



**Observatoire National des Discriminations
et de l'Égalité dans le Supérieur**

Rapport d'étude n° 23-04

**L'Indice de Diversité Patronymique :
enjeux, principes et applications**

Moussa KEDDACHE
(UGE, ERUDITE et TEPP)

Yannick L'HORTY
(UGE, ERUDITE et TEPP)

Décembre 2023



L'Indice de Diversité Patronymique : enjeux, principes et applications

Décembre 2023

Moussa KHEDDACHE¹, Yannick L'HORTY²

Résumé

Dans le contexte règlementaire français de prohibition des statistiques ethniques, il est pratiquement impossible de mesurer la diversité d'origine à l'échelle d'une entreprise, d'un établissement public ou privé, d'un secteur d'activité ou de tout autre groupement d'individus, qu'il soit de petite ou de grande taille. Il est donc hors de portée de produire des connaissances générales sur l'état de la diversité d'origine, son évolution dans le temps, sa géographie, ses liens avec les performances des organisations ou avec toute autre grandeur.

Cette note propose une méthode simple permettant de calculer un Indice de Diversité Patronymique (IDP) afin d'objectiver la diversité d'origine à l'échelle d'un groupement donné d'individus. L'indice proposé correspond au pourcentage de personnes dont le nom de famille n'est apparu de façon significative dans les registres de l'état civil qu'à partir de 1950. L'étude discute les avantages et inconvénients d'une telle approche et propose plusieurs applications afin de contribuer à la diffusion d'une mesure objective de la diversité d'origine à l'échelle des organisations.

Mots-clés : Diversité, discrimination, onomastique

JEL Codes: C67, H52, I23

Cette étude a bénéficié du soutien de l'ANR (Projet n°ANR-22-CE41-0018-01 ESTRADES) et de l'appui de l'Observatoire National des Discriminations et de l'Égalité dans le Supérieur (ONDES). Au moment de sa rédaction, Moussa Kheddache effectuait un stage de licence 3 au sein de la fédération TEPP et du laboratoire ERUDITE à l'université Gustave Eiffel. Nous remercions Lionel Espinasse et Patrick Simon pour leurs avis sur une première version de l'étude.

¹ Université Gustave Eiffel, Université Paris Est Créteil, ERUDITE, TEPP, F-77454, Marne-La-Vallée, France, moussakheddache@gmail.com

² Université Gustave Eiffel, Université Paris Est Créteil, ERUDITE, TEPP, F-77454, Marne-La-Vallée, France, yannick.lhorty@univ-eiffel.fr

Introduction

La situation de la France est assez paradoxale du point de vue de la mesure de la diversité d'origine et des discriminations. D'un côté, il existe de longue date un assez large consensus parmi les acteurs privés et publics sur la nécessité de promouvoir la diversité sociale, géographique ou culturelle sous toutes ses formes et de lutter contre les discriminations sur le marché du travail, sur celui du logement et dans tous les domaines. Une illustration en a été donnée avec la présentation fin janvier 2023 du Plan national de lutte contre le racisme, l'antisémitisme et les discriminations selon l'origine, pour la période 2023-2026, qui a associé onze ministres du gouvernement et autant de grands partenaires associatifs autour de la Première ministre, Mme Elisabeth Borne. D'un autre côté, contrairement à la plupart des autres pays, il est en pratique impossible de mesurer statistiquement la diversité à l'échelle des entreprises ou des administrations en France, ce qui rend caduque toute comparaison dans le temps et dans l'espace et, par contrecoup, toute possibilité de suivi et de pilotage dans ces organisations. Comment promouvoir la diversité si elle est impossible à mesurer ?

Les obstacles à la mesure de la diversité en France sont bien connus. Le cadre légal prohibe la production de données s'appuyant sur un référentiel ethno-racial, jugé incompatible avec le principe constitutionnel d'égalité « de tous les citoyens sans distinction d'origine, de race ou de religion » (premier alinéa de l'article premier de la constitution de 1958). Selon la Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, « il est interdit de collecter ou de traiter des données à caractère personnel qui font apparaître, directement ou indirectement, les origines raciales ou ethniques (...) ». Pour autant, comme le rappelait le rapport du *Comité pour la mesure de la diversité et des discriminations* (Héran, 2010), s'il est proscrit d'utiliser des référentiels de race supposée, de couleur de peau ou de sentiment religieux, il est envisageable dans le cadre de la statistique publique d'effectuer des traitements de données subjectives portant sur l'identité d'appartenance ressentie et de traiter des données objectives portant sur le nom, l'origine géographique et la nationalité.

Dans cette note, nous proposons un nouvel indice que nous avons appelé « Indice de Diversité Patronymique » (IDP). Il consiste à mesurer la diversité par la proportion de personnes dont le patronyme était absent des registres de l'état civil avant 1950. Même si la France est depuis très longtemps un pays de migration, la troisième vague historique d'immigration qui débute

autour des années 1950 est allée de pair avec une prépondérance des flux en provenance d’Afrique du nord et d’Afrique subsaharienne dans le contexte de la décolonisation et des importants besoins en main d’œuvre associés aux Trente Glorieuses. Les personnes issues de ces vagues migratoires, soit directement, soit indirectement par leurs ascendants, constituent en France l’essentiel des populations potentiellement discriminées par leur origine. Il s’agit de mesurer la diversité par la proportion des personnes dont les noms de famille signalent ce passé migratoire.

Contrairement aux travaux internationaux qui ont recours à une approche onomastique (Matéos, 2007 ; Mazières et Roth, 2018), nous n’utilisons aucun référentiel ethno-racial ce qui est en cohérence avec le cadre juridique français de prohibition des statistiques ethniques. La méthode que nous proposons s’inscrit de surcroît dans la tradition française de mesure de la diversité par l’origine au travers de l’ascendance démographique des personnes, signalée par leur nom de famille. Elle nous paraît néanmoins comparable dans ses performances heuristiques aux indices de diversité qui relèvent de la tradition anglo-saxonne d’auto-déclaration d’une origine ethno-raciale. Par ailleurs, l’IDP ne laisse aucune place à la subjectivité des acteurs contrairement aux approches par auto-déclaration. Surtout, il peut être calculé de façon standardisée à grande échelle dans l’ensemble des organisations privées ou publiques, à une fréquence régulière et de façon quasi-automatique tant à partir de données de stocks (listes des membres de telle ou telle entité) qu’à partir de données de flux (issues d’opérations de recrutement, d’examens ou de concours). Enfin, contrairement à de nombreuses procédures alternatives, l’IDP peut être calculé rapidement pour un coût presque nul.

La première section revient sur les principaux enjeux de la mesure de la diversité d’origine. La deuxième section présente les principes et la méthode de calcul de l’IDP. La troisième section discute les avantages et inconvénients de l’indice. La quatrième section propose plusieurs applications de l’IDP.

1. Les enjeux de la mesure de la diversité

1.1. Enjeux de dignité pour les personnes et de connaissance pour les sciences sociales

A défaut de pouvoir mesurer la diversité, c'est l'entrée par les discriminations qui prévaut dans la recherche en sciences sociales. Les discriminations constituent ainsi à la fois un délit et un objet de recherche. D'un point de vue juridique, une discrimination est une différence préjudiciable de traitement fondée sur un critère prohibé dans un domaine donné, tel que l'accès à l'emploi ou au logement. Lorsqu'il est constitué, le délit de discrimination fait encourir une peine maximale de trois ans d'emprisonnement et de 45 000 € d'amendes s'il est commis par une personne physique, portée à 225 000 € s'il relève d'une personne morale. Le code pénal énumère dans son article 225-1 près de 25 critères illicites³. Parmi tous les critères prohibés, près de la moitié sont étroitement associés à la diversité ethno-raciale : l'origine, l'apparence physique, la particulière vulnérabilité résultant de leur situation économique, le patronyme, le lieu de résidence, la capacité à s'exprimer dans une langue autre que le français, l'appartenance ou de la non-appartenance, vraie ou supposée, à une ethnie, une Nation, une prétendue race ou une religion déterminée.

Du point de vue de la recherche en sciences sociales, le caractère délictueux n'est pas prédominant dans l'intérêt porté aux discriminations. Il ne s'agit pas davantage de dénoncer le fait qu'elles contredisent le principe d'égalité et portent atteinte à la fois aux fondements du pacte républicain et à la dignité des personnes. Tout cela est vrai mais ne détermine pas l'intérêt des chercheurs et des chercheuses pour les discriminations, qui recouvre avant tout un enjeu de connaissance. Pour le sociologue, l'expérience des discriminations est une épreuve individuelle qui recouvre des enjeux collectifs. Elle nous renseigne sur ce que nous avons de commun, sur ce qui nous divise, sur notre manière de vivre ensemble et de faire société (Dubet, 2014). L'entrée par les discriminations est tout simplement une autre façon

³ « Constitue une discrimination toute distinction opérée entre les personnes physiques sur le fondement de leur origine, de leur sexe, de leur situation de famille, de leur grossesse, de leur apparence physique, de la particulière vulnérabilité résultant de leur situation économique, apparente ou connue de son auteur, de leur patronyme, de leur lieu de résidence, de leur état de santé, de leur perte d'autonomie, de leur handicap, de leurs caractéristiques génétiques, de leurs mœurs, de leur orientation sexuelle, de leur identité de genre, de leur âge, de leurs opinions politiques, de leurs activités syndicales, de leur capacité à s'exprimer dans une langue autre que le français, de leur appartenance ou de leur non-appartenance, vraie ou supposée, à une ethnie, une Nation, une prétendue race ou une religion déterminée » (Art. 225-1, code pénal).

d'étudier les sociétés contemporaines et leurs transformations, en se concentrant sur les processus de mise à l'écart de certains groupes sociaux. L'introduction de la catégorie conceptuelle de discrimination dans la sociologie française est d'ailleurs venue stimuler l'étude de rapports de pouvoir, en particulier ceux liés au genre et à l'ethnicité, jusque-là peu pris en compte dans le questionnement sur les inégalités. Dans un contexte intellectuel français dominé par l'analyse des inégalités de classe sociale, les travaux sociologiques sur les discriminations ont ainsi enrichi les perspectives. De plus, en plaçant la focale sur les processus discriminatoires, les recherches sur les discriminations ont aussi contribué à mieux comprendre la manière dont les inégalités sont produites (Bereni et Chappe, 2011). Ces recherches ont permis de dépasser la seule étude des discours et des idéologies qui avait longtemps prévalu dans les analyses du racisme. Elles se sont intéressées de près aux pratiques et aux interactions, même les plus routinières, dont le caractère répété et cumulé a de puissants effets inégalitaires.

L'enjeu de connaissance est tout aussi saillant pour l'économiste. Son intérêt pour les discriminations vient du fait qu'elles paraissent relever a priori d'un comportement irrationnel et d'une anomalie du marché. Les offreurs de biens et services qui les pratiquent se privent d'une source de profit et sont moins efficaces que ceux qui traitent la clientèle ou les candidats à l'embauche dans l'égalité. Selon le modèle séminal de Becker (1957), la discrimination ne devrait pas exister dans un contexte parfaitement concurrentiel. Mesurer et comprendre les discriminations constitue alors un moyen de fixer les limites de la représentation classique où des *homo economicus* échangent sur des marchés de concurrence pure et parfaite. C'est sans doute la raison pour laquelle l'économie des discriminations constitue aujourd'hui un chapitre à part entière de la nouvelle économie comportementale qui remet en question les hypothèses fondatrices de la théorie économique dominante (Gabuthy, Jacquemet et L'Haridon, 2021).

Si l'intérêt d'un nombre croissant de scientifiques de toutes disciplines est bien réel, il convient néanmoins de noter les différences, voire les retards, qui séparent la situation de la France de celles des pays disposant de référentiels ethno-raciaux auto-déclarés, dans la tradition anglo-saxonne. Aux Etats-Unis, mais aussi au Canada, au Royaume-Uni et dans de nombreux autres pays, les questionnaires des recensements invitent les répondants à se classer dans un ou plusieurs groupes pré-construits, dont la définition est très variable d'un pays à l'autre. Les

modalités combinent de façon apparemment hétéroclite des nationalités (Indien ou Pakistanais en Grande-Bretagne, Japonais au Canada), des couleurs de peau (Blanc, Noir, etc., au Canada), des origines ethno-culturelles (« Arabe » au Canada, Hispanic/Latino aux États-Unis), des origines régionales (« Caribéens » en Grande-Bretagne, « Sud-Asiatique » au Canada). Comme l'explique le rapport du COMEDD qui cite ces exemples, « l'auto-identification de l'appartenance ethno-raciale est avant tout une statistique bricolée qui a pour objectif de se situer au plus près des représentations sociales et des identifications significatives pour les individus » (Héran, 2010). Les nomenclatures sont étroitement liées à l'histoire nationale et au passé colonial et/ou esclavagiste propre à chaque pays, à l'historique des migrations et des minorités formant « des communautés ethniques plus ou moins construites ou imaginées ».

Ces référentiels auto-déclarés ont néanmoins rendu possible le développement d'une abondante littérature consacrée aux liens entre les performances économiques et la diversité ethno-raciale, mesurée par exemple à l'aide de l'indice de Blau (1977) qui correspond à la probabilité que deux individus pris au hasard appartiennent à des groupes ethniques différents. Cette littérature n'associe pas toujours positivement les deux ensembles de variables. Pour certains auteurs, plus une société est divisée sur le plan ethnique, moins elle sera en mesure de fournir des services publics et des infrastructures collectives et plus grand sera le risque de conflit entre communautés, divisées par leur origine ethnique (voir Alesina et al., 2003, et Fearon, 2003, sur les questions de mesure, et Alesina et La Ferrara, 2005, pour un exemple d'étude).

1.2. Enjeux de mesure statistique

Le recensement de la population et de nombreuses grandes enquêtes de la statistique publique permettent d'identifier la nationalité des individus, leur lieu de naissance et celui de leurs parents voire de leurs grands-parents. Une fois les individus ainsi identifiés, ces sources permettent de documenter des inégalités de situation entre groupes d'individus, en matière d'emploi, de revenu ou de conditions de vie. Il s'agit d'informations à la fois précieuses et insuffisantes pour approcher les questions de diversité ethno-raciale en France. Elles sont insuffisantes parce que ces sources sont dans l'incapacité d'approcher les questions de discriminations selon la couleur de la peau ou l'identité culturelle.

Une mesure directe des discriminations suppose de pouvoir observer des différences de traitement entre individus dans l'accès à un marché ou à une ressource, sur la base d'un critère prohibé. Même lorsque l'on dispose de données de candidatures, comme c'est le cas dans les travaux de recherche sur les discriminations dans la réussite aux concours de la fonction publique (Bréda et Hillion, 2016 ; Greenan *et al.* 2019), il est très difficile de réaliser des analyses toutes choses égales par ailleurs, d'autant que ces données ne couvrent pas tous les critères prohibés. De fait, la plupart des critères illicites qui sont en association étroite avec l'origine sont absents des bases de données existantes (c'est le cas de la couleur de la peau et de l'apparence physique ou encore du sentiment religieux).

Ce constat est assez ancien et il avait été posé en 2009 comme point de départ à la constitution par Yazid Sabeg, commissaire à la diversité et à l'égalité des chances, du comité pour la mesure de la diversité et des discriminations (Comedd), présidé par François Héran, alors directeur de l'INED. Le rapport du COMEDD, ou rapport Héran (2010), est intitulé « Inégalités et discriminations - Pour un usage critique et responsable de l'outil statistique : rapport du comité pour la mesure de la diversité et l'évaluation des discriminations ». Il n'a guère perdu selon nous de son acuité et de son actualité.

Le rapport du COMEDD souligne notamment que la statistique publique qui organise les opérations de recensement intègre d'ores et déjà une dimension ethnique, entendu au sens faible de « trace du pays d'origine », en repérant les nationalités d'origine et les lieux de naissance des personnes. Cette entrée permet de documenter avec précision les conditions de vie des immigré·es et de leurs descendant·es, les deuxièmes voire troisièmes générations. La statistique publique française est certes « aveugle à la couleur », mais elle n'est pas « aveugle à l'origine » et sa capacité de repérage des ascendances peut être présentée comme un substitut à la tradition anglo-saxonne d'identification raciale auto-déclarée par les personnes.

Les données des grandes enquêtes de la statistique publique qui portent sur des échantillons représentatifs, telles que les enquêtes TeO développées par l'INED et l'INSEE (Beauchemin *et al.*, 2016 ; 2023), apportent ainsi des éléments précieux à la connaissance de la diversité et au ressenti des discriminations. La comparaison des résultats de TeO1 et TeO2 montre en particulier que le sentiment de discrimination augmente entre 2008-2009 et 2018-2019 : près d'une personne sur cinq déclare avoir subi « des traitements inégaux ou des

discriminations », contre 14 % dix ans plus tôt. À caractéristiques égales, l'origine immigrée ou ultra-marine constitue le principal déterminant des expériences déclarées de discrimination (Lê *et al.*, 2022). Mais ces enquêtes documentent le ressenti des personnes discriminées qui ne sont pas toujours les mieux placées pour être témoins des discriminations. Dans le cadre d'un recrutement par exemple, un candidat non retenu peut penser avoir été discriminé alors qu'il n'a pas été sélectionné pour d'autres raisons. Lorsque les personnes sont effectivement discriminées, elles peuvent ne pas souhaiter se signaler en tant que telles et même si elles se signalent, le sentiment de discrimination n'est pas toujours une mesure objective de discrimination, en particulier parce qu'un·e discriminé·e potentiel·le n'est pas en position de pouvoir raisonner « toutes choses égales ». Enfin, les données issues d'enquêtes ont l'immense avantage d'être représentatives de la population générale mais elles ne sont pas désagrégeables au niveau d'un territoire ou d'une organisation donnée et elles ne permettent donc pas de documenter la diversité à ce niveau d'observation.

Dans la littérature internationale sur les discriminations, la méthode de référence est celle du test par correspondance, couramment appelé en France « testing », qui consiste à comparer les chances de succès de deux candidatures fictives, semblables en tous points à l'exception de la caractéristique dont on veut quantifier l'effet (Riach et Rich, 2002 ; Rooth, 2014 ; Bertrand et Duflo, 2017 ; Zschirnt et Ruedin, 2016 ; Quillian *et al.*, 2017 ; Neumark, 2018). Appliquée au marché du travail, cette approche consiste à sélectionner un petit nombre de professions et à fabriquer de toutes pièces des candidatures (CV et lettres de motivation), similaires en tout point, à l'exception de la caractéristique dont on souhaite examiner l'effet sur l'accès à un entretien d'embauche. Ensuite, ces candidatures fictives sont envoyées en réponse à plusieurs centaines d'offres d'emploi disponibles et correspondant aux profils construits, à l'insu des responsables de recrutement qui doivent ignorer faire l'objet d'une expérience. La comparaison des réponses reçues par les candidats fictifs permet alors de mettre en évidence l'existence de discriminations dans l'accès aux entretiens d'embauche. La méta-analyse de Lippens *et al.* (2021) couvre la quasi-totalité des tests de correspondance portant sur le marché du travail et ayant fait l'objet d'une publication académique en anglais entre 2005 et 2020. La moitié de ces études s'intéresse à la question de l'origine ou de la couleur de la peau et met en évidence l'existence de discriminations à l'encontre des minorités. Cette approche est utilisée en France depuis deux décennies pour mettre en

évidence l'existence de discriminations à l'embauche (Duguet et Petit, 2005 ; Duguet *et al.*, 2010 ; du Parquet et Petit, 2019). Un survol extensif des vingt études qui ont mesuré les discriminations à raison de l'origine en France, dans l'accès à l'emploi, souligne l'intensité, la généralité et la permanence des différences de traitement selon l'origine des personnes en France, faisant des discriminations ethno-raciales un véritable défi pour l'action publique (L'Horty et Petit, 2023).

Cette approche permet de produire une statistique au niveau d'une profession. Elle peut être répliquée pour obtenir des données couvrant un petit nombre de professions. Mais elle n'est ni représentative de la situation du marché du travail dans son ensemble, ni représentative de la situation d'une entreprise en particulier, fût-elle de grande taille. Appliquée à une entreprise unique, elle entraînerait un risque de détection trop important. Or le risque de détection est une problématique centrale lorsqu'on réalise des tests répétés sur les mêmes entreprises. A notre connaissance, une seule étude est parvenue à produire des statistiques représentatives au niveau entreprise sur un grand nombre d'entreprises (103 au total) à l'aide d'une approche multicanale combinant des tests en réponses à des offres et sous forme de candidatures spontanées (Challe et al. 2020).

Une autre limite importante du testing est qu'il ne permet au mieux de détecter des discriminations qu'à l'entrée dans les entreprises. Or le risque de discriminations existe aussi dans le déroulement des carrières au sein des organisations publiques ou privées, comme peut l'illustrer l'immense littérature sur les écarts de salaires entre les femmes et les hommes. Les tests par correspondance mesurent le rendement de la diversité dans les flux entrants de ressources humaines et ne documentent pas la situation du point de vue du stock des effectifs dans les organisations.

1.3. Enjeux de sécurité juridique

La mesure de la diversité à l'échelle micro-économique d'une entreprise pose en outre une question de sécurité juridique. Les données personnelles qui permettent de documenter les critères prohibés sont des données sensibles qui ne peuvent être construites qu'avec un grand nombre de précautions méthodologiques. C'est pourquoi, comme nous l'avons déjà mentionné, la Loi Informatique et libertés de 1978 « *interdit de collecter ou de traiter des données à caractère personnel qui font apparaître, directement ou indirectement, les origines*

raciales ou ethniques (...) ». Cette disposition a été reprise par le Règlement Général sur la Protection des Données (RGPD) qui encadre le traitement des données personnelles sur le territoire de l'Union européenne.

Toutefois, l'interdiction n'est pas totale et des possibilités existent. Dans un rapport de 2007, la CNIL s'est prononcée contre la production de statistiques construites à partir d'une nomenclature de catégories « ethno-raciales » tout en admettant la possibilité, dans le cadre de la statistique publique, d'études sur le ressenti des discriminations pouvant inclure des données sur l'apparence physique. De même, elle considère que l'analyse des prénoms, patronymes, nationalités et lieux de naissance des ascendants peut être utilisée dans certaines conditions pour tenter de révéler des pratiques discriminatoires.

Dans sa décision de conformité 2007-557 du 15 novembre 2007, relative à la Loi sur la maîtrise de l'immigration, à l'intégration et à l'asile, le Conseil constitutionnel a estimé que l'on ne pouvait méconnaître le principe énoncé par l'article 1^{er} de la Constitution, qui dispose que « *La France est une République indivisible, laïque, démocratique et sociale. Elle assure l'égalité devant la loi de tous les citoyens sans distinction d'origine, de race ou de religion* ». La définition, a priori, d'un référentiel ethno-racial serait contraire à la Constitution. Pour autant, il est parfaitement possible de traiter des données objectives, fondées sur le nom, l'origine géographique ou la nationalité antérieure à la nationalité française. Il en va de même pour des données subjectives, par exemple celles fondées sur le ressenti d'appartenance. L'indice que nous proposons repose sur la construction d'une variable binaire indiquant si le patronyme de la personne existait en France avant 1950. Il s'agit d'une donnée objective fondée sur le nom qui ne correspond en rien à un référentiel ethno-racial.

1.4. Enjeux pour l'action publique et privée

En France, la promotion de la diversité et la lutte contre les discriminations recouvrent également des enjeux du point de vue de l'action publique. Depuis la fin des années quatre-vingt-dix, la diffusion tardive du thème des discriminations dans les débats publics va progressivement modifier les perspectives antérieures centrées sur l'intégration des immigrés. Fassin (2002), dans son article consacré à « l'invention française de la discrimination », fait remonter symboliquement la rupture à l'année 1998 avec la publication par le Haut Conseil à l'intégration de son rapport intitulé « Lutte contre les discriminations :

faire respecter le principe d'égalité ». Au début des années 2000, les premiers changements du cadre législatif français vont répondre aux demandes de la Commission européenne de transposition de directives adoptées par le Conseil de l'Europe. En réponse à cette impulsion, la Loi n°2001-1066 du 16 novembre 2001 relative à la lutte contre les discriminations modifie le code du travail pour y introduire le délit de discrimination tandis que la loi n°2004-1486 du 30 décembre 2004 crée la Haute Autorité de Lutte contre les Discriminations et pour l'Égalité (HALDE), autorité administrative indépendante à laquelle le Défenseur des Droits succédera en 2011. Par la suite, plusieurs lois successives vont élargir la liste des critères prohibés. De façon parallèle, la fonction de Délégué interministériel à la lutte contre le racisme et l'antisémitisme (DILCRA) est créée en 2012 en vue d'animer le comité interministériel dont l'origine remonte à 2003. A partir de 2014, le ou la déléguée est rattachée au premier ministre et non au ministre de l'intérieur. Fin 2016, son champs d'intervention est étendu à la lutte contre la haine et les discriminations anti-LGBT. La DILCRAH pilote les plans triennaux qui sont déployés depuis 2012 et dont la dernière incarnation est le plan national de lutte contre le racisme, l'antisémitisme et les discriminations selon l'origine, pour la période 2023-2026.

Dans ce nouveau contexte institutionnel, les actions de promotion de la diversité se multiplient dans les entreprises. Dès le début des années 2000, elles signent des chartes, recrutent des responsables diversité puis s'engagent dans le label éponyme créé par l'État en 2008. Un petit nombre de dirigeants de grandes entreprises, tels Claude Bébéar et Yazid Sabeg, vont substituer à la dénonciation morale et juridique des discriminations un discours positif de légitimation de la diversité dans l'intérêt même des performances des entreprises (Bereni, 2009). L'idée d'une causalité de la diversité vers la productivité, très débattue par la recherche, fait l'objet d'un large consensus parmi les acteurs et est aussi partagée par le Gouvernement. En février 2022, Élisabeth Borne, alors ministre du Travail, de l'Emploi et de l'Insertion, l'exprimait clairement en déclarant « J'en suis convaincue, la diversité est une source de performance économique et sociale : c'est en permettant à des profils différents et complémentaires de travailler ensemble que se créent les dynamiques d'équipe, les synergies, les idées nouvelles ».

A la suite des attentats de janvier 2015, un comité interministériel à l'égalité et à la citoyenneté est constitué, dont les travaux vont être traduits par la loi n° 2017-86 du 27 janvier 2017 relative à l'égalité et à la citoyenneté. Elle prévoit notamment que « le Gouvernement

publie un rapport biennal sur la lutte contre les discriminations et la prise en compte de la diversité de la société française dans la fonction publique de l'État, la fonction publique territoriale et la fonction publique hospitalière. » (art. 158).

Si, depuis 1974, le gouvernement comportait de manière intermittente une ministre chargée de l'Égalité entre les femmes et les hommes ou des Droits des femmes, c'est en octobre 2018 que la secrétaire d'État chargée de l'Égalité entre les femmes et les hommes, Marlène Schiappa, voit son portefeuille élargi à la Lutte contre les discriminations. Puis, lorsqu'Elisabeth Moreno lui succède à partir de juillet 2020, l'intitulé de sa fonction au sein du gouvernement devient celui de ministre déléguée chargée de l'Égalité entre les femmes et les hommes, de la Diversité et de l'Égalité des chances. A partir de mai 2022, Isabelle Lonvis-Rome reprend cet intitulé jusqu'au remaniement de juillet 2023 où elle est remplacée par Bérangère Couillard sur le même périmètre fonctionnel. Depuis 2020, la diversité est donc incluse explicitement dans le champ d'intervention d'un secrétariat d'Etat ou d'un ministère.

L'expansion de l'action publique en faveur de la diversité nécessite de pouvoir mesurer les progrès dans ce domaine. A cette fin, un index « Diversité et Inclusion » est expérimenté par le gouvernement à partir en novembre 2021. Il repose sur une simple enquête par questionnaire individuel interrogeant les personnes volontaires dans des entreprises et administrations expérimentatrices, elles-mêmes volontaires. Le questionnaire comporte des interrogations sur la nationalité, la langue maternelle, sur les lieux de naissance et la catégorie socio-professionnelle des parents. Ce type d'indice déclaratif paraît conforme aux recommandations de la CNIL et aux exigences légales fixées par le RGPD en respectant strictement des règles de volontariat, d'anonymat ainsi que de confidentialité des données. Toutefois, l'enquête n'introduit pas de catégorie plus fine d'origine que celle des étrangers hors Union européenne. Elle contient des questions subjectives sur l'importance jouée par les origines sans distinguer entre les origines sociales, géographiques ou culturelles.

L'enquête a été administrée dans neuf très grandes organisations expérimentatrices, parmi lesquelles figurent Coca Cola ou Pôle Emploi. Cette expérimentation, inédite en France, a permis de soumettre le questionnaire de l'Index diversité et inclusion à près de 122 000 salariés et d'obtenir 38 400 réponses. Comme en témoigne ce nombre très élevé de questionnaires, l'index « Diversité et Inclusion » est davantage par sa construction un outil de promotion interne des actions en faveur de l'égalité et de la diversité qu'un véritable outil de

connaissance sur la diversité. Une enquête par questionnaires auprès de personnes volontaires dans des organisations volontaires introduit d'importants biais de sélection à la fois au niveau des organisations et des individus qui ne sont pas de nature à être contrôlés par les évaluateurs. Les réponses ne sont pas comparables d'une organisation à l'autre et elles ne permettent pas d'étudier de façon satisfaisante les liens entre la diversité et tout autre variable d'intérêt. De surcroît, l'expérimentation ne paraît pas en mesure d'être généralisée à des entreprises de plus petite taille.

2. Principes et mode de calcul

Il nous semble qu'un Indice de Diversité Patronymique peut apporter une réponse adaptée à ces différents enjeux. Il s'agit de fournir un instrument de mesure de la diversité qui puisse être déployé à une échelle à la fois macro ou micro-économique, qui puisse être utilisé par les chercheurs lorsqu'ils s'interrogent sur les effets de la diversité et qu'ils souhaitent les étudier de façon rigoureuse et objective, qui soit respectueux du cadre juridique autorisant la mesure de la diversité d'origine mais prohibant l'usage de référentiels ethno-raciaux, et qui puisse répondre également aux besoins de mesure des acteurs publics et privés en vue de promouvoir la diversité et lutter contre les discriminations.

Il existe une littérature internationale sur la mesure de la diversité à l'aide d'approches onomastiques. Dans un survol extensif, Matéos (207) recense 13 études qui utilisent des méthodologies robustes et comparables. Ces études ont recours à des référentiels ethno-raciaux qui distinguent plusieurs origines (asiatique, slaves, africaine, arabe, etc.) et qui sont construits de façon automatique par des méthodes d'analyses de données appliquées à des bases d'apprentissage. L'auteur met en avant la faible précision de ces études dont la sensibilité (fréquence des vrais positifs) varie entre 67 % et 95 % et la spécificité (fréquence des vrais négatifs) s'étend de 80 à 100 %. Il préconise d'élargir les bases d'apprentissage de façon à améliorer la précision de ces approches. Une illustration récente est donnée par l'étude de Mazières et Roth (2018) qui utilise comme base d'apprentissage une large banque de données de publications scientifiques, recensant à la fois les noms d'auteurs et les pays où ils travaillent (la base PubMed). Cette approche leur permet de construire un référentiel en sept catégories et de prévoir avec raison une origine nord-européenne dans 62,2 % des cas.

Relativement à ces travaux, l'approche que nous proposons a pour avantage de reposer sur une base d'apprentissage quasi-exhaustive (les registres de l'état civil), de produire des indicateurs très précis et de ne pas déployer de référentiel ethno-racial. En contrepartie, sa limite est d'être spécifique à la France et de ne pas distinguer les différentes origines au sein de la population issue de la diversité. L'idée est de comparer une mesure de la diversité en population générale à une mesure de la diversité calculée de la même façon pour une population particulière. Dans les deux cas, l'indice consiste à mesurer le pourcentage de personnes dont le nom de famille n'est apparu de façon significative dans les registres des naissances de l'état civil qu'à partir des années 1950. L'indice est un scalaire, qui prend la forme d'un pourcentage dont la lecture est simple : la part des personnes dont le nom de famille n'était pas répandu en France avant la décennie 1950.

L'indice utilise uniquement les noms de famille et aucune information personnelle. Il utilise les données du registre d'état civil pour distinguer au sein de l'ensemble des noms de famille, ceux qui étaient déjà présents avant 1950 et ceux qui ne sont apparus qu'à partir de la décennie 1950. La décennie 1950 est retenue afin d'isoler les personnes issues de la troisième grande vague migratoire dans le contexte de l'expansion économique des Trente Glorieuses et de la décolonisation. Des correctifs sont apportés au calcul de l'indice brut afin de se prémunir contre plusieurs biais éventuels. Nous développons successivement chacun de ces points.

2.1. L'IDP utilise les noms de familles et rien d'autre

Pour calculer l'IDP au niveau d'un groupe donné d'individus, le seul prérequis est de disposer d'une liste de noms de famille, assortie de leurs effectifs au sein du groupe. Aucune autre information n'est nécessaire. Les noms de familles constituent un signal évident de l'identité des ascendants de chaque individu et c'est l'information contenue dans ce signal qui est uniquement exploitée par l'indice.

Il ne s'agit pas d'utiliser les noms de famille à des fins d'enquêtes, afin d'éclairer éventuellement des inégalités de situation. Cet objectif serait impossible à réaliser dans la mesure où l'on ne dispose d'aucun élément identifiant permettant d'administrer une enquête, mis à part les patronymes et leurs fréquences. Une liste de noms de famille,

lorsqu'elle est réduite à une simple liste sans autre indication sur les personnes, ne permet pas de ré-identifier les personnes, même s'il s'agit de données nominatives. En outre, on peut mettre en question le caractère personnel de ces données, s'agissant de noms de famille, par nature collectifs. Les données ne relèvent donc pas du RGPD.

Un avantage de la méthode patronymique réside dans le fait de travailler sur des données déjà collectées pour d'autres finalités. Les noms de famille, comme les prénoms, le sexe ou l'âge, figurent dans la plupart des fichiers de gestion et des sources administratives. Par conséquent, le chercheur n'a pas à les collecter. D'où le faible coût de cette approche comparée à celui des grandes enquêtes par questionnaire.

La méthode n'utilise que les noms de famille et échappe aux limites des méthodes utilisant les prénoms, caractérisées par une grande instabilité des règles de transmission entre les générations et/ou les groupes d'origine, qui peuvent par ailleurs être influencés par les trajectoires sociales et individuelles d'intégration. En France, Coulmont et Simon (2019) mettent en évidence un mouvement de convergence entre population majoritaire et descendants d'immigrés en matière de choix des prénoms.

Par ailleurs, les patronymes ne font l'objet d'aucune classification manuelle ou semi-automatique par grande aire géographico-culturelle, qui laisse une marge variable à la subjectivité des codeurs et peut correspondre à une forme d'assignation identitaire. La classification que nous opérons est au contraire totalement automatique et ne laisse aucune prise à la subjectivité du codeur et à des a-priori identitaires.

Le procédé peut même paraître relativement grossier puisqu'il résume la diversité en une variable binaire distinguant les patronymes présents avant 1950 dans les registres de l'état civil des patronymes issus des migrations postérieures aux années 1950. Cette catégorisation est à l'évidence réductrice. Mais l'objectif est légitime puisqu'il s'agit de promouvoir la lutte contre les discriminations. L'idée est de reconstituer les catégories de perception et d'action des discriminants. Si ces catégories sont grossières (« arabes », « africain », « européen », etc.), voire binaires (« français » / « étrangers », « discriminés » / « non discriminés », etc.), il est pertinent que l'outil de catégorisation partage les mêmes propriétés.

2.2. Source des données : les fichiers des noms issus des registres de l'état civil

Les données patronymiques de la population générale sont issues du fichier des noms de famille de 1891 à 2000⁴ mis en ligne par l'INSEE en mai 2018 et dont les données sont issues du répertoire national d'identification des personnes physiques (le RNIPP). Le RNIPP est l'image des registres d'état civil qui est une source administrative ayant un caractère exhaustif : toute naissance survenue sur le territoire français fait l'objet d'une déclaration à l'état civil. Cette déclaration doit être faite dans les trois jours suivant l'accouchement, dans la commune de l'accouchement. Les renseignements sont demandés au déclarant, et chaque fois qu'il est possible, contrôlés d'après le livret de famille. Dans chaque commune, l'officier d'état civil enregistre tout au long de l'année les événements qui affectent le registre de l'état-civil en remplissant un questionnaire dit « Bulletin statistique de l'état civil » chaque fois qu'il dresse un acte sur ses registres. Il existe un bulletin par événement d'état civil : transcription, mention en marge, mariage, naissance, enfant sans vie, décès, ainsi que diverses précisions sur chaque catégorie de bulletin. Le fichier des noms est établi sur la base des bulletins de naissance des registres communaux d'état civil.

Le champ du fichier des noms diffusés par l'INSEE couvre l'ensemble des personnes nées entre 1891 et 2000 en France métropolitaine et celles nées entre 1900 et 2000 dans un département d'outremer ou à Saint-Pierre-et-Miquelon. Le fichier a été créé en 1946 et son informatisation date de 1970-1972. Les personnes nées avant 1946 et décédées avant 1972 n'y figurent pas. Pour chaque personne, le RNIPP contient plusieurs variables dont le nom patronymique devenu nom de famille depuis la loi n° 2002-234. Ce nom est celui de l'état civil et figure sur l'acte de naissance ainsi que sur les autres actes d'état civil (mariage, mention en marge, décès, etc.).

L'INSEE précise que le fichier ne contient que les noms qui ont été donnés trente fois ou plus sur l'ensemble de la période 1890-2000. Nous les dénommerons « noms fréquents ». Le fichier contient au total 218 982 différents noms fréquents. Tous les autres noms ont été regroupés sous AUTRES NOMS. Nous les appellerons « noms rares », parce qu'ils n'ont pas été donnés plus de trente fois sur l'ensemble de la période 1890-2000.

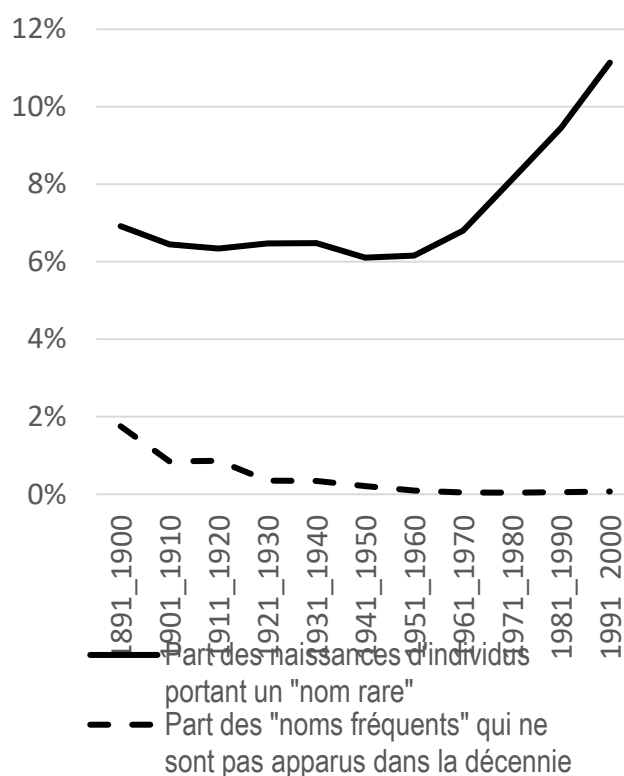
⁴ <https://www.insee.fr/fr/statistiques/3536630>

Les données sont fournies par tranches d'âge décennales, correspondant aux années de naissance telles qu'elles figurent au répertoire ou dans les registres d'état civil. La première décennie est celle de la période 1891-1900. Le fichier recense alors un total de 3 285 631 naissances parmi lesquelles 3 058 364 (soit 93,1 %) sont celles d'enfants portant l'un des 218 982 noms fréquents donnés sur l'ensemble de la période. Parmi ces noms, 57 659 (1,75 %) ne sont encore portés par aucun individu, tandis que 227 267 individus portent un nom rare, qui a été donné moins de trente fois sur toute la période.

Lors de la deuxième décennie, le nombre de naissances recensées par le fichier est de 4 959 983. Cet effectif représente 41,8 % de l'ensemble des naissances en France. L'échantillon n'est pas exhaustif car ne sont recensés que les personnes encore vivantes en 1946, date de la constitution du fichier. Parmi ces naissances, 4 639 988 (93,5 %) sont celles d'enfants portant des noms fréquents sur toute la période. Le fichier permet d'identifier 42 045 (0,8%) patronymes qui ne sont encore apparus sur la période et 319 995 naissances d'individus portant un nom rare.

Puis, de décennie en décennie, le nombre de naissances représentées dans le fichier va régulièrement augmenter jusqu'à atteindre l'exhaustivité dans la décennie 1940. Le nombre de naissance répertoriées va continuer à augmenter sous l'effet de la croissance de la natalité dans le contexte du *baby boom* jusqu'au milieu de la décennie 1960 où 8 875 827 naissances sont répertoriées. La part des noms fréquents répertoriés par le fichier qui ne sont pas encore donnés diminue jusqu'à s'annuler tandis que la part des noms rares augmente régulièrement à partir de 1950 jusqu'à atteindre 11,1% en fin de période (graphique 1).

Graphique 1. Evolution des noms fréquents et des noms rares au sein du fichier de noms de l'INSEE



Source : Insee, fichiers des noms – calculs des auteurs

Champ : France métropolitaine

Source des données : INSEE, fichiers des noms - RNIPP

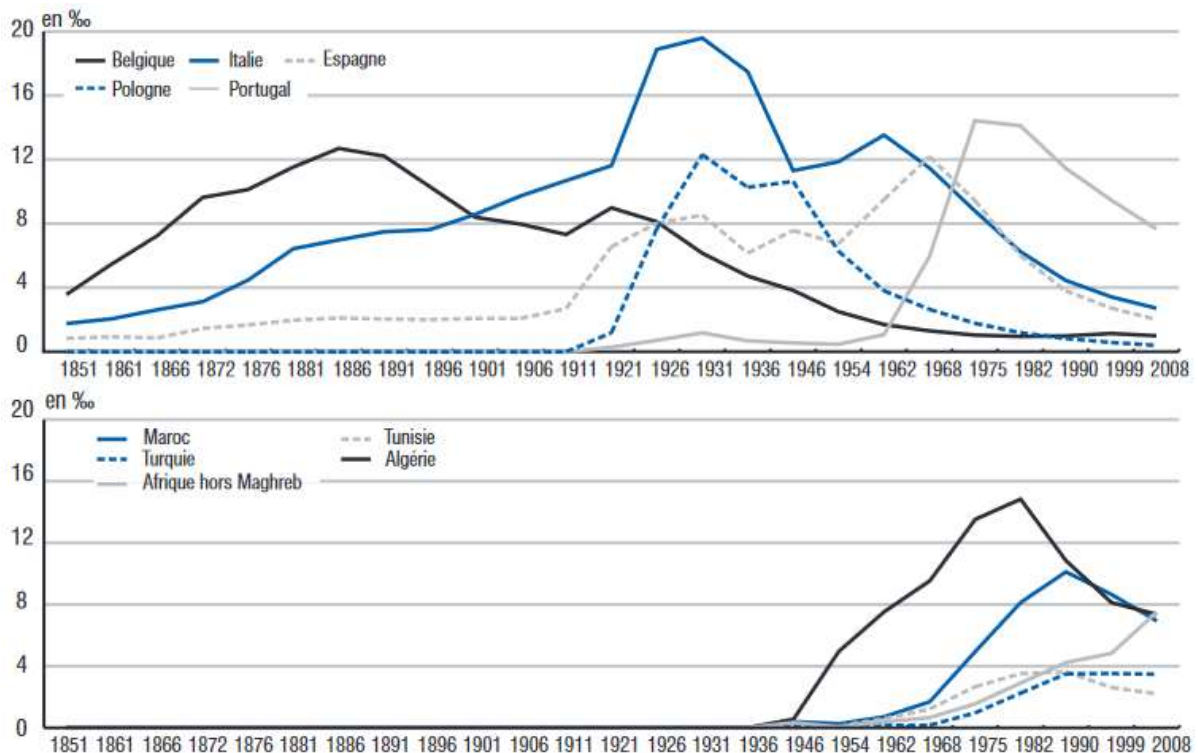
2.3. L'IDP adopte 1950 comme année pivot

Le principe est de construire une variable binaire permettant de coder l'ensemble des patronymes en les distinguant selon leur ancienneté de présence en France métropolitaine. Nous avons retenu 1950 comme année pivot. Les « patronymes issus de la diversité » correspondent ainsi à l'ensemble des patronymes apparus après 1950 dans le fichier des noms et absents auparavant.

Même si la France est depuis très longtemps un pays d'immigration et que les migrations se sont accélérées au XIX^{ème} siècle sous l'effet de la révolution industrielle, nous prenons 1950 comme date de rupture en relation avec la chronologie des migrations internationales qui ont pris la France pour destination. Dans un ouvrage classique de l'histoire des migrations, *Le creuset français* (1988), Gérard Noiriel distingue trois grandes vagues d'immigration en France depuis le XIX^{ème} siècle. La première se situe durant le Second Empire, vers la fin du XIX^{ème} siècle. La deuxième se produit dans l'entre-deux guerres, dans les années 1920. La troisième

vague a lieu après les années 1950. Comme le montrent les données du recensement (graphique 2), chaque vague élargit la géographie des origines des migrants. Surtout limitée à des mobilités transfrontalières au départ, en provenance de Belgique, d'Italie, d'Espagne, puis de Pologne et du Portugal, l'immigration va principalement concerner des populations originaires d'Afrique du Nord puis d'Afrique subsaharienne après la Seconde Guerre mondiale. Cette troisième vague correspond à une « immigration de travail » dans le contexte de forte croissance économique des Trente Glorieuses jusqu'en 1974 qui nécessite de forts besoins en main d'œuvre. Par la suite, la part des femmes va augmenter jusqu'à devenir prépondérante et les motifs des migrations deviennent de plus en plus familiaux.

Graphique 2. Vagues migratoires pour les nationalités les plus nombreuses en France depuis 1851



Champ : France métropolitaine

Lecture : au recensement de la population de 1931, les personnes de nationalité italienne représentent 20 millièmes (2%) de la population résidente en France

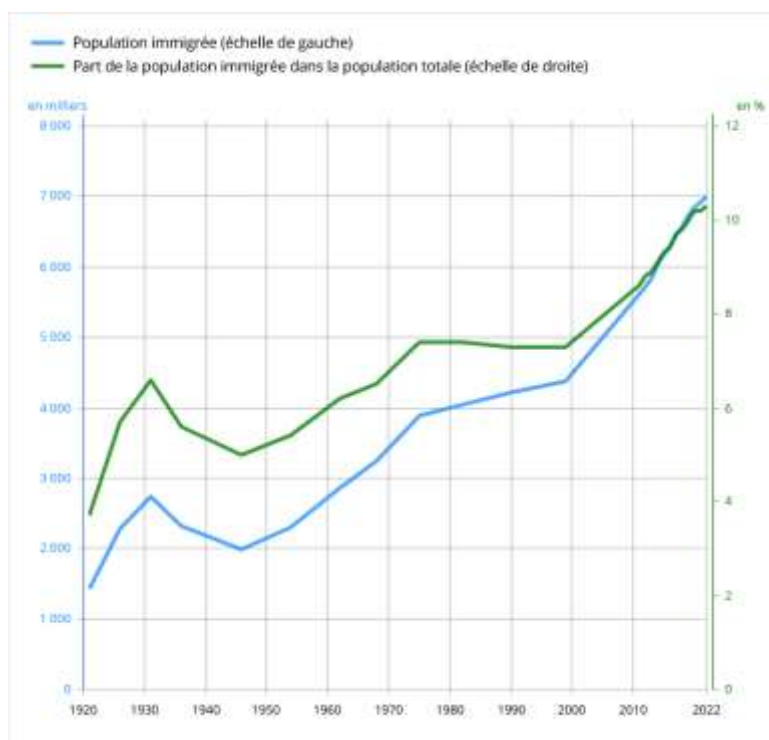
Source des données : INSEE, recensement de la population

Source de la publication : Insee référence, Immigrés et descendants d'immigrés en France, édition 2012.

La décennie 1950 correspond également au le début de la décolonisation qui va aller de pair avec des flux migratoires croissants vers la France hexagonale. Les principales dates qui structurent la chronologie des déclarations d'indépendance sont les suivantes : protectorat français du Maroc (2 mars 1956), Tunisie (20 mars 1956), Guinée (2 octobre 1958), Cameroun (1^{er} janvier 1960), Togo (27 avril 1960), Sénégal (20 juin 1960), Madagascar (26 juin 1960), Bénin (1^{er} août 1960), Niger (3 août 1960), Burkina Faso (5 août 1960), Côte d'Ivoire (7 août 1960), Tchad (11 août 1960), République Centrafricaine (13 août 1960), République du Congo (15 août 1960), Gabon (17 août 1960), Mauritanie (28 novembre 1960), Algérie (5 juillet 1962), Comores (6 juillet 1975), Djibouti (27 juin 1977).

Après avoir marquée un repli de 1930 à 1945, où le nombre d'immigrés en France métropolitaine atteint un point bas de deux millions de personnes et d'un peu moins de 5 % de la population de France métropolitaine, la population immigrée connaît une phase d'expansion à la fois absolue et relative jusqu'au milieu des années soixante-dix (graphique 3). Par la suite, pendant trois décennies, le nombre d'immigrés progresse à un rythme moindre et la part des immigrés dans la population totale diminue. Une remontée se produit dans les années 2000 et jusqu'en 2020 où le nombre d'immigrés atteint sept millions, soit 10,3 % de la population active. Dans cet ensemble, la part des immigrés en provenance d'Afrique du Nord et d'Afrique subsaharienne représente un peu moins de la moitié en 2021 contre un cinquième en 1968 (Rouhban et Tanneau, 2023).

Graphique 3. Evolution de la population immigrée en France métropolitaine



Champ : France métropolitaine de 1921 à 1982 ; France hors Mayotte de 1990 à 2013 et y compris Mayotte à partir de 2014.

Lecture : en 2022, la part de la population immigrée représente 10,3 % de la population totale.

Source : INSEE, recensements de la population (données réajustées en 2019 et 2020) et estimations de population.

3. Applications

La démarche retenue consiste en premier lieu à établir une liste de noms de famille issus de la diversité (au sens patronymique) à partir du fichier des noms de l'INSEE. Dans un deuxième temps, cette liste est appariée avec une liste donnée de noms de famille pour y calculer le pourcentage de personnes qui portent un nom de famille issu de la diversité, correspondant à l'IDP. Dans cette dernière section, nous détaillons la procédure de traitement des données avant de présenter le résultats d'une quinzaine d'applications.

3.1. Traitement des données

En premier lieu, nous distinguons au sein du fichier des noms de l'INSEE, parmi l'ensemble des 218 982 noms de famille présents plus de trente fois entre 1890 et 2000, les noms qui ne sont apparus de façon significative qu'après 1950. Un patronyme issu de la

diversité satisfait au moins l'un des deux critères suivants. Une première possibilité est d'être un patronyme totalement absent avant 1950 parmi les patronymes présents plus de trente fois entre 1890 et 2000. Ce type de patronyme n'apparaît donc qu'après 1950. On dénombre un total de 9 941 patronymes qui satisfont ce critère, soit 4,5 % de l'ensemble des patronymes.

Nous avons retenu également une deuxième possibilité : le patronyme est présent avant 1950 mais de façon marginale et il va connaître une très forte expansion à partir de 1950. Nous avons retenu un facteur quinze pour qualifier cette expansion : un patronyme présent avant 1950 qui voit ses effectifs multipliés par plus de quinze est classé comme patronyme issu de la diversité. On compte au total 6 828 noms de familles qui ont cette caractéristique, soit 3,1 % de l'ensemble des patronymes.

On dénombre donc un ensemble de 16 769 patronymes issus de la diversité, qui satisfont l'une ou l'autre de ces deux conditions exclusives, soit 7,6 % de l'ensemble des patronymes. Sur la dernière décennie 1990-2000, où l'on compte un total de 6 839 122 naissances, on dénombre 727 733 naissances d'enfants issus de la diversité patronymique, soit 10,6 % des naissances.

Nous avons utilisé les tables de mortalité publiées par l'INSEE pour les appliquer à toutes les naissances de façon à décompter la population vivante à chaque décennie. En 2000, la population vivante ainsi calculée, issue de l'addition de toutes les naissances auxquelles on applique des quotients de mortalité par âge révolu, est de 50,6 millions d'individus. Ce calcul exclu par construction les immigrés, qui sont au nombre de 4,5 millions en 2000, soit un total de 55,1 millions d'individus, peu éloigné des 58,9 millions de personnes recensées en 2000 par l'INSEE en France métropolitaine. Sur cette base, le nombre de personnes issues de la diversité patronymique est de 2,4 millions en 2000, soit 4,7 % des personnes nées en France.

Cette procédure de traitement des données est susceptible d'être ajustée. Le facteur quinze que nous avons retenu peut être discuté et amélioré. En pratique, la liste des patronymes issus de la diversité ainsi obtenue ne comprend pas uniquement des noms de famille dont la consonance évoque une origine d'Afrique du Nord ou d'Afrique subsaharienne. Elle contient aussi quelques noms de famille de consonance espagnole, italienne ou portugaise. Une lecture détaillée indique aussi que de nombreux patronymes sont d'origine turque et que d'autres sont d'origine asiatique.

Une limite importante de l'exercice est que c'est le nom de famille porté par le père qui est très majoritairement donné aux enfants. Des changements récents sont intervenus mais les nouvelles pratiques sont encore peu développées. Dès lors, les unions mixtes entre une personne issue de l'immigration, qu'elle soit de nationalité française ou non, et une personne française d'origine métropolitaine sont susceptibles d'affecter l'IDP. Lorsque dans l'union mixte, c'est le père qui est immigré ou d'ascendance immigrée, sa descendance est comptée dans les français issus de la diversité. Lorsque c'est la mère, sa descendance ne l'est pas. Cette limite est réelle mais elle ne doit pas être exagérée. D'une part, il est probable que les unions mixtes avec un père d'origine métropolitaine équilibrent celles avec une mère d'origine métropolitaine. D'autre part, si un biais existe, il est probable qu'il affecte de la même manière les différentes listes de noms sur lesquelles l'indice est mesuré et qu'il n'altère donc pas la mesure relative ou le classement des groupes du point de vue de leur IDP.

Une autre limite réside dans l'expansion du nombre des noms de famille de moins de trente occurrences qui est notable à partir de 1950. On compte au total 415 666 noms rares présents en plus entre la décennie 1990-2000 et la décennie 1940-1950. Il est probable que les noms issus de la diversité soient surreprésentés dans ces patronymes et que la mesure de la diversité patronymique s'en trouve sous-évaluée. Ce risque est réel et il ne peut être évité. Mais à nouveau il ne nous paraît de nature à altérer une lecture relative de l'indice lorsque des groupes sont comparés les uns aux autres.

3.2. Exemples d'application

Afin de tester la pertinence de l'indice, nous avons recherché des listes de patronymes dans différents domaines professionnels, sportifs et politiques dont les noms de famille des membres avaient un caractère public. Les sources utilisées figurent en annexe 1. La démarche et ses résultats peuvent être comparés à ceux obtenus par Mazières et Roth (2018) qui utilisent 15 bases de données pour appliquer leur calculateur onomastique à une grande variété de groupes socio-professionnels (figure 2, page 68).

Ces groupes constituent autant d'échantillons de taille et de composition variées, par âge, sexe et localisation. Nous avons apparié chacun de ces échantillons avec les fichiers des noms de l'INSEE afin de pouvoir calculer la valeur de l'IDP. Un travail préalable a été de mettre les

listes au même format que celle de l'INSEE. Nous écrivons les patronymes en capital, sans accent, sans double tiret avant de procéder à l'appariement des fichiers. Les patronymes comportant des apostrophes, des traits d'unions, des espaces, de même que les patronymes composés doivent être surveillés et harmonisés avant appariement. Ensuite, nous appliquons strictement la même procédure de traitement des données et de calcul de l'indice, parfaitement harmonisée d'un échantillon à l'autre.

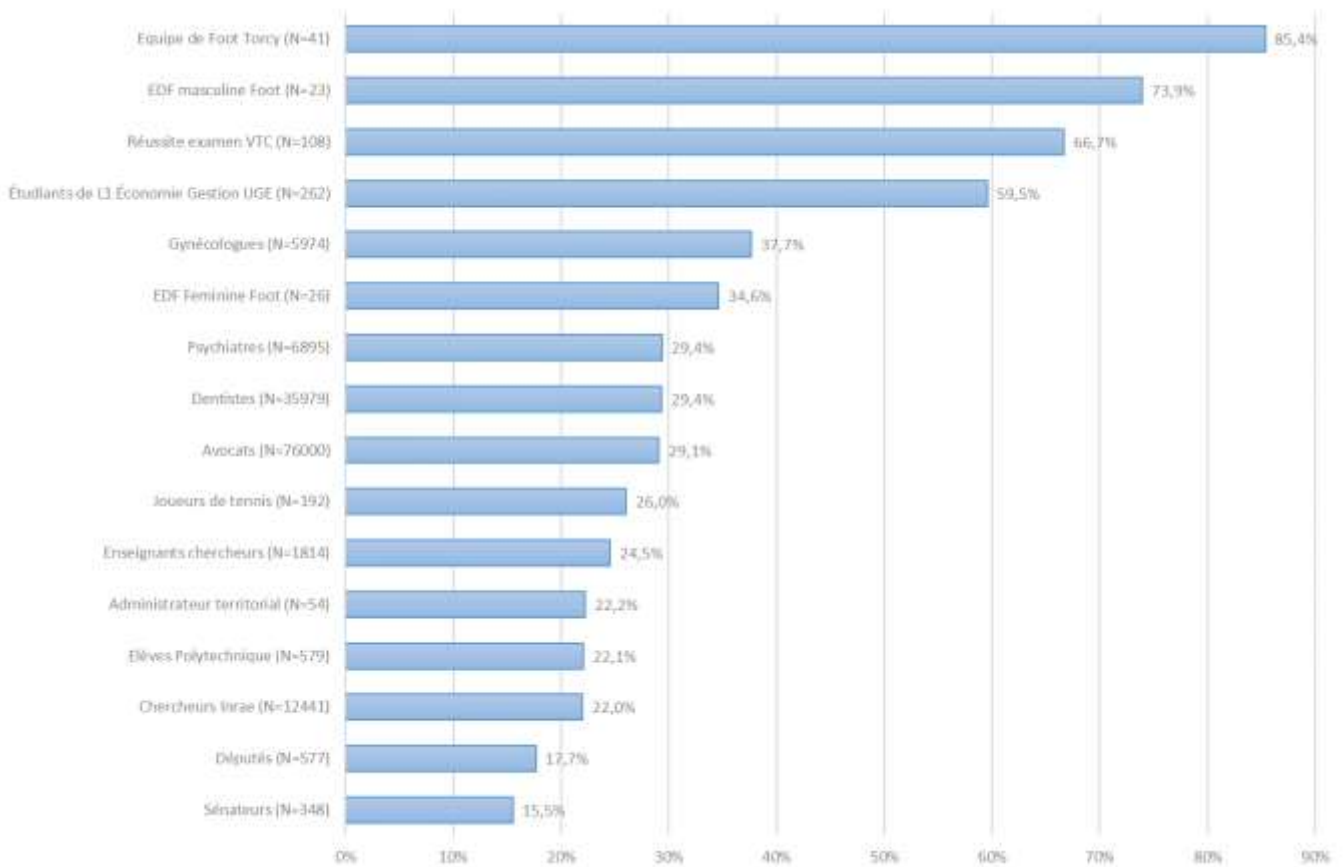
Les échantillons relèvent de trois domaines distincts. En premier lieu, nous avons effectué des estimations de l'IDP pour différentes institutions publiques où une problématique de représentativité de la société civile peut être posée : le Sénat, l'Assemblée Nationale, l'administration territoriale. Les résultats montrent que les sénateurs constituent un groupe moins ouvert à la diversité que les députés et que les lauréats d'un concours d'administrateur territorial sont un peu plus diversifiés, même si ces trois échantillons figurent globalement parmi les moins diversifiés.

En second lieu, nous avons exploré différentes professions pour lesquelles il nous a été possible d'accéder à des listes publiques. Les chercheurs et chercheuses de l'INRAE constituent une population moins diversifiée que celle des enseignant-es-chercheur-euses qui elle-même est moins diversifiée que celle des avocat-es, puis des dentistes, des psychiatres et des gynécologues. Les lauréat-es d'un examen d'aptitude à la profession de conducteur-riche de véhicule de tourisme avec chauffeur (VTC) constitue une profession parmi les plus diversifiées.

En troisième lieu, nous avons comparé différentes populations d'élèves et d'étudiant-es. Les élèves admis-es ou en liste d'attente à l'école polytechnique sont très peu diversifié-es lorsqu'on les compare aux étudiant-es d'un premier cycle universitaire.

Enfin, nous avons aussi exploré la diversité sportive en mettant en évidence que le classement ATP des meilleurs joueurs de tennis était beaucoup moins diversifié que la composition de l'équipe de France de football féminin qui est elle-même moins diversifiée que l'équipe de France masculine. L'équipe locale de la ville de Torcy parvient à surclasser l'équipe de France en matière de diversité patronymique.

Graphique 4. Calculs de l'IDP Evolution de la population immigrée en France métropolitaine



Conclusion

En France, les débats publics sur la diversité manquent structurellement de mesures. Pour autant, la prohibition constitutionnelle des référentiels ethno-raciaux ne rend pas impossible le suivi statistique de la diversité d'origine. Dans cette note, nous proposons un nouvel indice permettant d'évaluer la diversité patronymique à l'échelle d'un groupement humain de petite ou de grande taille en exploitant uniquement une liste de noms de famille, sans aucune autre information de caractère personnelle. Nous avons détaillé la procédure de calcul de cet indice de diversité patronymique (IDP) et nous présentons une quinzaine d'applications sur des groupements variés par leur taille et leur nature.

Si la méthode de calcul est évidemment perfectible, elle présente suffisamment d'avantages et trop peu de défauts pour ne pas être testée sur une plus grande échelle. Il nous semble que l'indice pourrait constituer une solution économique pour les organisations publiques ou

privées qui souhaitent piloter leur propre diversité. Il peut s'avérer utile dans des domaines aussi divers que la réussite au baccalauréat ou aux concours de la fonction publique, en matière de recrutement, d'égalité professionnelle selon l'origine, dans les progressions de carrière. Au-delà du monde du travail, l'IDP peut être utile dans tous les domaines de la vie sociale où prévalent les enjeux d'égalité entre citoyens selon leur origine. Ils sont nombreux et les acteurs ont besoin d'être mieux équipés pour affronter les défis auxquels ils sont confrontés.

Les perspectives pour les chercheurs et chercheuses en sciences humaines et sociales et dans toutes les disciplines intéressées par les problématiques d'égalité et de lutte contre les discriminations sont là aussi très importantes. La possibilité de disposer d'une mesure partagée de la diversité peut ouvrir des portes vers des domaines encore inexplorés.

Bibliographie

- Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., and Wacziarg, R. (2003). "Fractionalization". *Journal of Economic Growth* 8: 155– 94.
- Alesina, A., and La Ferrara, E., (2005). "Ethnic Diversity and Economic Performance. *Journal of Economic Literature* 43 (3): 762– 800.
- Beauchemin, C., Hamel, C., & Simon, P. (2016). *Trajectoires et origines: enquête sur la diversité des populations en France*. Ined éditions.
- Beauchemin, C., Ichou, M., Simon, P., & Dutreuilh, C. (2023). « Trajectories and Origins 2 (2019–2020): A Survey on Population Diversity in France ». *Population*, 78(1), 11-28.
- Becker, Gary S. (1957). *The Economics of Discrimination*. The University of Chicago Press.
- Bereni, L. « « Faire de la diversité une richesse pour l'entreprise ». La transformation d'une contrainte juridique en catégorie managériale », *Raisons politiques*, vol. 35, no. 3, 2009, pp. 87-105.
- Bereni, L. & Chappe, V. (2011). « La discrimination, de la qualification juridique à l'outil sociologique ». *Politix*, 94, 7-34.
- Bertrand, M., & Duflo, E. (2017). Chapter 8—Field Experiments on Discrimination. In A. V. Banerjee & E. Duflo (Éds.), *Handbook of Economic Field Experiments* (Vol. 1, p. 309-393). North-Holland.
- Blanchard, E. (2018). *Histoire de l'immigration algérienne en France*. La Découverte.
- Blau, P. M., (1977). *Inequality and Heterogeneity*. New York: Free Press.
- Breda, T., & Hillion, M. (2016). Teaching accreditation exams reveal grading biases favor women in male-dominated disciplines in France. *Science*, 353(6298), 474-478.
- Caselli, F., and Coleman, W. J. II. (2006). "On the Theory of Ethnic Conflict." *CEPR Discussion Paper* 5622.
- Challe L., S. Chareyron, Y. L'Horty et P. Petit (2020), « Discrimination dans le recrutement des grandes entreprises: une approche multicanal », Rapport de recherche TEPP (2020-01).
- Chareyron, S. & L'Horty Y., (2023) « Discriminations ethno-raciales dans l'accès au logement social : un test des guichets d'enregistrement », TEPP Rapport de recherche, n°23-3.
- Collier, Paul. 2001. "Implications of Ethnic Diversity. *Economic Policy* 16 (32): 129– 66.
- Coulmont, B., & Simon, P. (2019). Quels prénoms les immigrés donnent-ils à leurs enfants en France ?. *Population & Sociétés*, (4), 1-4.
- Du Parquet, L., & Petit, P. (2019), « Discrimination à l'embauche : retour sur deux décennies de testing en France », *Revue Française d'Economie*, Vol. 34, n°1.
- Dubet F. (2014). « Discriminations : égalité, reconnaissance et retours du refoulé », *La nouvelle revue des sciences sociales*, 3, p. 315-349.
- Duguet E., N. Leandri, Y. L'Horty, P. Petit. (2010). "Are young French jobseekers of ethnic immigrant origin discriminated against? A controlled experiment in the Paris area", vol. 99-100, pp. 187-215.

- Duguet, E. and P. Petit (2005). "Hiring discrimination in the French financial sector: an econometric analysis on field experiment data". *Annals of Economics and Statistics*, vol. 78, pp. 79-102.
- Easterly, William, and Ross Levine. 1997. "Africa's Growth Tragedy: Policies and Ethnic Divisions". *Quarterly Journal of Economics* 112 (4): 1203– 50.
- Fassin, Didier. « L'invention française de la discrimination », *Revue française de science politique*, vol. 52, no. 4, 2002, pp. 403-423.
- Fearon, James D. 2003. "Ethnic and Cultural Diversity by Country". *Journal of Economic Growth* 8: 195– 222.
- Gabuthy, Y. Jacquemet, N., & L'Haridon, O. (2021). *Économie comportementale des politiques publiques*. La Découverte. Collection Repères, 128 p.
- Greenan, N., Lanfranchi, J., L'Horty, Y., Narcy, M., & Pierné, G. (2019). Do competitive examinations promote diversity in civil service?. *Public Administration Review*, 79(3), 370-382.
- Harrison, David A., and Katherine J. Klein. 2007. "What's the Difference? Diversity Constructs as Separation, Variety, or Disparity in Organizations". *Academy of Management Review* 32 (4): 1199– 228.
- Haut Conseil à l'intégration, , rapport au Premier ministre, Paris, La Documentation Française, 1998
- Héran, F. (2010). « Inégalités et discriminations - Pour un usage critique et responsable de l'outil statistique : rapport du comité pour la mesure de la diversité et l'évaluation des discriminations ». 272 pages.
- L'Horty, Y., & Petit, P. (2023). Mesurer des discriminations ethno-raciales en France: l'apport des testing. *Appartenances & Altérités*, (3).
- Lê, J., Rouhban, O., Tanneau, P., Beauchemin, C., Ichou, M., & Simon, P. (2022). « En dix ans, le sentiment de discrimination augmente, porté par les femmes et le motif sexiste ». *Insee Première*, (1911), 1-4.
- Lippens, L., Vermeiren S., and Baert, S., (2021). « The State of Hiring Discrimination: A Meta-Analysis of (Almost) Recent Correspondence Experiments », IZA DP n°14966
- Mateos, P. (2007). "A review of name-based ethnicity classification methods and their potential in population studies". *Population, Space and Place*, 13(4), 243-263.
- Mazières, A., & Roth, C. (2018). "Large-Scale Diversity Estimation Through Surname Origin Inference". *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 139(1), 59–73.
- Miguel, Edward, and Mary Kay Gugerty. 2005. "Ethnic Diversity, Social Sanctions, and Public Goods in Kenya". *Journal of Public Economics* 89: 2325– 68.
- Neumark, D. (2018). "Experimental research on labor market discrimination". *Journal of Economic Literature*, vol. 56, pp. 799–866.
- Noiriel, G. (1988). *Le creuset français : histoire de l'immigration, XIXe-XXe siècles*. Paris: Editions du Seuil, 1871-1971.

- Ottaviano, Gianmarco I. P., and Giovanni Peri. 2006. "The Economic Value of Cultural Diversity: Evidence from US cities. *Journal of Economic Geography* 6: 9– 44.
- Pitts, David W. 2005. "Diversity, Representation and Performance: Evidence about Race and Ethnicity in Public Organizations. *Journal of Public Administration Research and Theory* 15 (4): 615– 31.
- Quillian, L., D. Pager, O. Hexel, and A. H. Midtbøen (2017). "Meta-analysis of field experiments shows no change in racial discrimination in hiring over time". *Proceedings of the National Academy of Sciences*, vol. 114, no. 41, pp. 10870–10875.
- Riach, P. A., & Rich, J. (2002). "Field Experiments of Discrimination in the Market Place". *The Economic Journal*, 112(483), F480-F518.
- Richard, O. C. (2000). "Racial Diversity, Business Strategy, and Firm Performance: A Resource-Based View. *Academy of Management Journal* 43 (2): 164– 77.
- Rooth D-A., (2014). "Correspondence testing studies". *IZA World of Labor*, pp. 1-58.
- Rouhban, O., et Tanneau P. (2023). « Une situation des descendants d’immigrés plus favorable que celle des immigrés », in *Immigrés et descendants d’immigrés*, INSEE Références, édition 2023.
- Zschirnt, E. and D. Ruedin (2016). "Ethnic discrimination in hiring decisions: A meta-analysis of correspondence tests 1990-2015". *Journal of Ethnic and Migration Studies*, vol. 42, pp. 1115-1134.

ANNEXE. Source des données utilisées pour les applications

« Chercheurs Inrae » : www.inrae.fr

« Classement ATP Tennis France » : https://www.fft.fr/sites/default/files/2023-06/JoueursNumerotes_Juin23.pdf

« Concours Administrateur » : <https://www.cnfpt.fr/sites/default/files/document/1668701599/liste-candidats-admis-au-concours-dadministrateur-territorial-2022.pdf>

« EDF Féminine Football » : <https://www.fff.fr/selection/3-equipe-de-france-feminine/derniere-selection.html>

« EDF masculine Football » : <https://www.fff.fr/selection/2-equipe-de-france/derniere-selection.html>

« Polytechnique » : <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000046357587>

« Dentistes », « Psychiatres » et « Gynécologues » : www.ameli.fr

« Avocat » : <https://www.data.gouv.fr/fr/datasets/annuaire-des-avocats-de-france/>

« Député » : <https://www2.assemblee-nationale.fr/deputes/liste/alphabetique>

« Joueur de Foot Torcy » : <https://www.foot-national.com/data/679-club-effectif-Torcy.html>

« Sénateur » : <https://www.senat.fr/senateurs/senatl.html>

« VTC » : <https://www.cm-toulouse.fr/files/Taxis-et-VTC/Liste-ADMIS-Session-Aout-Sept-2022-CMA31.pdf>



Rapports d'étude de l'ONDES

22-01. Discrimination dans l'accès aux masters : une évaluation expérimentale.

Sylvain Chareyron, Louis-Alexandre Erb et Yannick L'Horty

22-02. Les écarts de rémunération entre les femmes et les hommes : une étude à l'échelle d'une Université

Rahma Bensalem

22-03. Le harcèlement scolaire à l'encontre des LGBTQ+ : Une enquête par questionnaire

Mickaël Jardin

22-04. Origine ou couleur de la peau ? Anatomie des discriminations à l'embauche dans le secteur du prêt-à-porter

Dianké Tchabo

23-01. Sélection à l'entrée en master : les effets du genre et de l'origine

Sylvain Chareyron, Berlanda Desuza Fils-Aimé, Yannick L'Horty

23-02. L'orientation universitaire explique l'essentiel des inégalités de genre sur le marché du travail

Louis-Alexandre Erb

23-03. Les actions pour l'égalité des établissements d'enseignement supérieur : un état des lieux

Yannick L'Horty, Philippe Liotard, Romane Masternak, Aude Stheneur