

Annexe méthodologique à l'article « Quel prénom les immigrés donnent-ils à leurs enfants ? »

Baptiste Coulmont et Patrick Simon

Les données utilisées dans l'article proviennent de l'enquête Trajectoires et Origines. Les caractéristiques de l'enquête TeO sont décrites sur le site de l'enquête (<https://teo1.site.ined.fr/fr/>) et dans l'ouvrage publié par les éditions de l'Ined en 2016 (<https://www.ined.fr/fr/publications/editions/grandes-enquetes/trajectoires-et-origines/>).

Les informations sur les prénoms sont collectées pour les 21800 personnes enquêtées, ainsi que pour toutes les personnes vivant dans leur ménage et pour leurs enfants ayant décohabité. Le nombre total de prénoms disponibles est ainsi de 106.597, portés par différentes personnes constituant les ménages ou directement enquêtées et qui ne sont pas toutes prises en compte dans l'étude. Pour éviter les doubles comptes, notre champ d'étude est en effet circonscrit aux personnes enquêtées et à leurs enfants, répartis en générations vis-à-vis de l'immigration : immigrés, descendants d'immigrés, petits enfants d'immigrés et personnes sans lien à l'immigration sur trois générations. Pour reconstituer ces groupes nous avons utilisé les informations issues de la base principale portant sur les personnes ayant répondu au questionnaire (n=21 767) qui se répartissent en 8458 immigrés, 8164 descendants d'immigré(s) et 5145 personnes d'ascendance française sur deux générations (dont 1363 originaires des Dom et leurs enfants nés dans l'hexagone). L'échantillonnage de l'enquête TeO a été conçu spécifiquement pour sur-représenter les immigrés, leurs descendants nés en France, les natifs d'un Dom et leurs descendants nés dans l'hexagone, ce qui explique ces effectifs importants pour des groupes habituellement peu représentés dans les enquêtes en population générale.

Les catégories de population

Nous avons utilisé dans l'article les catégories d'origine des populations produites pour l'enquête TeO dont on trouve un descriptif détaillé dans le chapitre 1 de l'ouvrage (<https://books.openedition.org/ined/776>).

- **Les immigrés** sont identifiés comme nés étrangers dans un pays étranger (pays de naissance hors de France y compris les Dom et qui ne sont pas Français à la naissance) ; l'origine des immigrés est attribuée à partir du pays d'origine (et non de la nationalité à la naissance).
- **Les descendants d'immigré(s)** sont identifiés comme nés en France d'au moins un parent immigré. Ils peuvent avoir un ou deux parents immigrés (48% des descendants ont un parent français). L'origine des descendants est celle de leur(s) parent(s) immigré(s). Dans le cas où les deux parents sont immigrés et sont nés dans des pays différents (ce qui représente 7% des descendants ayant deux parents immigrés et 4% de l'ensemble des descendants), il a été décidé de choisir le pays du père comme origine du descendant.

Ne sont pas définis comme descendant direct d'immigré les enfants nés à l'étranger et arrivés jeunes en France

- Les **petits-enfants d'immigrés** n'ont pas été identifiés dans les exploitations de l'enquête conduites avant cet article¹ et les décisions de classification ont été les nôtres. Sont considérés comme petits-enfants d'immigré(s) les personnes nées en France dont aucun des parents n'est immigré et dont un moins un parent est descendant d'immigré. Cette opération nécessite de reconstituer les liens familiaux pour construire une base de données renseignant sur l'origine des deux parents et des quatre grands-parents. Il est important de rappeler la clause d'éligibilité liée au statut migratoire des parents, car cela exclut de l'univers des petits enfants d'immigré(s) les personnes dont l'un des parents est immigré même si l'autre parent est descendant d'immigré. Dans ce cas-là, ces personnes sont simultanément petits-enfants d'immigré(s) et descendant direct d'immigré, mais ne sont pas retenues dans notre groupe de petits-enfants d'immigré(s).

Le repérage des petits enfants d'immigrés dans les bases de l'enquête et l'attribution d'une origine

Dans l'enquête TeO nous disposons d'informations sur les parents, le premier conjoint, le conjoint actuel, et les enfants de la personne enquêtée. C'est à partir de ces informations que nous distinguons les petits-enfants d'immigrés. Par exemple, la variable « IMMIP » indique si le père de la personne enquêtée est immigré, « GROUP2CA » donne, pour le conjoint actuel, le statut par rapport à la migration.

Le travail consiste alors à identifier, pour chaque enfant des personnes enquêtées, les caractéristiques de ses père et mère, et celles de ses grands-parents paternels et maternels. La situation la plus simple est celle d'un ménage dans lequel la personne enquêtée cohabite avec ses enfants et l'autre parent de ses enfants, car les liens de parenté sont bien identifiés. Quand les enfants ne cohabitent pas, ou quand les parents se sont séparés, si le lien entre la personne enquêtée et ses enfants est toujours présent, il est plus complexe de bien identifier l'autre parent. On s'appuie donc sur les dates de début et de fin des relations conjugales et sur les dates de naissance des enfants pour identifier l'autre parent.

Si l'un des deux parents est immigré, alors l'enfant est identifié comme un descendant d'immigré. Si les deux parents appartiennent à la population majoritaire, alors l'enfant est identifié comme appartenant à la population majoritaire. Si l'un des parents est descendant d'immigré et si l'autre parent est aussi descendant d'immigré ou appartient à la population majoritaire, alors l'enfant est identifié comme un petit-enfant d'immigrés.

Les 21 767 enquêté.e.s ont généré une base « enfants » (les enfants qui vivent avec eux ou qu'ils ont élevé) avec 31476 observations. Ces enfants se distribuent de la manière suivante dans les catégories de lien à la migration :

¹ A l'exception d'un court cadrage dans le chapitre 1 de l'ouvrage.

statut de l'enfant	nombre
immigré	2454
descendant d'immigré	17264
petit-enfant d'immigré	5515
sans origine immigrée directe	5370
(renseignement précis manquant)	873
Total	31476

L'attribution d'une origine ethno-culturelle pour les petits-enfants d'immigré(s) est plus complexe que pour les descendants d'immigré(s). Le principe est d'attribuer le pays de naissance du ou des grands-parents immigrés comme origine des petits-enfants. Cependant, le nombre et la diversité des origines des grands-parents complique l'allocation d'une origine unique ou principale pour les petits-enfants. Comment arbitrer si plusieurs origines sont combinées, à concurrence de 4 origines différentes potentiellement ?

Seulement 4% des descendants d'immigrés ont nécessité des choix de classification, et le principe de patrilinéarité qui a été suivi pour trancher dans les cas de parents immigrés d'origine différente n'est plus cohérent pour les petits enfants. Le tableau ci-dessous fournit une illustration de ces combinaisons pour le cas des personnes ayant une origine maghrébine : 73% d'entre elles ont au moins 2 grands-parents qui ne sont pas d'origine maghrébine et tout choix conduirait à favoriser une origine sur une autre. La logique patrilinéaire (l'origine du grand-père paternel prévalant sur les autres par exemple) écarte de manière arbitraire 3 autres choix possibles dont l'influence sur l'identité familiale et la vie sociale des personnes ne peut être distinguées valablement.

Nombre de grands-parents immigrés du Maghreb pour les personnes ayant au moins un grand-parent immigré du Maghreb

Nombre grands-parents	Répartition (en %)	Effectifs (non pondérés)
1	38	654
2	35	597
3	8	132
4	20	339
Total	100	1722

Pour les besoins de nos analyses de la transmission des prénoms, nous avons privilégié une approche inclusive : tout petit-enfant ayant au moins un grand-parent immigré d'une origine particulière (du Maghreb ou d'Europe du Sud par exemple), quel que soit leur nombre (de 1 à 4) est considéré comme petit-enfant d'immigré(s) de cette origine (maghrébin ou européen du sud par exemple). De la sorte nous couvrons l'ensemble du groupe plutôt qu'une sélection.

Notre classification a pour conséquence que les mêmes personnes petits-enfants d'immigré(s) sont assignées à plusieurs origines. Cela conduit à des analyses séparées puisque les mêmes individus sont pour beaucoup d'entre eux simultanément d'ascendance maghrébine, d'ascendance française et d'ascendance espagnole pour prendre l'une des combinaisons possibles. De ce fait, quand nous estimons la part des prénoms spécifiques parmi les petits-enfants maghrébins, une partie de ces derniers sont aussi partie prenante du groupe des petits-enfants d'Europe du Sud. Nos « petits-enfants d'immigrés du Maghreb » ont par exemple, pour un cinquième d'entre eux environ, au moins un grand-parent d'Europe du Sud, parfois le grand-père paternel.

Distinguer les Français nés à l'étranger et leurs descendants des immigrés et de leurs descendants

Il n'existe pas de base de sondage satisfaisante pour pouvoir tirer un échantillon représentatif des descendants d'immigré(s) vivant en France, en particulier le recensement ne comporte pas d'information sur l'origine des parents. Aussi l'identification des descendants d'immigrés dans TeO s'est effectuée à partir des bulletins de naissance de personnes recensées où sont connus les pays de naissance de chacun des parents et de la constitution d'une base de sondage *ad hoc* d'où ont été tirés des enquêtés de différentes origines.

Les parents nés à l'étranger ne sont pas tous des immigrés puisqu'une partie d'entre eux peuvent avoir eu la nationalité française à la naissance. Nous les définissons dans TeO comme des français de l'étranger, ce qui recouvre des situations très variées, mais dont une partie très importante concerne les Français vivant dans l'ancien empire colonial, et tout particulièrement les Français d'Algérie qui forment le groupe numérique le plus important parmi ces Français nés à l'étranger. La distinction entre enfant d'immigrés et enfants de Français nés à l'étranger n'est pas identifiable à partir de l'information disponible dans le bulletin de naissance. Pour cette raison, le tirage de l'échantillon de personnes potentiellement descendantes d'immigré(s) de l'enquête TeO a été large pour pouvoir atteindre les objectifs numériques assignés à l'enquête. Les enquêtrices et enquêteurs ont ainsi rencontré des personnes qui avaient des parents nés à l'étranger sans savoir à quelle catégorie de population elles appartenaient, cette information étant pendant la passation du questionnaire, lorsque les questions relatives à l'origine des parents sont posées.

Une difficulté particulière pour établir la qualité de descendant d'immigré concerne les personnes dont les parents sont nés dans des anciennes colonies françaises (c'est-à-dire qui sont nés avant l'indépendance de ces colonies). Le régime de nationalité pratiqué dans les colonies, relativement complexe et qui connaît des variations selon les territoires concernés, introduit une distinction entre la nationalité française attribuée aux indigènes qui ne pouvaient détenir la nationalité de leur pays qui n'existait pas en tant qu'entité indépendante et la citoyenneté française, réservée aux Français de naissance et aux étrangers naturalisés. Dans le cas de l'Algérie, bien que formellement les anciens indigènes comme les colons (et les Juifs d'Algérie ayant été collectivement naturalisés par le décret Crémieux en 1870) disposaient tous de la nationalité française, celle-ci était dépourvue des attributs de la

citoyenneté pour les anciens indigènes qui étaient distingués par leur statut personnel musulman. Il est difficile pour les enfants des anciens sujets coloniaux de répondre à la question de savoir si leurs parents étaient « Français de naissance » avant les indépendances et les réponses sont relativement hétérogènes.

Pour déterminer au mieux la position des personnes nées en dans l'empire colonial, l'enquête TeO a innové et ajouté une question supplémentaire pour les parents français de naissance dans ces colonies libellée comme suit : « Au moment de l'indépendance ou dans les années qui ont suivi, votre père [mère] a 1) pris la nationalité du pays devenu indépendant (ou la double nationalité), 2) est resté français, 3) était déjà décédé, 4) n'était pas encore né, 9) ne sait pas ». Grâce à cette précision il a été possible d'écarter du groupe de descendant d'immigré(s) les personnes qui ont interprété la nationalité à la naissance de leur parent d'une façon formelle en attribuant une nationalité française aux parents qui étaient alors des sujets coloniaux et qui, après l'indépendance, ont pris la nationalité du pays. En toute hypothèse à de très rares exceptions toujours possibles, cette clause a conduit à placer les descendants de Français de l'empire colonial, et tout particulièrement d'Algérie, dans la catégorie des descendants de Français de l'étranger.

Parmi les personnes nées en France dont un parent est né en Algérie, près de la moitié peuvent être considérées comme des descendants de rapatriés (i.e des enfants de Français d'Algérie, parfois d'origine espagnole, italienne ou maltaise, ainsi que de Français musulmans ayant opté pour la nationalité française à l'indépendance).

Les méthodes de classification des prénoms

Plutôt que de nous livrer à une classification manuelle des prénoms à partir de nos connaissances des champs sémantiques dont ils relèvent, qui aurait été inductive et forcément en partie subjective, nous avons confié l'opération à une entreprise spécialisée allemande (Humpert & Schneiderheinze) avec laquelle Patrick Simon avait déjà travaillé pour l'échantillonnage de l'enquête TIES (*The Integration of the European Second generation*) en 2007. L'algorithme utilisé par cette entreprise pour classer les prénoms consiste à utiliser une base de référence comprenant les prénoms et noms d'habitants de plus de 50 pays (comprenant plusieurs millions d'entrées : 37 millions en 2000 mais bien plus aujourd'hui) et de procéder à une procédure d'appariement probabiliste entre les prénoms de la base TeO et cette base de référence. Ces appariements ne fonctionnent pas uniquement par recoupement exact, mais utilisent des inférences fondées sur des schèmes linguistiques pour compléter les informations manquantes ou déviantes. Les catégories utilisées dans les classifications sont géographico-nationales, mais proposent également des regroupements par aires culturelles ou des langues infra ou supra-nationales (voir nomenclatures ci-dessous).

Humpert & Schneiderheinze ont procédé à des corrections sur la liste des prénoms fournis pour supprimer les fautes de frappe, les espaces superflus, les doubles prénoms sans espace, etc, dont on trouvera une synthèse en fin d'annexe. 1831 recodages ont été opérés sur les

196 597 prénoms, aboutissant à identifier 236 enregistrements impropres à la classification et 282 non reconnaissables pour une classification. Plusieurs séries de classification (CAT0, CAT1, CAT2, CAT3) ont été produites, correspondant à des codages successifs de probabilité décroissante. CAT0 est une classification signalétique qui n'est pas utilisable en tant que telle. CAT1 est la classification de référence que nous avons utilisée car elle épuise l'information : CAT2 et CAT3 n'apportent rien de plus.

Nous avons opéré des regroupements sur la base de CAT1 pour aboutir à la classification utilisée dans le cadre de l'article avec le format suivant :

Europe et Balkans=

100, 126, 128, 132, 139, 142, 148, 149, 157, 611, 121, 122, 125, 138, 144, 154, 152, 160, 162, 165, 166, 169, 430

France=**129**

Europe Sud (prénoms latins)= **134, 137, 153, 161, 323, 327, 661**

Prénoms Anglophones et internationaux = **135, 168, 368, 697**

Prénoms Arabo-musulmans = **221, 252, 224, 622, 422, 425, 439, 450, 641, 643, 662**

Afrique sub-saharienne= **231-251, 262-283, 621, 623**

Asie= **427, 431-437, 442, 446, 449, 457-479, 644, 645**

Divers= **441, 651, 696**

Inclassables= **0, 698**

Codes de classification des prénoms :

CAT0 :

CODE	Description
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

CAT1

100	Germany	231	Ivory Coast	462	Philippines
121	Albania	232	Nigeria	467	Korea
122	Bosnia	238	Ghana	476	Thailand
125	Bulgaria	243	Kenia	479	China
126	Danmark	244	Comores	611	Scandinavia
128	Finland	246	Democratic Republic of Congo	621	Africa except for Northern Africa ³
129	France	249	Madagascar	622	Maghreb ⁴
134	Greece	251	Mali	623	West Africa ⁵
135	Ireland	252	Morocco	641	Arabic speaking States ⁶
137	Italy	262	Cameroon	643	Middle East ⁷
138	Serbo-Croat	263	South Africa	644	Tibet
139	Latvia	269	Senegal	645	Far East Indochina ⁸
142	Lithuania	283	Togo	651	Polynesia
144	Republic of Macedonia	323	Argentina	661	Spain, Portugal, Latin America, Philippines, Angola, Mozambique ⁹
148	Netherlands	327	Brazil	662	Muslim without any further classification
149	Norway	368	USA	695	first name typical for a certain country (only CAT0)
152	Poland	422	Armenia		
153	Portugal	425	Azerbaijan		
154	Romania	427	Myanmar		
157	Sweden	430	Georgia		
160	Russia	431	Sri Lanka ²		
161	Spain	432	Vietnam		
162	Czechia	436	India		
163	Turkey	437	Indonesia		
165	Hungary	439	Iran (Farsi names)	696	First names common in many countries (not only European), some very widely used (Maria), some less (Alina)
166	Ukraine	441	Israel		
168	Great Britain (English/Welsh)	442	Japan		
169	White Russia	446	Cambodia		
221	Algeria	449	Laos		
224	Eritrea/Ethiopia (Amharic names)	450	Kyrgyzsan	697	International names
		457	Mongolia	698	unclassified
		461	Pakistan	999	Error (no name)

² Sinhalese and Tamil names

³ Maghreb, Libya, Egypt

⁴ Morocco, Algeria, Tunisia

⁵ Benin, Burkina Faso, Ivory Coast, Gambia, Ghana, Guinea, Guinea-Bissau, Liberia, Mali, Niger, Nigeria, Senegal, Sierra Leone, Togo

⁶ Libya, Egypt, Jordan, Syria, Iraq, Palestine, Arabian Peninsula

⁷ Iran, Afghanistan, Pakistan, Bangladesh, Indian Muslims

⁸ China, Myanmar, Thailand, Vietnam, Cambodia, Laos, Korea

⁹ Spanish, Portuguese, in Latin America also Italian first names

Table 1: Error codes with examples

Error code	description	number	original entry (PRENOM)	corrected
1	letter missing	54	ABDELKDER	Abdelkader
2	wrong letter	40	ABDELOIHAB	Abdelouhab
3	letter too much	81	ALEXXANDRE	Alexandre
4	containing a non-letter	13	WILLIA%M	William
5	letters mixed up	56	ABDELGAHNI	Abdelghani
6	name cut off (>15 characters)	171	FLORENTINE ELEO	Florentine Eleonore
7	blank within the name	37	DAV ID	David
8	containing a number	54	SIVARAJASINGAMI	Sivarajasingam
9	too much blanks within double first name	88	EVELYNE SARAH	Evelyne Sarah
10	blank missing	11	ESPERANCESYLVIE	Esperance Sylvie
11	surname (not corrected)	20	DUROVIC	-
12	umlaut missing	608	AYGUL	Aygül
15	non-name-entry added	4	JOAQUIM (GP)	Joaquim
16	initial added	4	LOUIS LUCIEN B	Louis Lucien
17	apostrophe missing	31	M BAREK	M'Barek
18	name cut off (<15 characters)	4	MARIE PASCA	Marie Pascale
19	other error (not corrected)	21	Z.MINO	-
21	"Jean" not written in full	16	J FRANCOIS	Jean Francois
89	probably an error, but not verifiable	282		
96	character string without meaning	26	FE...	-
97	only number	54	3	-
98	only initial	96	F	-
99	no first name	60	ENFANT 1	-