. Sont grisés les prénoms qui, d'après la nomenclature de l'INED, ne sont pas « latins ».

Champ : individus du champ retenu (personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008) qui sont immigrés ou descendants d'immigrés d'Europe du Sud, données pondérées, N = 2 704.

Source : TeO (voir Annexe 5B, p. 183).

NB : Conformément aux règles du secret statistique du [CASD](#) et de l'[INSEE](#), et pour qu'aucune information ne permette d'identifier directement ou indirectement une personne physique (ni d'inférer de nouvelles informations sur elle), les résultats obtenus sur des effectifs faibles (<5) sont notés « s » (comme « secret »).

- **Les prénoms des habitants originaires du Maghreb, par « génération »**

Si l'on examine maintenant le Top 20 des prénoms des habitants originaires du Maghreb, les résultats concernant les prénoms des immigrés (G1) n'apparaissent guère surprenants (tableau 7.4). Ceux des enfants d'immigrés (G2) le sont légèrement plus, puisqu'ils comprennent le prénom « Daniel ». Mais les résultats les plus surprenants concernent les petits-enfants d'immigrés (G3) du Maghreb.

Tout d'abord, en métropole en 2008 « Nicolas » ferait bien partie des prénoms les plus portés par les petits-fils d'immigrés du Maghreb, comme indiqué dans le tableau de Coulmont et Simon (2019, p. 3). Sur l'échantillon TeO retenu pondéré (représentatif des personnes de 18-50 / 18-60 ans en ménage ordinaire et de leurs enfants de 0-17 ans habitant en France métropolitaine en 2008), « Nicolas » ne serait certes pas le 2<sup>e</sup> prénom le plus fréquent chez les petits-fils d'immigrés du Maghreb, mais le 4<sup>e</sup>, ce qui reste étonnant. Mais plus généralement, d'après l'échantillon TeO retenu pondéré, parmi les 40 prénoms les plus courants chez les petits-enfants d'immigrés (G3) du Maghreb, on trouverait notamment : chez les garçons, non seulement « Nicolas », mais aussi « Jérémy » et « Benoît » ; et chez les filles, « Virginie ». On peut s'étonner que Coulmont, Simon et l'INED aient publié un tableau de résultats dans lequel les prénoms des petits-enfants d'immigrés du Maghreb présentés comme les plus fréquents ne reposent que sur moins d'une vingtaine d'observations, mais des aléas d'échantillonnage ne sauraient expliquer la présence de prénoms aussi contre-intuitifs sur un effectif pondéré de 335 individus (tableau 7.4), qui repose sur un effectif non pondéré de 1 490 individus (tableau 5.3).

En outre, d'après l'enquête TeO, ce sont *plus de 3 % des petits-fils d'immigrés du Maghreb* qui se prénommeraient « Nicolas » (tableau 7.4). Ainsi, parmi les petits-fils d'immigrés du Maghreb, le prénom « Nicolas » serait beaucoup plus fréquent (3,5 %) que ne l'est le prénom « Mohamed » (0,8 %). Comment un prénom comme « Nicolas » pourrait-il faire partie des prénoms les plus fréquents chez les petits-fils d'immigrés du Maghreb, et même être porté par plus de 3 % d'entre eux, sans avoir *jamais* été rencontré par *aucun* des dizaines de proches (y compris plusieurs descendants d'immigrés du Maghreb) à qui j'ai mentionné ce résultat depuis avril 2019 ?

Au total, le Top 20 des prénoms des petits-enfants d'immigrés du Maghreb (tableau 7.4) apparaît si contre-intuitif et si peu conforme aux connaissances établies sur les prénoms des descendants d'immigrés du Maghreb (Annexe 5A, p. 147, Annexe 6A, p. 267 et Annexe 6B, p. 285) qu'il n'est pas interdit de faire un pas de côté et de se demander : qu'est-ce qui pourrait produire de tels résultats ? Ne serait-il pas possible que l'enquête TeO, malgré la bonne qualité de son échantillonnage et de sa variable « prénom », contienne des erreurs ? Quand un chercheur « tombe » sur de tels résultats, il lui revient d'exercer son métier et de chercher à comprendre ce qu'il se passe.

**Tableau 7.4. Top 20 des prénoms des habitants originaires du Maghreb selon leur « génération », en France en 2008 (enquête TeO)**

		G1 Maghreb (N = 1 043)			G2 Maghreb (N = 1 777)			G3 Maghreb (N = 335)		
		Prénom	Effectif non pondéré	% pondéré	Prénom	Effectif non pondéré	% pondéré	Prénom	Effectif non pondéré	% pondéré
TOP 20 DES PRÉNOMS MASCULINS	1	Mohamed	103	9,5 %	Mohamed	105	3,6 %	Karim	s	4,5 %
	2	Ahmed	30	3,2 %	Karim	55	2,2 %	Nassim	9	3,6 %
	3	Rachid	20	2,1 %	Farid	38	2,0 %	Farès	8	3,5 %
	4	Mohammed	16	1,8 %	Mehdi	48	1,6 %	Nicolas	10	3,5 %
	5	Abdelkader	18	1,5 %	Rachid	27	1,5 %	Jérémy	s	2,7 %
	6	Abdellah	12	1,4 %	Samir	32	1,3 %	Mourad	s	2,5 %
	7	Khalid	10	1,3 %	Kamel	21	1,2 %	Yanis	19	1,9 %
	8	Mourad	13	1,3 %	Adam	35	1,1 %	Mehdi	16	1,7 %
	9	Mustapha	16	1,3 %	Daniel	s	1,1 %	Kevin	9	1,2 %
	10	Ali	16	1,3 %	Sofiane	36	1,0 %	Yacine	8	1,0 %
	11	Brahim	9	1,2 %	Yanis	37	1,0 %	Ilyès	6	1,0 %
	12	Hamid	9	1,2 %	Yassine	31	0,9 %	Rayan	7	0,9 %
	13	Hicham	10	1,1 %	Saïd	20	0,9 %	Benoît	s	0,9 %
	14	Saïd	13	1,1 %	Ahmed	21	0,9 %	Driss	6	0,8 %
	15	Omar	10	1,0 %	Amine	36	0,9 %	Mohamed	9	0,8 %
	16	Walid	s	1,0 %	Youssef	25	0,9 %	Rayane	6	0,8 %
	17	Nabil	7	1,0 %	Yacine	20	0,7 %	Sofian	6	0,7 %
	18	Driss	12	1,0 %	Nordine	17	0,7 %	Adam	7	0,7 %
	19	Hassan	8	0,9 %	Ali	15	0,6 %	s	s	0,7 %
	20	Karim	6	0,9 %	Samy	22	0,6 %	s	s	0,7 %
TOP 20 DES PRÉNOMS FÉMININS	1	Fatima	70	6,0 %	Sarah	100	3,2 %	Sarah	25	5,0 %
	2	Khadija	27	3,1 %	Nadia	73	2,8 %	Sophia	s	3,0 %
	3	Samira	16	2,5 %	Myriam	74	2,8 %	Sabrina	s	2,6 %
	4	Fatiha	28	2,5 %	Fatima	61	1,9 %	Lina	17	2,4 %
	5	Aïcha	23	1,9 %	Yamina	15	1,7 %	Inès	17	2,3 %
	6	Malika	17	1,6 %	Inès	61	1,6 %	Sofia	9	1,7 %
	7	Naima	18	1,5 %	Malika	31	1,5 %	Celia	11	1,5 %
	8	Nadia	14	1,5 %	Sonia	39	1,4 %	Jessica	7	1,1 %
	9	Zohra	17	1,5 %	Sabrina	37	1,3 %	Leïla	6	1,1 %
	10	Karima	17	1,4 %	Nora	28	1,2 %	Myriam	10	1,0 %
	11	Samia	18	1,4 %	Samira	25	1,2 %	Anaïs	9	1,0 %
	12	Houria	13	1,2 %	Leïla	31	1,2 %	Virginie	s	0,9 %
	13	Rachida	14	1,2 %	Houria	5	1,2 %	Melissa	8	0,9 %
	14	Meriem	7	1,1 %	Anissa	33	1,0 %	Lilia	5	0,9 %
	15	Saïda	13	1,1 %	Samia	28	1,0 %	Aïcha	s	0,9 %
	16	Fatma	13	1,0 %	Yasmina	25	0,9 %	Lisa	7	0,8 %
	17	Bouchra	8	1,0 %	Aïcha	22	0,8 %	Emma	8	0,8 %
	18	Yamina	12	1,0 %	Lina	29	0,8 %	s	s	0,8 %
	19	Jamila	8	0,9 %	Yasmine	28	0,6 %	Lola	5	0,8 %
	20	Najat	10	0,9 %	Karima	20	0,6 %	s	s	0,8 %

Lecture : le prénom masculin le plus porté par les immigrés (G1) du Maghreb est « Mohamed », porté par 9,5 % d'entre eux. Sont grisés les prénoms qui, d'après la nomenclature de l'INED, ne sont pas « arabo-musulmans », comme « Nicolas ».

Champ : individus du champ retenu (personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008) qui sont immigrés ou descendants d'immigrés du Maghreb, données pondérées, N = 3 155.

Source : TeO (voir Annexe 5B, p. 183).

NB : Conformément aux règles du secret statistique du [CASD](#) et de l'[INSEE](#), et pour qu'aucune information ne permette d'identifier directement ou indirectement une personne physique (ni d'inférer de nouvelles informations sur elle), les résultats obtenus sur des effectifs faibles (<5) sont notés « s » (comme « secret »).

NB : une version corrigée de ce tableau 7.4 figure au tableau 7.15.

❖ **La reproduction « au plus près » du tableau page 3**

Pour y voir plus clair, je reproduis le tableau de Coulmont et Simon (2019, p. 3) sur 3 échantillons différents (échantillon retenu, échantillon maximal et échantillon composé des enquêtés et de tous leurs enfants vivants), et selon 2 variantes de pondération (avec ou sans), soit au total  $3 \times 2 = 6$  combinaisons d'hypothèses envisageables (tableaux 7.5a à 7.5f). Pour rappel, le tableau à reproduire est le suivant :

PUBLIÉ				
Tableau 7.5. « Top 3 des prénoms par sexe et générations, Europe du Sud ou Maghreb », tel que publié par Coulmont et Simon 2019, p. 3 : champ non indiqué, pondération non indiquée, N = non indiqué				
	Europe du Sud		Maghreb	
	Homme	Femme	Homme	Femme
<b>Immigrés</b>	José Antonio Manuel	Maria Marie Ana	Mohamed Ahmed Rachid	Fatima Fatiha Khadija
<b>Descendants d'immigrés</b> <sup>270</sup>	Jean David Alexandre	Marie Sandrine Sandra	Mohamed Karim Mehdi	Sarah Nadia Myriam
<b>Petits-enfants d'immigrés</b>	Thomas Lucas Enzo	Laura Léa Camille	Yanis Nicolas Mehdi	Sarah Inès Lina

<sup>270</sup> Coulmont et Simon (2019) appellent ici « descendants d'immigrés » les seuls *enfants* d'immigrés. Pour éviter toute confusion, je renomme ci-dessous cette catégorie « enfants d'immigrés » (tableaux 7.5a à 7.5f). Plus généralement, le présent document distingue d'une part les « enfants d'immigrés » (G2), d'autre part les « petits-enfants d'immigrés » (G3), et enfin les « descendants d'immigrés » (G2 et G3 réunis).

### Reproduction « au plus près » du tableau publié page 3

CHAMP RETENU, DONNÉES PONDÉRÉES				
Tableau 7.5a. « Top 3 des prénoms par sexe et générations, Europe du Sud ou Maghreb », N = 42 208				
	Europe du Sud		Maghreb	
	Homme	Femme	Homme	Femme
Immigrés	José	Maria	Mohamed	Fatima
	Antonio	Marie	Ahmed	Khadija
	Manuel	Isabelle	Rachid	Samira
Enfants d'immigrés	Daniel	Marie-Carmen	Mohamed	Sarah
	David	Elisabeth	Karim	Nadia
	Philippe	Evelyne	Farid	Myriam
Petits-enfants d'immigrés	Jean-Baptiste	Anne-Laure	Karim	Sarah
	Samuel	Emilie	Nassim	Sophia
	Thomas	Carinne	Farès	Sabrina
<div>Résultat reproduit = résultat publié</div> <div>Résultat reproduit ≠ résultat publié</div>				

CHAMP RETENU, DONNÉES NON PONDÉRÉES				
Tableau 7.5b. « Top 3 des prénoms par sexe et générations, Europe du Sud ou Maghreb », N = 42 208				
	Europe du Sud		Maghreb	
	Homme	Femme	Homme	Femme
Immigrés	José	Maria	Mohamed	Fatima
	Antonio	Marie	Ahmed	Fatiha
	Manuel	Isabelle	Rachid	Khadija
Enfants d'immigrés	David	Isabelle	Mohamed	Sarah
	Alexandre	Sylvie	Karim	Myriam
	Philippe	Nathalie	Mehdi	Nadia
Petits-enfants d'immigrés	Thomas	Laura	Yanis	Sarah
	Enzo	Léa	Mehdi	Inès
	Nicolas	Julie	Nicolas	Lina
<div>Résultat reproduit = résultat publié</div> <div>Résultat reproduit ≠ résultat publié</div>				

NB : je colore en vert les prénoms que chaque reproduction permet de retrouver à l'identique du tableau publié ; je colore en rouge les prénoms qui figurent dans une reproduction alors qu'ils ne figurent pas dans le tableau publié ; et je conserve sans couleur les prénoms que chaque reproduction permet de retrouver mais pas au même rang que dans le tableau publié.

NB : ce champ est représentatif des personnes de 18-50 / 18-60 ans en ménage ordinaire et de leurs enfants de 0-17 ans, habitant en France métropolitaine en 2008.

### Reproduction « au plus près » du tableau publié page 3

CHAMP MAXIMAL, DONNÉES PONDÉRÉES				
Tableau 7.5c. « Top 3 des prénoms par sexe et générations, Europe du Sud ou Maghreb », N = 83 045				
	Europe du Sud		Maghreb	
	Homme	Femme	Homme	Femme
Immigrés	José	Maria	Mohamed	Fatima
	Manuel	Maria Inès	Ahmed	Khadija
	Antonio	Carmelle	Rachid	Aicha
Enfants d'immigrés	Pierre	Isabelle	Mohamed	Nadia
	François	Christine	Karim	Sarah
	Philippe	Elisabeth	Mehdi	Fatima
Petits-enfants d'immigrés	Sébastien	Elodie	Yanis	Inès
	Maxime	Mélanie	Nicolas	Sarah
	Thomas	Virginie	Karim	Audrey
<div>Résultat reproduit = résultat publié</div> <div>Résultat reproduit ≠ résultat publié</div>				

CHAMP MAXIMAL, DONNÉES NON PONDÉRÉES				
Tableau 7.5d. « Top 3 des prénoms par sexe et générations, Europe du Sud ou Maghreb », N = 83 045				
	Europe du Sud		Maghreb	
	Homme	Femme	Homme	Femme
Immigrés	José	Maria	Mohamed	Fatima
	Manuel	Marie	Ahmed	Aicha
	Antonio	Rosa	Ali	Malika
Enfants d'immigrés	David	Sandrine	Mohamed	Nadia
	Alexandre	Nathalie	Karim	Sarah
	Philippe	Sandra	Mehdi	Fatima
Petits-enfants d'immigrés	Thomas	Laura	Yanis	Sarah
	Nicolas	Léa	Mehdi	Inès
	Lucas	Julie	Mohamed	Lina
<div>Résultat reproduit = résultat publié</div> <div>Résultat reproduit ≠ résultat publié</div>				

NB : je colore en vert les prénoms que chaque reproduction permet de retrouver à l'identique du tableau publié ; je colore en rouge les prénoms qui figurent dans une reproduction alors qu'ils ne figurent pas dans le tableau publié ; et je conserve sans couleur les prénoms que chaque reproduction permet de retrouver mais pas au même rang que dans le tableau publié.

NB : ce champ n'est présenté qu'à titre indicatif, il n'est pas représentatif des personnes de 18-50 / 18-60 ans en ménage ordinaire et de leurs enfants de 0-17 ans, habitant en France métropolitaine en 2008.

### Reproduction « au plus près » du tableau publié page 3

CHAMP ENQUÊTÉS ET TOUS LEURS ENFANTS VIVANTS, DONNÉES PONDÉRÉES				
Tableau 7.5e. « Top 3 des prénoms par sexe et générations, Europe du Sud ou Maghreb », N = 52 903				
	Europe du Sud		Maghreb	
	Homme	Femme	Homme	Femme
Immigrés	José	Maria	Mohamed	Fatima
	Antonio	Marie	Ahmed	Khadija
	Manuel	Isabelle	Rachid	Samira
Enfants d'immigrés	David	Elisabeth	Mohamed	Sarah
	Daniel	Sylvie	Karim	Nadia
	Sébastien	Marie-Carmen	Samir	Fatima
Petits-enfants d'immigrés	Sébastien	Elodie	Nicolas	Inès
	Maxime	Mélanie	Yanis	Sarah
	François	Virginie	Karim	Noelia
<div>Résultat reproduit = résultat publié</div> <div>Résultat reproduit ≠ résultat publié</div>				

CHAMP ENQUÊTÉS ET TOUS LEURS ENFANTS VIVANTS, DONNÉES NON PONDÉRÉES				
Tableau 7.5f. « Top 3 des prénoms par sexe et générations, Europe du Sud ou Maghreb », N = 52 903				
	Europe du Sud		Maghreb	
	Homme	Femme	Homme	Femme
Immigrés	José	Maria	Mohamed	Fatima
	Antonio	Marie	Ahmed	Fatiha
	Manuel	Isabelle	Rachid	Khadija
Enfants d'immigrés	David	Sandrine	Mohamed	Sarah
	Philippe	Sandra	Karim	Nadia
	Alexandre	Isabelle	Mehdi	Fatima
Petits-enfants d'immigrés	Thomas	Laura	Yanis	Sarah
	Nicolas	Léa	Mehdi	Inès
	Enzo	Julie	Nicolas	Lina
<div>Résultat reproduit = résultat publié</div> <div>Résultat reproduit ≠ résultat publié</div>				

NB : je colore en vert les prénoms que chaque reproduction permet de retrouver à l'identique du tableau publié ; je colore en rouge les prénoms qui figurent dans une reproduction alors qu'ils ne figurent pas dans le tableau publié ; et je conserve sans couleur les prénoms que chaque reproduction permet de retrouver mais pas au même rang que dans le tableau publié.

NB : ce champ n'est présenté qu'à titre indicatif, il n'est pas représentatif des personnes de 18-50 / 18-60 ans en ménage ordinaire et de leurs enfants de 0-17 ans, habitant en France métropolitaine en 2008.

De ces résultats (tableaux 7.5a à 7.5f), je tire plusieurs enseignements :

- tout d'abord, aucune de ces reproductions « au plus près » ne permet de reproduire les résultats publiés par Coulmont et Simon (2019) à l'identique : comme leur article n'indique pas le champ sur lequel ils ont calculé leurs résultats, et comme les programmes informatiques qu'ils m'ont finalement communiqués sont incomplets (Annexe 5A, p. 147), je continue d'ignorer le champ sur lequel leur tableau est calculé ;
- ensuite, quel que soit le champ considéré, les résultats les plus proches de ceux des auteurs sont obtenus sur données non pondérées ; j'en conclus que leur tableau pourrait être issu de calculs non pondérés — un choix qu'on a peine à envisager<sup>271</sup> et que leur article ne mentionne pas et ne justifie pas ;
- en outre, la reproduction qui produit les résultats les plus proches de ceux des auteurs est réalisée non pas sur l'échantillon ici retenu (représentatif des personnes de 18-50 / 18-60 ans en ménage ordinaire et de leurs enfants de 0-17 ans habitant en France métropolitaine en 2008), mais sur l'échantillon des enquêtés et de *tous* leurs enfants vivants (tableau 7.5f) ; j'en conclus que le tableau de l'article pourrait être issu de calculs réalisés sur un champ proche de celui-là — qui inclut des doubles comptes et déforme donc les prénoms des petits-enfants d'immigrés du Maghreb en France métropolitaine en 2008 ;
- mais surtout, le prénom « Nicolas » apparaît fréquemment dans le Top 3 des prénoms des petits-fils d'immigrés du Maghreb (tableaux 7.5b, 7.5c, 7.5e et 7.5f), à tel point que dans la reproduction sur l'échantillon pondéré composé des enquêtés et de tous leurs enfants vivants, le prénom « Nicolas » se retrouve comme le prénom *le plus fréquemment* porté en France en 2008 par les petits-fils d'immigrés du Maghreb (tableau 7.5e) !

❖ **À la recherche de petits-fils d'immigrés du Maghreb prénommés « Nicolas » : auprès des experts des migrations et chez les couples « mixtes »**

Si, parmi les personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008, plus de 3 % des petits-fils d'immigrés du Maghreb se prénommaient « Nicolas » (tableau 7.4), on peut imaginer que les experts des migrations en connaîtraient un certain nombre. Pour savoir ce qu'il en est, j'ai personnellement contacté 36 chercheurs (historiens, anthropologues, sociologues, démographes, géographes, politistes, statisticiens, linguistes et juristes) qui sont spécialistes des immigrés et/ou de leurs descendants en France, et je leur ai demandé s'ils avaient personnellement connaissance de petits-fils d'immigrés du Maghreb prénommés « Nicolas ». Parmi ces 36 chercheurs, 25 m'ont répondu et, parmi eux, *une seule* m'a indiqué connaître *un seul* descendant d'immigrés du Maghreb prénommé « Nicolas », mais il s'agit d'un *fils* (né au début des années 1970) et non pas d'un *petit-fils* d'immigrés du Maghreb. Sur les 25 chercheurs qui m'ont répondu, aucun ne m'a donc indiqué connaître un seul petit-fils d'immigrés du Maghreb prénommé « Nicolas ».

Si en 2008 plus de 3 % des petits-fils d'immigrés du Maghreb se prénommaient « Nicolas », on peut envisager que ce prénom aurait été attribué — au moins en partie — au sein de couples « mixtes », composés d'un descendant d'immigrés du Maghreb et d'un membre du groupe « majoritaire », plutôt qu'au sein de couples formés de deux personnes originaires du Maghreb.<sup>272</sup> Et si tant de petits-fils d'immigrés du Maghreb se prénommaient effectivement

<sup>271</sup> Les statistiques descriptives doivent être calculées sur échantillon pondéré pour tenir compte, notamment, (i) du suréchantillonnage des descendants d'immigrés, qui exige de les sous-pondérer, et (ii) des variations de taux de non-réponse d'une population à l'autre.

<sup>272</sup> En effet, même dans les unions entre un(e) descendant(e) d'immigrés d'Algérie et un(e) descendant(e) de harkis, « ne se posent pas [...] les enjeux liés au choix du prénom des enfants » (Fabbiano 2009, p. 104).



« Nicolas », on peut imaginer que, depuis les années 1980, les recherches menées par entretiens auprès des couples « mixtes » auraient rapporté le phénomène. Or, ces recherches à propos des prénoms que les couples « mixtes » attribuent à leurs enfants ont observé des couples qui ont donné à leurs enfants des prénoms comme :

- « Karim », « Nassim », « Yanis » (ou « Yannis »), « Mehdi » (ou « Medhi ») « Kévin », et « Adam », <sup>273</sup> qui d'après l'enquête TeO sont respectivement les 1<sup>er</sup>, 2<sup>e</sup>, 7<sup>e</sup>, 8<sup>e</sup>, 9<sup>e</sup> et 18<sup>e</sup> prénoms les plus portés par les petits-fils d'immigrés du Maghreb en 2008 (tableau 7.4), mais aussi « Anis », « Camille » (dérivé de « Kamil »), « Elias », <sup>274</sup> « Enzo », « Gabriel », « Hedi », « Ismaël », « Sami », « Selim », « Tarik » <sup>275</sup> et « Théo » <sup>276</sup> ;
- « Sarah » (ou « Sara »), « Sophia », « Sabrina », « Lina », « Inès », « Leïla » et « Myriam », qui d'après l'enquête TeO sont respectivement les 1<sup>er</sup>, 2<sup>e</sup>, 3<sup>e</sup>, 4<sup>e</sup>, 5<sup>e</sup>, 9<sup>e</sup> et 10<sup>e</sup> prénoms les plus portés par les petites-filles d'immigrés du Maghreb en 2008 (tableau 7.4), mais aussi « Athénéa », « Camille », « Ève », « Floria », « Julia », « Kenza », <sup>277</sup> « Léa », « Lili », « Linda », « Luna », « Lydia », « Lydie », « Nadia », « Nora », « Rachel », « Sonia », « Victoria », « Yasmin » et « Zoé », ces prénoms « mixtes », « ambigus » ou « passe-partout » étant sensiblement plus nombreux parmi les prénoms féminins que masculins <sup>278</sup> ;
- mais aucune trace de « Nicolas », « Jérémy », « Benoît » ou « Virginie » (tableau 7.4).

Plus généralement, selon ces recherches par entretiens, il apparaît que les couples mixtes composés d'un descendant d'immigrés du Maghreb et d'un natif sans ascendant immigré

<sup>273</sup> Chez 24 couples mixtes interrogés en France en 2005-2007 et formés d'un conjoint originaire du monde arabo-musulman et d'un autre de la population native sans ascendant immigré, « Les prénoms qui connaissent des équivalences dans les trois religions monothéistes sont très prisés, notamment ceux de l'Ancien Testament (Adam, Gabriel pour les garçons ou Eve, Sara ou Myriam pour les filles). Ces derniers sont toutefois plutôt choisis dans leurs transcriptions adoptées par la langue majoritaire ; par exemple Myriam et non pas Meriem ou Gabriel et non pas Djibril » (Collet 2019, p. 161).

<sup>274</sup> Pour des négociations conjugales en France entre un père arabe et une mère française consistant à savoir si le prénom « Elias » (dérivé du prénom du prophète hébreu « Elie ») est marocain, tunisien ou juif, voir Auger et Moïse 2007, p. 3-4. Plus généralement, sur les discussions de couples mixtes franco-arabes relatives au choix du prénom des enfants, voir Auger et Moïse 2000.

<sup>275</sup> Sur le changement de prénom d'un Libanais devenu français (de « Tarek » à « Julien Tarek ») pour éviter d'éventuelles discriminations sur le marché du travail, voir Odasso 2016, p. 74-75.

<sup>276</sup> Parmi 11 couples de la fin des années 1990 formés d'un enfant d'immigrés d'Algérie et d'un enfant de la population native sans ascendant immigré, Leïla indique, sur le choix du prénom de son fils : « je me suis dit que en France pour l'intégration, il valait mieux avoir un nom français ou un nom (mixte)... on a hésité, parce qu'on s'est dit "s'il a une tête d'arabe et qu'il a un prénom français, ça va mal passer, et l'inverse aussi..." donc moi j'avais déjà accroché sur un prénom, c'est Riwan : c'est un prénom berbère et breton en même temps ! j'avais bien accroché sur ce prénom-là, et puis une copine m'a parlé de Théo... donc j'ai dit "s'il est blanc on va l'appeler Théo, et s'il est bis on va l'appeler Riwan !" et en fait à la maternité, ils te demandent tout de suite le nom, donc on a dit Théo, heureusement il était blanc (rire) ! et ça lui va très bien ! » (Laffort 2003, p. 131).

<sup>277</sup> Parmi 11 couples de la fin des années 1990 formés d'un enfant d'immigrés d'Algérie et d'un enfant de la population native sans ascendant immigré, Fatima indique : « moi je connais des amis, des couples mixtes, ils ont essayé de trouver un prénom original, entre les deux : Kenza ! il y a eu beaucoup de Kenza ! et puis à la dernière minute, ils ont choisi Léa ! » (Laffort 2003, p. 253).

<sup>278</sup> Parmi 20 couples franco-maghrébins en région parisienne en 2005-2006, Fouad indique : « Ce qui m'a le plus aidé dans le choix du prénom c'était l'intérêt des enfants. On a rapidement trouvé Nadia qui était un petit peu les deux. Par contre pour Vincent c'était moins évident parce que tous les prénoms étaient marqués et je voulais qu'il soit à l'aise dans la cour quand on l'appelle. Son prénom ne doit pas lui rappeler qu'il est... qu'il n'ait aucun frein au niveau de sa scolarité ou au niveau de son intégration » (Puzenat 2008, p. 121). Parmi 5 couples mixtes franco-arabes en France dans les années 2000, on observe les prénoms masculins (Elias\*2, Camille dérivé de Kamil) et féminins (Julia, Victoria, Flora, Rachel, Camille, Zoé, Leïla, Yasmin) (Auger et Moïse 2007).

tendraient à attribuer non pas des prénoms « français » non ambigus comme « Nicolas »,<sup>279</sup> mais plutôt certains prénoms « arabo-musulmans »<sup>280</sup> ou, peut-être plus encore, des prénoms « passe-partout »,<sup>281</sup> c'est-à-dire des prénoms qui pourraient aussi bien passer pour « arabo-musulmans » que pour « anglophones » ou « latins » voire « français ».

De cette revue de la littérature par entretiens sur les prénoms attribués aux enfants des couples « mixtes », je conclus qu'aucune recherche sur le sujet n'a jamais mis en évidence l'attribution — ni a fortiori l'attribution massive — du prénom « Nicolas ». Bien au contraire, ce prénom « français » non ambigu ne ressemble pas au type de prénom typiquement choisi par ces couples « mixtes ». Et de fait, si « La prénomination des enfants dans les couples mixtes relève de logiques particulières oscillant entre souci d'intégration et transmission identitaire »

<sup>279</sup> Seule Streiff-Fénart (1990) observe que parmi les prénoms attribués à 576 enfants de couples mixtes franco-maghrébins nés à Marseille en 1984-1985, 58 % ont un prénom « français » et 19 % un prénom « arabe » (et 16 % un prénom « ambigu », et 7 % un prénom « autre ») (Streiff-Fénart 1990, p. 10). En outre, parmi 11 couples formés d'un enfant d'immigrés d'Algérie et d'un enfant de la population « majoritaire » à la fin des années 1990, Saïd indique : « tu as remarqué que mes enfants s'appelaient Laurent et Sabrina, qui ne sont pas du tout des prénoms... enfin Sabrina si, c'est un nom arabe... alors Laurent pourquoi ? c'est un peu le sort qui en a décidé ainsi... quelques jours avant notre mariage, on a eu un accident de voiture avec mon beau-père qui est mort dans l'accident... et j'aimais mon beau-père énormément, c'est quelqu'un pour qui j'avais énormément de respect et d'affection... et donc j'avais dit — qu'on ait un garçon ou une fille — qu'on l'appellerait Laurent ou Laurence et il a un deuxième prénom, Aïssa, pour équilibrer... [...] [Quant à Sabrina,] on voulait un prénom de compromis... moi je ne voulais pas un prénom arabe, Séverine non plus... je ne voulais pas un prénom bien français et Séverine non plus... donc on a coupé la poire en deux, et ça a satisfait tout le monde... [...] je n'aurais pas appelé mon fils Laurent si la circonstance ne s'était pas posée... » (Laffort 2003, p. 214). En France, certains Berbères ou descendants de Berbères convertis au christianisme, qui rejettent l'islam pour protester contre les politiques d'arabisation, ont toutefois pu attribuer à leurs enfants des prénoms chrétiens, y compris « français ».

<sup>280</sup> « La tendance à attribuer un prénom pris à la culture paternelle est encore plus forte quand un enfant sans religion a un père d'origine islamique ; un prénom « musulman » est alors parfois le seul, ou le principal, élément identitaire « visible » qui rattache l'enfant à la culture d'origine de son père » (Varro 1990, p. 309).

<sup>281</sup> Parmi 17 couples binationaux habitant en France et en RDA en 1986-1988, environ la moitié de leurs 29 enfants ont des prénoms « passe-partout », pouvant provenir de la culture du père comme de celle de la mère, environ 1/6<sup>e</sup> ont un prénom marqué du côté maternel et environ 1/3 ont un prénom marqué du côté paternel (Varro 1990, p. 309). Streiff-Fénart (1990) catégorise comme « passe-partout » ou ambigu (plutôt que « français » ou « arabo-berbères ») des prénoms « bibliques comme Sarah ou Myriam et de nombreux prénoms féminins, qui sans être des prénoms musulmans classiques font partie du stock actuel des prénoms au Maghreb (comme Sabrina ou Ines), ou d'autres encore dont l'origine arabe est dissimulée en France derrière une autre origine étrangère plus prestigieuse (slave, anglo-américaine ou parfois italienne) : ainsi en est-il de Nadia, Sonia, Linda (diminutif de Dalinda), Sophia. Nous avons aussi classé dans cette rubrique certains prénoms masculins qui permettent de jouer sur l'ambiguïté d'une identité phonétique entre des prénoms arabes usuels et des diminutifs américanisés : c'est le cas notamment de Hedi (Eddy) et de Sami (Sammy) » (Streiff-Fénart 1990, p. 9). Pour une analyse du sens du prénom « Jean-Karim » pour un porteur en France en 1992, voir Varro 1994. Chez les couples franco-maghrébins dans les années 1980 et 1990, la recherche de compromis et de prénoms « passe-partout » « explique la mode de certains prénoms mixtes comme Linda, Léa, Lina ou Inès pour les filles, Théo, Enzo, Kévin ou encore Yanis pour les garçons. [...] Et citant Augustin Barbara, il ajoute :] Tarik, Karim, Nassim, Ismaël, s'ils marquent une appartenance orientale, ne disqualifient pas autant que des prénoms comme Mohammed, Tahar, automatiquement écartés à cause de leur connotation fortement islamique » (Laffort 2003, p. 32). D'après la synthèse récente de Collet (2019), il arrive que « Les deux conjoints choisissent délibérément des prénoms difficilement identifiables culturellement. Ces derniers seront alors originaux dans le vrai sens du terme, ils peuvent provenir d'un stock de prénoms arabes anciens, voire oubliés, ou sinon, ils seront inventés ou réinventés, jouant sur des sonorités orientales ou méditerranéennes. Kenza est ainsi redevenue à la mode, mais on rencontre aussi Inès, Lydie, Lili, Lydia, Luna, Athénéa. Pour les garçons, Medhi ou Sami ont eu leur heure de gloire dans les années 1990, alors que dans les années 2000, c'est le prénom masculin Yanis ou Yannis qui connaît un franc succès. Plus récemment, ce sont des prénoms comme Ilian ou Ilies qui apparaissent. Plutôt d'origine grecque, ces derniers sont pris pour des prénoms maghrébins ou, comme disent les enquêtés, « il passe bien », au même titre que Selim ou Anis. D'autres couples encore vont inventer des prénoms ou adopter une orthographe d'une autre tradition culturelle comme Gil ou Alan » (Collet 2019, p. 164). L'intérêt de prénoms « passe-partout » n'a rien de spécifique aux personnes originaires du Maghreb en France, et se retrouve par exemple chez des immigrés ou minorités linguistiques comme la minorité russophone en Finlande (Eskola et Hämäläinen 2019).

(Collet 2019, p. 166), quelle identité maghrébine ou arabe ou berbère ou musulmane serait transmise par le choix d'un prénom « français » non ambigu comme « Nicolas » ?

❖ **À la recherche de petits-fils d'immigrés du Maghreb prénommés « Nicolas » : dans l'annuaire et sur les réseaux sociaux**

Sur l'échantillon retenu pondéré (représentatif des personnes de 18-50 / 18-60 ans en ménage ordinaire et de leurs enfants de 0-17 ans habitant en France métropolitaine en 2008), l'enquête TeO compte 10 « Nicolas » originaires du Maghreb (3 G2 et 6 G3, nés en moyenne en 1983), contre 17 « Nicolas » originaires d'Europe du Sud (12 G2 et 5 G3, nés en moyenne en 1987). Dans ce contexte, et vu la bonne qualité de l'échantillonnage TeO et de sa variable « prénom » (tableau 5.5, figures 5.6 à 5.10, et figures 6.10 et 6.11), y compris pour le prénom « Nicolas » lui-même (figures 7.1a et 7.1b), on devrait pouvoir trouver dans l'annuaire et sur les réseaux sociaux environ 1,5 à 2 fois moins de « Nicolas » originaires du Maghreb que d'Europe du Sud. Pour estimer si c'est bien le cas, je compare le nombre de « Nicolas » qui portent le nom de famille le plus fréquent de chacun des 3 pays du Maghreb (Alaoui au Maroc, Saidi en Algérie et Trabelsi en Tunisie) avec le nombre de « Nicolas » qui portent le nom de famille le plus fréquent de chacun des 3 pays d'Europe du Sud (Rossi en Italie, Garcia Espagne et Santos au Portugal) (Taggart 2020 renseigne les noms de famille les plus fréquents par pays, à partir de [sources](#) de recensements et généalogiques). Cette méthode a un inconvénient : elle n'est valide que si, notamment, la proportion d'immigrés et de descendants d'immigrés du Maghreb en France qui s'appelle Alaoui, Saidi ou Trabelsi est la même que la proportion d'immigrés et de descendants d'immigrés d'Europe du Sud en France qui s'appelle Rossi, Garcia ou Santos ; or, j'ignore si tel est le cas. C'est pourquoi les résultats des analyses suivantes ont une valeur indicative, plutôt que décisive.

L'annuaire téléphonique des « Pages blanches » compte 1 « Nicolas Saidi », 1 « Nicolas Alaoui » et 0 « Nicolas Trabelsi » (Figure 7.6a), contre 44 « Nicolas Rossi », 146 « Nicolas Garcia » et 85 « Nicolas Silva » (Figure 7.6b).<sup>282</sup> La relative rareté des « Nicolas » qui portent un nom de famille du Maghreb serait-elle due à la rareté des noms de famille du Maghreb dans l'annuaire ? Non, puisqu'on y trouve 95 « Mohamed Saidi », 55 « Mohamed Alaoui » et 44 « Mohamed Trabelsi » (Figure 7.6c). Si l'enquête TeO compte 1,5 à 2 fois plus de « Nicolas » originaires d'Europe du Sud que du Maghreb, l'annuaire en compte donc environ 138 fois plus.

Les réseaux sociaux grand public qui comptent au moins 1 « Nicolas Saidi », 1 « Nicolas Alaoui » ou 1 « Nicolas Trabelsi » (ce qui n'est pas le cas de « Copains d'avant », qui n'en compte aucun) comptent de 12 fois plus de « Nicolas » originaires d'Europe du Sud que du Maghreb (Instagram) jusqu'à 85 fois plus (comptes Facebook associés à un individu venant de France ou y habitant), en passant par 57 fois plus (comptes Skype localisés en France) (tableau 7.7). Les réseaux sociaux professionnels qui comptent au moins 1 « Nicolas » qui porte un nom de famille du Maghreb (ce qui n'est pas le cas de Academia, qui n'en compte aucun) comptent de 52 (Viadeo) à 196 (LinkedIn) fois plus de « Nicolas » originaires d'Europe du Sud que du Maghreb (tableau 7.7). Quant aux naissances, mariages et décès en France depuis 1970 enregistrés dans la base de données généalogiques Geneanet, ils concernent 63 fois plus souvent un « Nicolas » originaire d'Europe du Sud que du Maghreb (tableau 7.7). Quant aux données exhaustives de décès en France depuis 1970, elles concernent 41 fois plus souvent un « Nicolas » originaire d'Europe du Sud que du Maghreb (tableau 7.8).

---

<sup>282</sup> Je comptabilise manuellement les individus (i) dont au moins le premier prénom indiqué est exactement « Nicolas » ou « Mohamed » (à l'exclusion de tous les « Nicolás », « Nicola », « Mohammed », etc.) et (ii) dont au moins un nom de famille indiqué est exactement « Saidi », « Alaoui », « Trabelsi », « Rossi », « Garcia » ou « Silva » (y compris tout nom de famille composé incluant ces noms, comme « El Alaoui » et « Da Silva »).

**❖ À la recherche de petits-fils d’immigrés du Maghreb prénommés  
« Nicolas » : dans les médias**

Peut-être les 3 % de petits-fils d’immigrés du Maghreb qui, en 2008, étaient prénommés « Nicolas » (tableau 7.4) sont-ils mentionnés dans les médias ? Parmi 11 titres de la presse écrite nationale française depuis le 1<sup>er</sup> janvier 1970 (*L’AGEFI Quotidien*, *Aujourd’hui en France*, *La Correspondance économique*, *La Croix*, *Les Échos*, *Les Échos — Avis financiers*, *Le Figaro*, *L’Humanité*, *Libération*, *Le Monde* et *La Tribune*), pas un seul n’a publié un seul article contenant « Nicolas Saidi », « Nicolas Alaoui » ou « Nicolas Trabelsi », alors qu’ils ont publié 70 articles mentionnant « Nicolas Rossi », « Nicolas Garcia » ou « Nicolas Silva » et 24 articles mentionnant « Mohamed Saidi », « Mohamed Alaoui » ou « Mohamed Trabelsi » (tableau 7.8). Beaucoup plus généralement, parmi 775 titres de la presse française depuis 1970, on compte 735 fois plus de mentions de « Nicolas » originaires d’Europe du Sud que du Maghreb (tableau 7.8). Les archives de l’Institut national de l’audiovisuel (INA) comptent quant à elles 22 fois plus d’audios, vidéos, pubs et créations web mentionnant des « Nicolas » originaires d’Europe du Sud que du Maghreb, même si ces effectifs sont faibles (tableau 7.8).

Parmi les moteurs de recherche sur Internet qui comptent au moins 1 « Nicolas Saidi », 1 « Nicolas Alaoui » ou 1 « Nicolas Trabelsi » (ce qui n’est pas le cas des pages françaises de Google Scholar, qui n’en comptent aucun), certains comptent 8 fois plus de réponses aux requêtes relatives aux « Nicolas » originaires d’Europe du Sud que du Maghreb (Bing et Yahoo), d’autres 399 fois plus (Google) (tableau 7.8). Ni en français ni en anglais l’encyclopédie Wikipedia ne compte une seule mention d’un « Nicolas Saidi », « Nicolas Alaoui » ou « Nicolas Trabelsi » (tableau 7.8). Et parmi les dizaines de milliers d’ouvrages publiés en français depuis 1970,<sup>283</sup> on ne compte aucune mention de « Nicolas Saidi », « Nicolas Alaoui » ou « Nicolas Trabelsi » (tableau 7.9a), contre plusieurs dizaines de mentions de « Nicolas Rossi » et de « Nicolas Garcia » (mais pas de « Nicolas Silva ») (tableau 7.9b).

---

<sup>283</sup> Le corpus francophone de [Ngram Viewer](#), qui comprend d’environ 5 000 ouvrages annuels dans les années 1970 à plus de 12 000 ouvrages annuels dans les années 2000 (Héran 2015, p. 530), a déjà été utilisé avec profit dans des publications de démographie (Héran 2015).

## Fréquence d'occurrence de « Nicolas » au nom de famille du Maghreb, dans l'annuaire téléphonique en avril 2021

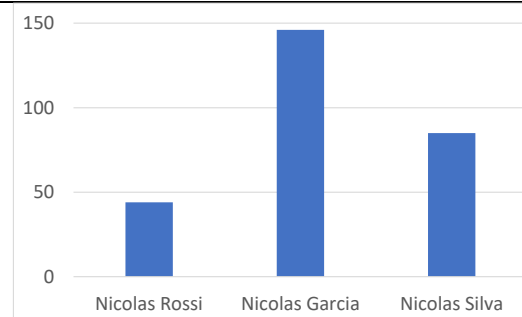
**Figure 7.6a. « Nicolas » au nom de famille du Maghreb**



Champ de la requête : France entière.

Source : <https://www.pagesjaunes.fr/pagesblanches>. Requête [ici](#).

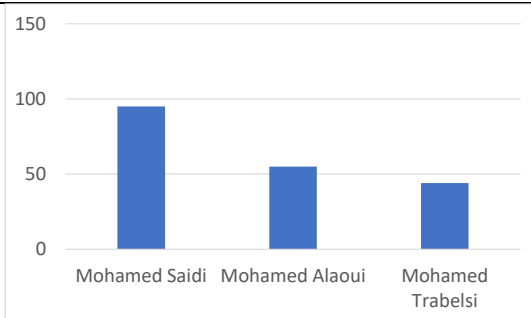
**Figure 7.6b. « Nicolas » au nom de famille d'Europe du Sud**



Champ de la requête : France entière.

Source : <https://www.pagesjaunes.fr/pagesblanches>.

**Figure 7.6c. « Mohamed » au nom de famille du Maghreb**



Champ de la requête : France entière.

Source : <https://www.pagesjaunes.fr/pagesblanches>.

**Tableau 7.7. Fréquence d'occurrence de « Nicolas » au nom de famille du Maghreb, parmi les utilisateurs des réseaux sociaux en France en avril 2021**

		Grand public				Professionnel			Généalogie
		Facebook (France)	Instagram	Skype (France)	<a href="#">Copains d'avant</a>	<a href="#">LinkedIn</a> (France)	<a href="#">Viadeo</a>	<a href="#">Academia</a> (« people »)	<a href="#">Geneanet</a> (France)
« Nicolas » au nom de famille du Maghreb	Nicolas Saidi	2	11	1	0	2	2	0	2
	Nicolas Alaoui	1	0	0	0	0	0	0	0
	Nicolas Trabelsi	0	0	0	0	0	0	0	0
« Nicolas » au nom de famille d'Europe du Sud	Nicolas Rossi	40	43	8	41	52	17	50	25
	Nicolas Garcia	74	41	28	135	213	56	1 024	87
	Nicolas Silva	142	45	21	62	127	30	463	13
« Mohamed » au nom de famille du Maghreb	Mohamed Saidi	23	25	98	42	73	76	63	60
	Mohamed Alaoui	24	54	56	29	57	121	52	14
	Mohamed Trabelsi	61	54	63	19	56	52	48	28

NB : Je comptabilise manuellement les individus (i) dont au moins le premier prénom indiqué est exactement « Nicolas » ou « Mohamed » (à l'exclusion de tous les « Nicolás », « Nicola », « Mohammed », etc.) et (ii) dont au moins un nom de famille indiqué est exactement « Saidi », « Alaoui », « Trabelsi », « Rossi », « Garcia » ou « Silva » (y compris tout nom de famille composé incluant ces noms, comme « El Alaoui » et « Da Silva »). Sur Facebook et Instagram, la requête porte sur le nom mais aussi sur le pseudo. Sur Facebook, Skype et LinkedIn, la requête est restreinte aux comptes dont il est indiqué que l'individu associé vient de France ou y habite.

**Tableau 7.8. Fréquence d'occurrence de « Nicolas » au nom de famille du Maghreb,  
dans les médias en France jusqu'en avril 2021**

		Presse écrite depuis 1970 ( <a href="#">Europresse</a> )		<a href="#">Audiovisuel</a> (INA, 2021)				Moteurs de recherche Internet (2021)				Encyclopédie Wikipedia (2021)		<a href="#">Décès en France depuis 1970</a>
		Nationale	Totale	Audios	Vidéos	Pubs	Web	<a href="#">Google</a>	<a href="#">Google scholar</a>	<a href="#">Bing</a>	<a href="#">Yahoo</a>	<a href="#">français</a>	<a href="#">anglais</a>	
« Nicolas » au nom de famille du Maghreb	Nicolas Saidi	0	4	0	0	1	0	837	0	38 200	38 300	0	0	1
	Nicolas Alaoui	0	0	0	0	0	0	309	0	44	43	0	0	0
	Nicolas Trabelsi	0	1	0	0	0	0	95	0	18	18	0	0	0
« Nicolas » au nom de famille d'Europe du Sud	Nicolas Rossi	0	198	2	5	1	0	21 600	23	42 600	42 600	2	0	11
	Nicolas Garcia	70	3 441	1	10	0	0	387 000	82	203 000	203 000	57	29	29
	Nicolas Silva	0	37	0	3	0	0	86 000	6	71 400	71 600	1	4	1
« Mohamed » au nom de famille du Maghreb	Mohamed Saidi	7	242	0	1	0	0	54 700	116	64 500	64 400	24	122	47
	Mohamed Alaoui	3	39	0	2	0	0	88 900	88	26 600	26 600	6	4	4
	Mohamed Trabelsi	14	107	0	1	0	0	138 000	86	26 300	26 400	157	17	24

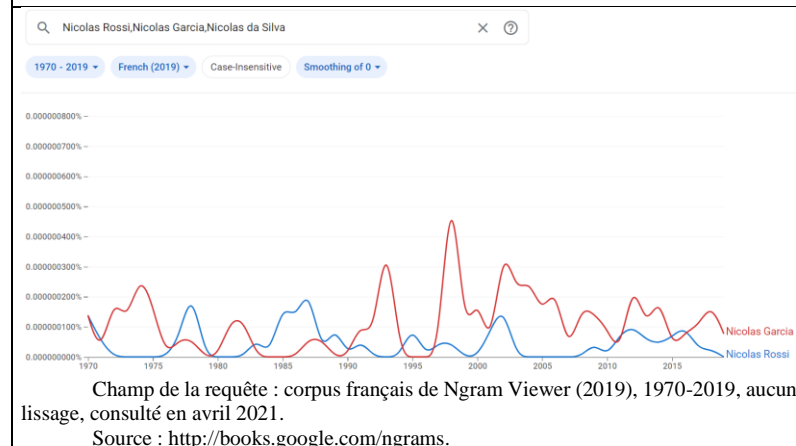
NB : Je reporte le nombre de réponses aux requêtes « Nicolas Saidi », « Nicolas Alaoui », « Nicolas Trabelsi », « Nicolas Rossi », etc.

## Fréquence d'occurrence de « Nicolas » au nom de famille du Maghreb, dans les livres publiés en français de 1970 à 2019

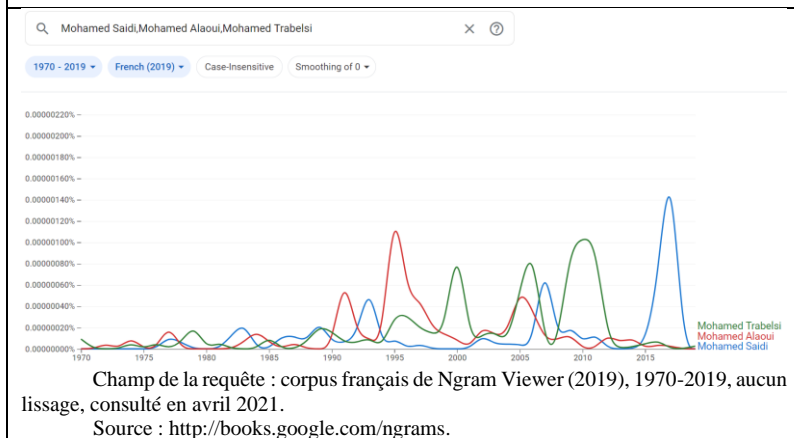
**Figure 7.9a. « Nicolas » au nom de famille du Maghreb**



**Figure 7.9b. « Nicolas » au nom de famille d'Europe du Sud**



**Figure 7.9c. « Mohamed » au nom de famille du Maghreb**





Au total, ni l'intuition commune, ni les experts des migrations qui ont bien voulu répondre à mes questions, ni les recherches par entretiens menées depuis les années 1980 auprès des couples « mixtes », ni surtout l'annuaire téléphonique ou les réseaux sociaux en 2021, ni la presse écrite, ni la radio et la télévision françaises, ni les ouvrages publiés en français depuis 1970, ni les moteurs de recherche Internet ou l'encyclopédie Wikipedia, ni même les décès enregistrés à l'état civil, ne semblent compter suffisamment de « Nicolas » originaires du Maghreb (par rapport aux « Nicolas » originaires d'Europe du Sud) pour qu'en 2008 « Nicolas » ait pu faire partie des prénoms les plus courants parmi les petits-fils d'immigrés du Maghreb, ou pour qu'à cette même date plus de 3 % des petits-fils d'immigrés du Maghreb aient pu se prénommer « Nicolas » — loin de là (tableaux 7.6 à 7.9). Alors, que se passe-t-il ? Cette accumulation d'indices conduit à envisager que le tableau 7.4 et le tableau de Coulmont et Simon (2019, p. 3) contiennent une erreur, qui proviendrait des bases de données de l'enquête TeO.

Au fil de mes investigations, j'ai toutefois été amené à entrer personnellement en contact avec un petit-fils d'immigrés du Maghreb prénommé « Nicolas » : il s'agit d'un petit-fils d'immigrés (G3) du Maghreb dont les deux grands-parents paternels (G1) sont des immigrés d'Algérie, et dont le père (G2) portait un prénom arabo-musulman mais en a changé suite à sa conversion au christianisme orthodoxe. C'est dans ce contexte qu'avec sa compagne ils ont choisi, pour leur fils (G3), le prénom « Nicolas » (ainsi qu'un second prénom arabo-musulman). Même ce rare exemple apparaît révélateur : pour qu'un enfant d'immigrés du Maghreb choisisse de prénommer son fils « Nicolas », des conditions très particulières semblent requises, comme en l'occurrence une conversion au christianisme (orthodoxe).

#### ❖ Quelle erreur l'enquête TeO pourrait-elle contenir ?

Parvenu à ce stade, me voilà conduit à envisager que les bases de données d'une enquête de la statistique publique — qui ont été exploitées depuis une dizaine d'années par des dizaines de chercheurs qui en ont tiré [des dizaines de publications](#) — contiendraient une erreur qui n'aurait pas été signalée. On peut envisager plusieurs types d'erreur.

Tout d'abord, on pourrait en principe envisager une erreur d'échantillonnage (plan de sondage et/ou pondération). Mais comme on l'a vu tout au long du présent travail (du tableau 5.5 à la figure 5.10, puis de la figure 6.10 à la figure 6.11, et de nouveau sur la figure 7.1), les résultats obtenus à partir de l'échantillon TeO ressemblent de près aux résultats obtenus à partir des données exhaustives de l'état civil. Concrètement, non seulement rien n'accrédite la thèse d'une erreur d'échantillonnage, mais bien au contraire tout semble indiquer que l'échantillonnage de l'enquête TeO est de (très) bonne qualité. On pourrait alors envisager une erreur de (re)pondération. Serait-ce parce qu'un tout petit nombre de « Nicolas » originaires du Maghreb sont massivement pondérés que les « Nicolas » originaires du Maghreb paraissent si nombreux ?<sup>284</sup> Les 17 « Nicolas » de l'échantillon retenu non pondéré de l'enquête TeO qui sont originaires du Maghreb (qui composent, une fois pondérés, les 10 « Nicolas » originaires du Maghreb : 3 G2 et 6 G3) ont un poids moyen de 625 (d'un minimum de 84 à un maximum de 5 147), tandis que les 42 191 autres individus du champ retenu de l'enquête TeO non pondérée ont un poids moyen de 1 112 (d'un minimum de 8 à un maximum de 59 701). Ce n'est donc pas la pondération qui génère ce résultat. Pour expliquer l'erreur du tableau 7.4, on pourrait aussi envisager que la variable « prénom » comporte une erreur. Mais là encore, on a vu tout au long du présent travail que cette variable de l'enquête TeO est de (très) bonne qualité,

<sup>284</sup> On a vu (Annexe 6B, p. 285) que certaines observations, comme l'apparente prévalence du prénom « Murielle » chez les personnes originaires de Turquie, ne sont effectivement dues qu'à la forte pondération d'un nombre d'individus très restreint (<5).

sans quoi elle ne permettrait pas de reproduire à l'identique des dizaines de résultats issus des prénoms attribués à l'état civil (fichier INSEE des prénoms).

Par conséquent, l'erreur provient plus probablement d'une erreur sur la principale autre variable utilisée pour construire le tableau 7.4, à savoir la variable d'origine des individus (ORIG\_IND, variable 11 du tableau 5.1, dont le tri à plat est présenté dans le tableau 5.3). Cette variable d'origine des individus est elle-même construite (correctement d'après moi) à partir des variables (possiblement corrompues) GROUP2 (statut d'ego par rapport à la migration en 14 positions), GROUP2CA (statut détaillé du conjoint actuel d'ego par rapport à la migration) et GROUP2PC (statut détaillé du premier conjoint d'ego par rapport à la migration), qui sont des variables « de nature calculée » (Insee 2011, p. 241-243).

Précisément, on peut envisager trois sources d'erreur différentes sur ces variables :

- **1<sup>e</sup> possibilité : au stade du calcul des variables GROUP2, GROUP2CA et GROUP2PC, le groupe d'exploitation de l'enquête TeO aurait de bonne foi commis une erreur de calcul**

Un examen de l'algorithme de création de la variable GROUP2, tel que présenté dans le [Guide d'utilisation et le dictionnaire des codes de l'enquête TeO 2008](#) (Insee 2011, p. 241), révèle effectivement une erreur<sup>285</sup> : à la 9<sup>e</sup> ligne en partant de la fin du programme (Insee 2011, p. 241), on lit « sinon si Group1=1 », alors qu'il aurait fallu écrire « sinon si Group1=5 ». J'ai prévenu les méthodologues de l'INSEE, qui m'ont indiqué que si le dictionnaire des codes contient effectivement une coquille, la variable semble avoir été calculée correctement. Au passage, j'observe que face à mes questions l'attitude de l'INSEE, contrairement à celle de l'INED (Chapitre 4, p. 15), a été d'une irréprochable honnêteté et transparence : ils m'ont répondu immédiatement, et sur le fond, ce qui les a conduits à admettre l'erreur du dictionnaire des codes, tout en m'indiquant que la base de données n'était vraisemblablement pas contaminée, même si cela reste à confirmer auprès de l'INED. Pour expliquer comment ces « Nicolas » de l'enquête TeO peuvent se retrouver catégorisés comme petits-fils d'immigrés du Maghreb, il convient vraisemblablement d'envisager un autre type d'erreur.

- **2<sup>e</sup> possibilité : au stade de la passation du questionnaire TeO, des enquêtés auraient fourni de bonne foi des réponses erronées aux questions utilisées pour calculer les variables GROUP2, GROUP2CA et GROUP2PC**

Parmi les questions posées aux enquêtés pour déterminer s'ils sont ou non descendants d'immigrés (plutôt que de faire partie de la population « majoritaire », sans ascendant immigré), certaines sont relativement complexes à comprendre, et/ou la réponse vraie est objectivement complexe à connaître, sans compter qu'elle peut être difficile à se remémorer. Pour déterminer si un individu est ou non immigré — c'est-à-dire, né étranger à l'étranger —, la question difficile est de connaître sa nationalité à la naissance. À cette difficulté s'en ajoute une autre, spécifique aux individus originaires d'Algérie qui sont nés avant l'indépendance (1962) : en effet, les anciens « indigènes » d'Algérie sont nés de nationalité française (sans être pour autant citoyens), ce qui devrait suffire à ne pas les catégoriser comme « immigrés », mais la convention adoptée par l'enquête TeO — que je ne conteste pas — veut que si la nationalité algérienne leur a été attribuée à l'indépendance, avant qu'ils migrent vers la France métropolitaine, ils soient malgré tout catégorisés comme « immigrés », les distinguant en cela des anciens Français et Européens d'Algérie, de nationalité française suite à l'indépendance parce que de statut personnel de droit civil et qui sont, à ce titre, catégorisés comme « rapatriés ».<sup>286</sup>

---

<sup>285</sup> En revanche et d'après ma compréhension, les algorithmes de création des variables GROUP2CA et GROUP2PC ne révèlent pas d'erreur (Insee 2011, p. 242-243).

<sup>286</sup> Selon le [Guide d'utilisation et le dictionnaire des codes de l'enquête TeO 2008](#), les rapatriés de l'ex-empire colonial français en Afrique du Nord (et leurs descendants) font partie du groupe « majoritaire », qui se distingue, notamment, des immigrés (et de leurs descendants) (Insee 2011, p. 5). Dans l'enquête TeO les descendants de

Rappelons que parmi les individus nés en Algérie coloniale à partir de 1865 (et jusqu'en 1962), les « indigènes » (colonisés, éventuels futurs « immigrés ») et les « Français d'Algérie » (colons, futurs « rapatriés ») sont (presque) tous nés français (Blévis 2001). En effet, les indigènes devaient être considérés comme de nationalité française, sans quoi l'État aurait dû reconnaître une nationalité algérienne — et donc la réalité de l'occupation coloniale (Blévis 2001, p. 566). Ce qui distingue juridiquement les « indigènes » des (futurs) « rapatriés » français, ce n'est donc pas leur nationalité à la naissance, mais leur statut civique et politique : contrairement aux colons, qui sont non seulement de nationalité française mais aussi pleinement citoyens, les « indigènes » ne sont pas pleinement citoyens (Blévis 2001). Précisément, ce qui distinguait juridiquement les « indigènes » des (futurs) « rapatriés », c'est leur statut personnel : contrairement aux colons français mais aussi espagnols ou italiens, dits « de statut civil de droit commun » (soumis au Code civil), les « indigènes » étaient « de statut civil de droit local » (soumis à un autre code de statut personnel), à moins qu'ils aient renoncé à leur statut personnel pour le droit civil ou qu'ils aient accédé au statut de citoyen français par décret. À l'indépendance de l'Algérie (1962), seules les personnes de « statut civil de droit commun » ont été reconnues comme ayant la nationalité française, qu'elles résident en France ou en Algérie (Massicot 1986, p. 540-542). Les personnes « de statut civil de droit local » résidant en Algérie ont perdu la nationalité française. Quant aux personnes « de statut civil de droit local » qui résidaient en métropole ou dans les DOM, elles ont pu souscrire une déclaration de reconnaissance de la nationalité française jusqu'en 1967. Rappelons aussi que, parmi les individus nés dans le Maroc et la Tunisie à l'époque des protectorats (jusqu'en 1956), les « protégés » (colonisés, éventuels futurs « immigrés ») sont (presque) tous nés de nationalité marocaine ou tunisienne, tandis que les colons sont (presque) tous nés français, espagnols ou italiens. Leur nationalité à la naissance distingue bien les « protégés » des colons français et européens, mais là encore, ce qui fait la différence entre d'une part les « protégés » et d'autre part les colons français ou européens, c'est leur statut personnel.

De fait, le droit colonial de la nationalité est une matière d'une très grande complexité. La difficulté à distinguer les descendants d'immigrés et les descendants de rapatriés est d'ailleurs bien connue de la statistique publique, qu'il s'agisse du recensement (Simon 1998, p. 546-553) ou de l'échantillon démographique permanent (Couto 2013).<sup>287</sup> Au fond, seuls quelques experts du droit colonial de la nationalité et de la citoyenneté, comme Laure Blévis (2001, 2004), connaissent suffisamment le domaine pour s'y retrouver. En conséquence, il ne

---

rapatriés sont définis comme les « Personnes nées en France métropolitaine dont au moins un parent est né Français dans un ancien territoire colonial avant sa décolonisation » ; en cas de double appartenance, « La priorité est donnée à l'ascendance immigrée, à défaut à l'ascendance DOM, et dans le cas où l'un des parents est rapatrié et l'autre est immigré ou natif de DOM » ([site TeO](#)).

<sup>287</sup> Concernant l'échantillon démographique permanent (EDP), « Les biais liés à l'identification par le nom et le prénom peuvent être de deux ordres. D'abord, des pieds-noirs ont pu, à tort, être qualifiés de "musulmans". Nous ne sommes pas en mesure d'évaluer l'ampleur de cette confusion que l'on peut cependant considérer comme limitée car, pour opérer son classement, l'administration statistique s'est appuyée sur une liste des "prénoms musulmans les plus usuels relevés sur les registres d'état civil en Algérie" et, en l'occurrence, il semble peu probable que les Européens aient opté pour ce type de prénom (Spire et Merllié, 1999 : 119-130). Toutefois, la frontière taxinomique entre les citoyens israélites et les « indigènes musulmans » reste poreuse. En conséquence, certains citoyens israélites ont été probablement évincés de la catégorie des pieds-noirs. Inversement, une part des anciens "indigènes musulmans" a pu échapper à l'identification des agents recenseurs. Cependant, des corrections ont été apportées aux classements faits par les agents de l'INSEE en 1968, en observant les réponses des individus aux questions sur la nationalité dans les recensements ultérieurs. Effectivement, si à l'une des dates suivantes (1975, 1982, 1990, 1999), les ressortissants de l'ancienne colonie déclaraient disposer ou avoir disposé de la nationalité algérienne, ils étaient écartés de la catégorie "pieds-noirs". Les "indigènes musulmans" ne portant pas de nom ni de prénom à consonance arabe ou berbère et s'étant toujours identifiés comme Français, demeurent dans l'échantillon, mais on suppose, d'une part, que leur nombre est restreint par rapport à l'effectif des pieds-noirs et, d'autre part, que leurs caractéristiques sociales avoisinent, dans ce cas particulier, celles des Européens d'Algérie » (Couto 2013, p. 96).

serait pas particulièrement surprenant que des enquêtés fournissent de bonne foi des informations erronées, qui auraient pour conséquence que les variables GROUP2, GROUP2CA et GROUP2PC les catégorisent à tort comme des descendants d'immigrés du Maghreb alors qu'ils n'en sont pas.

Dans ce contexte, une hypothèse qui vient immanquablement à l'esprit est qu'**une partie des « Nicolas » qui sont catégorisés comme petits-enfants d'immigrés du Maghreb pourraient être, non pas des petits-enfants d'immigrés du Maghreb, mais des petits-enfants de rapatriés français ou européens de l'ex-empire colonial français en Afrique du Nord, autrement dit des petits-enfants de pieds-noirs.** Or, pour que l'enquête TeO catégorise à tort un *rapatrié* (d'Algérie, du Maroc ou de Tunisie) comme un *immigré* (du Maghreb), il suffirait :

- que l'enquêté, qui est un enfant de rapatriés français, ait indiqué à tort que son parent né en Afrique du Nord est *né étranger* alors qu'en réalité il est *né français* (NATPNIV2 et/ou NATMNIV2 renseignés à tort différents de 1101 et 9999)<sup>288</sup> ;
- ou que l'enquêté, qui est un enfant de rapatriés français, ait indiqué à tort que son parent né français en Algérie, au Maroc ou en Tunisie a *pris la nationalité de ce pays à l'indépendance* alors qu'en réalité il a *conservé sa nationalité française* suite à l'indépendance (N\_INDPER et/ou N\_INDMER renseignés à tort égaux à 1).<sup>289</sup>

<sup>288</sup> Comme l'écrivent les concepteurs de l'enquête TeO, « le point le plus sensible dans le recueil de la nationalité concerne celle des parents à la naissance. C'est en effet à partir de cette information que peuvent se différencier les descendants d'immigrés venant des anciennes colonies et les descendants de « rapatriés », c'est-à-dire des Français (anciens colons, pour la plupart) qui ont quitté les colonies pour (re)venir s'installer en France métropolitaine » (Beauchemin et al. 2016, p. 38). Si les questions sont formulées le plus simplement possible (« Quelle était sa nationalité [celle du père et/ou de la mère de l'enquêté] à sa naissance ? »), la réponse n'en est pas moins complexe à connaître pour certains enquêtés, dans un contexte où des questions préalables distinguaient la nationalité « à la naissance », « par réintégration », « par naturalisation », « par mariage » et « par déclaration ou option à la majorité ou avant » (Insee 2008, p. 17-23). On peut d'ailleurs se demander pourquoi le questionnaire TeO ne semblait pas inclure d'alerte automatique lorsqu'un enquêté indiquait que l'un de ses parents, né en Algérie avant 1962, était *né algérien*, à une époque où une telle nationalité n'existait tout simplement pas. Quoi qu'il en soit, parmi les 13 grands-parents immigrés du Maghreb des 11 petits-fils d'immigrés du Maghreb prénommés « Nicolas » que compte l'échantillon maximal TeO non pondéré, 11/13 sont renseignés comme étant nés étrangers.

<sup>289</sup> Dans le questionnaire de l'enquête TeO, « Lorsque le(s) parent(s) est (sont) né(s) à l'étranger, l'enquêté doit indiquer dans un premier temps si ce parent avait la nationalité française à la naissance. Une seconde question est ensuite posée aux enquêtés dont le parent est né dans une ancienne colonie française pour identifier s'il a pris la nationalité du pays devenu indépendant. Dans ce cas, le parent est considéré comme immigré, et l'enquêté est alors défini comme descendant d'immigré. Seule l'enquête TeO permet de faire la distinction, avec cette précision, entre descendants d'immigrés et d'anciens colons » (Beauchemin et al. 2016, p. 38). De même, « Le parent né Français de naissance dans une colonie qui a pris la nationalité du pays à l'indépendance est considéré comme immigré et l'enquêté est donc considéré descendant d'immigré(s) et pas descendant de rapatrié(s) » ([site TeO](#)). Là encore si la question est formulée le plus simplement possible (« Au moment de l'indépendance ou dans les années qui ont suivi, votre mère / père : A pris la nationalité du pays devenu indépendant / Est resté français / Était déjà décédé / N'était pas encore né / Ne sait pas »), la réponse n'en est pas moins complexe à connaître pour certains enquêtés, comme l'anticipaient d'ailleurs les concepteurs de l'enquête eux-mêmes, qui ont accompagné leur question du paragraphe suivant : « Si gêne ou hésitation : Les questions de nationalité et de citoyenneté sont très compliquées, c'est pourquoi je dois demander des précisions. Rappel : indépendance de l'Algérie en 1962, de la Guinée en 1958, de Djibouti en 1977, des Comores en 1975 et des autres pays d'Afrique en 1960. Indépendance du Vietnam en 1949. ». En outre, l'absence de la variable COLOE du dictionnaire des codes m'empêche de savoir si elle inclut le Maroc et la Tunisie parmi les anciennes colonies françaises : si tel n'était pas le cas, cela accentuerait encore le risque que TeO manque d'identifier des rapatriés du Maroc et de Tunisie qui auraient conservé la nationalité française à l'indépendance de ces pays. À toutes ces complexités s'ajoutent des complexités de calcul des variables, qui rendent plus difficile encore de débrouiller la situation : pourquoi les enquêtés qui sont français par réintégration sont-ils catégorisés comme immigrés (GROUP1=1) alors que leurs conjoints actuel et passé qui sont français par réintégration sont catégorisés comme membres du groupe « majoritaire » (GROUP1CA=5 et

Dans ces deux cas, non seulement l'enquêté serait catégorisé par TeO comme enfant d'immigrés (G2) du Maghreb plutôt que comme enfant de rapatriés d'Afrique du Nord, mais en outre le parent de l'enquêté serait catégorisé comme immigré (G1) du Maghreb plutôt que comme rapatrié d'Afrique du Nord, et surtout les enfants de l'enquêté seraient catégorisés comme petits-enfants d'immigrés (G3) du Maghreb plutôt que comme petits-enfants de rapatriés d'Afrique du Nord. Voilà qui pourrait éventuellement expliquer pourquoi des individus catégorisés par l'enquête TeO comme petits-enfants d'immigrés du Maghreb portent des prénoms qui semblent plus courants chez les petits-enfants de pieds-noirs.

- **3<sup>e</sup> possibilité : au stade de l'exploitation de l'enquête TeO, Coulmont et Simon n'ont pas utilisé toutes les variables nécessaires pour éviter que certains rapatriés européens d'Afrique du Nord (et leurs descendants) n'apparaissent comme des immigrés du Maghreb (et leurs descendants)**

Un examen de ces sujets fait apparaître que, pour que le tableau de Coulmont et Simon (2019) et le tableau 7.4 catégorisent à tort un *rapatrié d'Afrique du Nord* comme un *immigré du Maghreb*, il suffirait aussi :

- que l'enquêté, qui est un enfant de rapatriés européens d'Afrique du Nord (disons d'origine espagnole, italienne ou allemande), ait indiqué à *raison* que son parent né en Afrique du Nord y est *né espagnol, italien ou allemand* : comme d'une part cet individu est né étranger à l'étranger et donc immigré, et que d'autre part il est né en Afrique du Nord, le tableau de Coulmont et Simon et le tableau 7.4 le catégoriseraient à tort comme immigré du Maghreb (et tous ses descendants comme descendants d'immigrés du Maghreb), alors qu'il est né étranger en Afrique du Nord à une époque où l'Afrique du Nord était française — avant d'être rapatrié de l'Afrique du Nord jusqu'en métropole au moment où l'Afrique du Nord a cessé d'être française.

Au total, diverses erreurs commises par plusieurs participants et à plusieurs stades de l'enquête ont pu conduire l'enquête TeO ou les auteurs de l'article à catégoriser à tort des descendants de rapatriés de l'ex-empire colonial français en Afrique du Nord comme des descendants d'immigrés. En l'occurrence, parmi les 13 grands-parents immigrés du Maghreb des 11 petits-fils d'immigrés du Maghreb prénommés « Nicolas » que compte l'échantillon maximal TeO non pondéré (les  $4 \times 11 - 13 = 31$  autres grands-parents des 11 petits-fils d'immigrés du Maghreb prénommés « Nicolas » n'étant pas des immigrés du Maghreb), 11/13 sont renseignés comme étant nés étrangers et 2/13 sont renseignés comme ayant pris la nationalité du pays à l'indépendance. Quoi qu'il en soit, je ne peux identifier quelles erreurs ont été commises et je ne peux les corriger. Seuls les méthodologues de l'INSEE et de l'INED pourront le faire, eux qui ont toutes les informations nécessaires.<sup>290</sup> Pour ce qui me concerne, je propose de *contourner* les problèmes de catégorisation des immigrés et des rapatriés et d'utiliser d'autres variables, plus fiables, pour examiner si les « Nicolas » catégorisés comme petits-fils d'immigrés du Maghreb ne seraient pas, en réalité, des petits-fils de pieds-noirs.

On notera toutefois que pour opérer une distinction juridique nette entre immigrés du Maghreb (et leurs descendants) et rapatriés d'Afrique du Nord (et leurs descendants), il faudrait demander si, pendant la période coloniale, l'enquêté né dans un pays d'Afrique du Nord ou ses ascendants étaient « de statut civil de droit local » (immigrés, ex-colonisés arabes et berbères)

---

GROUP1PC=5)? Quoi qu'il en soit, parmi les 13 grands-parents immigrés du Maghreb des 11 petits-fils d'immigrés du Maghreb prénommés « Nicolas » que compte l'échantillon maximal TeO non pondéré, 2/13 sont renseignés comme ayant pris la nationalité du pays à l'indépendance.

<sup>290</sup> Ces informations incluent les programmes de création des variables « calculées », la littérature grise relative à l'enquête et les informations relatives à des variables absentes du dictionnaire des codes. Voir les deux notes de bas de page ci-dessus.

ou « de statut civil de droit commun » (rapatriés, ex-colons français et européens). Vu la complexité d'une telle question, la meilleure solution reste peut-être de poser aux enquêtés nés dans un pays d'Afrique du Nord et arrivés en France jusqu'à l'indépendance la question suivante : au moment de votre arrivée en France, étiez-vous (votre ascendant était-il) un rapatrié, ou un Algérien (ou futur Algérien), un Marocain, un Tunisien ? Cette question est assez proche de celle posée par l'IFOP, qui demande tout simplement : « Êtes-vous pied-noir ? » et « Un ou plusieurs de vos parents / grands-parents sont-ils / étaient-ils pieds-noirs ? » (Fourquet 2014).

## Annexe 8 : Le « 23 % » de la figure 2 page 3

Cette Annexe 8 comprend les tableaux de résultats du début du Chapitre 8, p. 46.

**Tableau 8.1a. Proportion de prénoms spécifiques par origine et « génération » en France en 2008 (enquête TeO) — champ retenu**

		Données pondérées				Données non pondérées			
		N	G1	G2	G3	N	G1	G2	G3
ORIGINE GÉOGRAPHIQUE DES IMMIGRÉS ET DE LEURS DESCENDANTS	<b>Groupe « majoritaire »</b>	31 775	-	-	-	6 437	-	-	-
	<b>DOM</b>	616	79 %	64 %	46 %	2 311	81 %	61 %	47 %
	<b>Europe du Sud</b>	2 704	58 %	12 %	8 %	7 763	56 %	14 %	10 %
	<b>Autre Europe</b>	1 538	56 %	23 %	19 %	3 737	55 %	32 %	26 %
	<b>Maghreb</b>	<del>3 155</del> 2 881	<del>93 %</del> 94 %	<del>70 %</del> 76 %	<del>37 %</del> 49 %	<del>9 477</del> 8 646	<del>93 %</del> 94 %	<del>71 %</del> 75 %	<del>34 %</del> 45 %
	<b>Autre Afrique</b>	1 025	24 %	20 %	9 %	5 479	32 %	29 %	14 %
	<b>Turquie</b>	415	86 %	71 %	39 %	2 794	87 %	78 %	45 %
	<b>Asie</b>	577	46 %	10 %	5 %	3 136	51 %	11 %	4 %
	<b>Ailleurs</b>	403	-	-	-	1 073	-	-	-
	<b>Non réponse</b>	0	-	-	-	1	-	-	-
	<b>Total</b>	42 208	-	-	-	42 208	-	-	-

Lecture : 58 % des immigrés d'Europe du Sud portent un prénom spécifique (« latin »), d'après la nomenclature de l'INED.

Champ : individus du champ retenu (personnes de 18-50 / 18-60 ans en ménage ordinaire et leurs enfants de 0-17 ans habitant en France métropolitaine en 2008) qui sont G1 ou G2 ou G3, N = 42 208. Ce champ est représentatif des personnes de 18-50 / 18-60 ans en ménage ordinaire et de leurs enfants de 0-17 ans, habitant en France métropolitaine en 2008.

NB : Les estimations concernant les personnes originaires du Maghreb (barrées, car elles comprennent de probables rapatriés et descendants de rapatriés d'Afrique du Nord) sont remplacées par des estimations sur le champ des personnes originaires du Maghreb qui ne sont probablement pas rapatriés ou descendants de rapatriés d'Afrique du Nord.

Source : TeO (voir Annexe 5B, p. 183).



**Tableau 8.1b. Proportion de prénoms spécifiques par origine et « génération » en France en 2008 (enquête TeO) — champ maximal**

		Données pondérées				Données non pondérées			
		N	G1	G2	G3	N	G1	G2	G3
<b>ORIGINE GÉOGRAPHIQUE DES IMMIGRÉS ET DE LEURS DESCENDANTS</b>	<b>Groupe « majoritaire »</b>	60 956	-	-	-	17 818	-	-	-
	<b>DOM</b>	1 208	82 %	63 %	49 %	3 922	80 %	62 %	47 %
	<b>Europe du Sud</b>	5 380	54 %	11 %	8 %	12 295	55 %	13 %	10 %
	<b>Autre Europe</b>	2 901	55 %	23 %	22 %	5 670	52 %	33 %	26 %
	<b>Maghreb</b>	6 403	92 %	72 %	39 %	17 810	92 %	74 %	35 %
	<b>Autre Afrique</b>	2 046	23 %	25 %	19 %	10 632	34 %	31 %	14 %
	<b>Turquie</b>	855	83 %	74 %	59 %	5 453	85 %	78 %	52 %
	<b>Asie</b>	1 225	40 %	13 %	2 %	6 176	48 %	16 %	3 %
	<b>Ailleurs</b>	775	-	-	-	1 935	-	-	-
	<b>Non réponse</b>	1 297	-	-	-	1 334	-	-	-
	<b>Total</b>	83 045	-	-	-	83 045	-	-	-

Lecture : 54 % des immigrés d'Europe du Sud portent un prénom spécifique (« latin »), d'après la nomenclature de l'INED.

Champ maximal des individus TeO qui sont G1 ou G2 ou G3, N = 83 045. NB : ce champ n'est présenté qu'à titre indicatif, il n'est pas représentatif des personnes de 18-50 / 18-60 ans en ménage ordinaire et de leurs enfants de 0-17 ans, habitant en France métropolitaine en 2008. En outre, les estimations concernant les personnes originaires du Maghreb comprennent de probables rapatriés et descendants de rapatriés d'Afrique du Nord, d'environ 1 % chez les immigrés (G1) à environ 26 % chez les petits-enfants d'immigrés (G3) du Maghreb.

Source : TeO (voir Annexe 5B, p. 183).



**Tableau 8.1c. Proportion de prénoms spécifiques par origine et « génération » en France en 2008 (enquête TeO) — champ enquêtés et tous leurs enfants vivants**

		Données pondérées				Données non pondérées			
		N	G1	G2	G3	N	G1	G2	G3
<b>ORIGINE GÉOGRAPHIQUE DES IMMIGRÉS ET DE LEURS DESCENDANTS</b>	<b>Groupe « majoritaire »</b>	39 056	-	-	-	8 111	-	-	-
	<b>DOM</b>	814	79 %	63 %	48 %	2 874	81 %	62 %	47 %
	<b>Europe du Sud</b>	3 834	58 %	11 %	7 %	9 646	56 %	14 %	10 %
	<b>Autre Europe</b>	2 144	56 %	27 %	21 %	4 583	55 %	33 %	26 %
	<b>Maghreb</b>	4 019	93 %	71 %	38 %	11 876	93 %	73 %	34 %
	<b>Autre Afrique</b>	1 287	24 %	20 %	19 %	6 949	32 %	28 %	14 %
	<b>Turquie</b>	493	86 %	72 %	39 %	3 416	87 %	78 %	45 %
	<b>Asie</b>	741	46 %	13 %	3 %	4 057	51 %	16 %	4 %
	<b>Ailleurs</b>	515	-	-	-	1 390	-	-	-
	<b>Non réponse</b>	0	-	-	-	1	-	-	-
	<b>Total</b>	52 903	-	-	-	52 903	-	-	-

Lecture : 58 % des immigrés d'Europe du Sud portent un prénom spécifique (« latin »), d'après la nomenclature de l'INED.

Champ : enquêtés et tous leurs enfants vivants en France métropolitaine en 2008 qui sont G1 ou G2 ou G3, données pondérées, N = 52 903. NB : ce champ n'est présenté qu'à titre indicatif, il n'est pas représentatif des personnes de 18-50 / 18-60 ans en ménage ordinaire et de leurs enfants de 0-17 ans, habitant en France métropolitaine en 2008. En outre, les estimations concernant les personnes originaires du Maghreb comprennent de probables rapatriés et descendants de rapatriés d'Afrique du Nord, d'environ 1 % chez les immigrés (G1) à environ 26 % chez les petits-enfants d'immigrés (G3) du Maghreb.

Source : TeO (voir Annexe 5B, p. 183).