



**HAL**  
open science

# Les liens de la caste. Quantifier des identités sociales à l'aide de questions ouvertes

Mathieu Ferry

► **To cite this version:**

Mathieu Ferry. Les liens de la caste. Quantifier des identités sociales à l'aide de questions ouvertes. 2019. hal-03456902

**HAL Id: hal-03456902**

**<https://sciencespo.hal.science/hal-03456902>**

Preprint submitted on 12 Jan 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# OSC *Papers*



n° 2019-1 Mai 2019  
version française

Les liens de la caste

Quantifier des identités sociales à l'aide de questions ouvertes

Mathieu Ferry (OSC, LSQ)

## Introduction

Depuis les recensements coloniaux, l'administration statistique indienne ne prend plus en compte l'appartenance de caste pour décrire les propriétés sociales des individus et des ménages. Ce manque de données statistiques à la disposition des chercheurs en sciences sociales contraste avec une littérature mettant en évidence la pertinence de la caste dans la structure sociale indienne (Jaffrelot et Naudet, 2013 ; Vaid, 2014). Si les méthodes ethnographiques fournissent indéniablement des éléments précieux dans la compréhension de la société indienne, l'absence de statistiques exclut une vision synthétique de la réalité sociale des castes.

Pour clarifier, le terme de caste recouvre deux réalités différentes. D'abord, il fait référence à une quadripartition hiérarchique du monde entre les varnas, opposant graduellement les castes suivant leur niveau de pureté rituel. Ce système est issu du *Rig Veda*, un texte sacré de l'hindouisme. Dans ce schéma idéal, les Brahmanes, traditionnellement considérés comme des prêtres, sont au sommet de la hiérarchie, les Kshatriyas arrivent au second rang ; ils sont censés être des guerriers. Suivent ensuite les Vaishyas, traditionnellement des commerçants. Les Shudras, qui seraient des travailleurs agricoles, des ouvriers et des artisans, sont considérés bien en dessous des trois premières varnas, les hautes castes « deux fois nées ». En dehors de ce système, les intouchables rassemblent des hors castes impures. Mais cette catégorisation est davantage idéologique qu'une description de la réalité. Dans le monde social, on entend par caste la jati, une communauté définie comme un groupe social héréditaire, endogame et caractérisée par un métier traditionnel. Varnas et jatis sont liées mais ne correspondent pas parfaitement. En effet, les jatis revendiquent des positions dans la hiérarchie des varnas, qui peuvent ou non être contestées par d'autres jatis. Ce qu'une jati prétend être du point de vue de la varna n'est nécessairement pas ce que ses pairs considèrent comme tel et les étiquettes sociales varient temporellement et géographiquement. L'identification contestée de la caste en fait une catégorie centrale de la stratification sociale indienne et alimente encore les débats théoriques sur sa conceptualisation, cependant sans sources quantitatives empiriques.

Le vide statistique de la caste est parfois étonnamment comblé dans les politiques publiques par des données d'archives des recensements coloniaux datant de près de 80 ans. Ceux-ci sont ensuite par exemple

mobilisés pour justifier la mise en œuvre de politiques de discrimination positive, afin d'attribuer des statuts spécifiques aux castes les plus défavorisées dans la société, les jatis dites « *backward* » (« retardées »). De nombreux chercheurs en sciences sociales (y compris l'auteur de cet article, voir Ferry, Naudet et Roueff, 2018) utilisent ensuite ces catégories de discrimination positive comme *proxy* de la caste dans les analyses statistiques. Mais dans l'ensemble, cette situation rend difficile la conduite de travaux de recherche appropriés pour les sociologues de la stratification sociale.

Les grandes enquêtes contemporaines par échantillonnage représentatif collectent toutefois des données sur l'identité de caste par le biais de questions ouvertes. Celles-ci sont présentes dans l'Indian Human Development Survey (IHDS, Desai et al., 2018), une enquête longitudinale menée depuis 2004, et dans l'enquête nationale sur la santé de la famille (NFHS) menée depuis 1992. L'absence de classification des réponses dans une nomenclature par les enquêteurs pose cependant problème. Afin de mobiliser ces données, le chercheur doit lui-même classer les réponses, ce qui apparaît d'autant plus difficile qu'il n'existe pas de schéma de caste standardisé. La présente recherche vise alors à fournir une méthode pour pallier ce manque. Ce faisant, la conceptualisation théorique de la caste n'est pas oubliée et la pertinence conceptuelle de la classification sera évaluée. Mais nous privilégions d'abord l'auto-identification des personnes interrogées et essayons de tracer de façon inductive les groupes de caste les plus homogènes avec un algorithme de classification automatique, ce qui permet d'éviter des regroupements manuels extrêmement chronophages.

Les frontières statistiques ne correspondent jamais parfaitement aux frontières sociales, dans la mesure où la catégorisation résume un phénomène complexe de manière simplifiée. Mais nous soutenons dans le même temps que des outils statistiques peuvent être mobilisés pour explorer les frontières pertinentes entre les identités sociales tout en évitant de créer des chimères statistiques. Une approche statistique et sociologique de la catégorisation permet de s'affranchir d'une vision figée de la caste, en particulier de sa caractéristique hiérarchique (Dumont, 1967), mise en cause car s'appuyant trop fortement sur une vision idéologique de cette dernière (les varnas), plutôt que sur les jatis observées (Dirks, 2001). La distinction établie par Desrosières (1993) entre « mesurer » et « quantifier » est ainsi importante dans notre approche. Alors que la première action se rapporte à une métrologie réaliste, où une identité réelle serait mesurée, la seconde rappelle les

conventions mobilisées lors de la catégorisation du monde social. La « mise en nombre » du monde social, réalisée à la fois par les sciences sociales et par l'action gouvernementale, vise à décrire une réalité objective, mais aussi à « mettre en forme » cette réalité (Desrosières et Didier, 2014). En oubliant la performativité de l'outil statistique, l'analyse court le risque d'être réalisée au prix d'une abstraction du fait social, masqué par l'opération de quantification. Au contraire, en exposant clairement les conventions adoptées dans la classification, nous pouvons comprendre *a posteriori* la réalité sociale incorporée dans les catégories sociales mesurées par des outils statistiques. Par conséquent, au lieu de désespérer de l'absence de catégories institutionnalisées de la caste, le présent exercice offre une occasion unique de questionner le caractère de la caste et sa saillance dans la structure sociale.

Nous proposons ici de poser les jalons d'une quantification de la caste en nous centrant sur la population hindoue de l'Uttar Pradesh. La variabilité régionale des identités de caste nous amène à nous concentrer sur un État particulier, et l'Uttar Pradesh a l'avantage d'être dans une région du nord de l'Inde où le système des castes est considéré comme proche du modèle des varnas. Nous verrons cependant que l'auto-identification de caste par les ménages n'est que marginalement proche de cette classification<sup>1</sup>.

Dans la suite de l'article, nous rappelons tout d'abord que la mesure statistique de la caste est liée à l'histoire coloniale indienne, qui marque fortement les débats contemporains sur sa quantification. Ce rappel historique nous amène à discuter des arguments contre la mesure de la caste et à développer notre position sur la nécessité de prendre en compte statistiquement les identités de caste. L'algorithme, associant une classification automatique et un codage manuel, est ensuite présenté, avant d'en évaluer sa pertinence. Les catégories de caste apparaissent homogènes du point de vue des identités énoncées ; la nomenclature est efficace lorsqu'elle est comparée à un « gold standard » et les tests de validité montrent qu'elle quantifie adéquatement notre conceptualisation de la caste. En conclusion, nous discutons des

1. Il nous faut aussi préciser que la restriction à la population hindoue est avant tout pragmatique car, si l'on ne veut pas nier une structure de caste au sein des minorités religieuses chrétiennes (Roberts, 2016) ou musulmanes (Ahmad, 1973), les données ne la mesurent pas avec précision. Les taux de non-réponse sont souvent élevés parmi ces populations et les réponses font souvent référence à des subdivisions religieuses (par exemple, sunnite ou chiites) plutôt qu'à des appartenances de caste. Les autres minorités religieuses indiennes (Sikhs ou Zoroastriens) ne sont que très marginalement présentes dans l'Uttar Pradesh.

implications de notre méthode de classification et des utilisations futures de cette nomenclature.

## 1. Leçons pour une quantification de la caste dans l'Inde contemporaine

Si notre objectif est de montrer la pertinence de l'utilisation d'enquêtes statistiques pour étudier sociologiquement les identités de caste, il faut d'abord rappeler que l'exercice de quantification de la caste est lié à l'histoire coloniale de l'Inde. Les recensements, ainsi que l'anthropologie coloniale, ont indéniablement rigidifié une représentation du système de castes, et les outils statistiques ont eu des effets performatifs dans le monde social. Cette situation explique la réticence actuelle des sciences sociales à mobiliser l'appareil statistique pour étudier la caste.

### a. Les outils de production du savoir colonial

Lorsque le premier recensement britannique a été introduit dans l'Inde coloniale en 1871, les comptages administratifs n'étaient pas entièrement nouveaux sur le sous-continent et ils étaient basés sur des expériences pré-statistiques dans les provinces, administrées soit par les autorités britanniques, soit par les autorités locales (les Moghols en particulier). Ces chiffres étaient destinés à mesurer la production, en particulier agricole, afin de fixer le taux d'imposition optimal. Si le recensement scientifique de 1871, qui en réalité a eu lieu de 1867 à 1873, n'a pas suscité d'opposition de la population (hormis les révoltes fiscales localisées, déjà observées au cours du XIXe siècle), mais plutôt de l'étonnement et l'ironie de la part des interrogés face à la curiosité des enquêteurs, dont les questions n'allaient pas forcément de soi (Lardinois, 1996).

Ce recensement a en effet introduit des questions sur la religion, la secte et la caste, dans un but de comptabilisation et d'énumération. Les recensements constituaient des outils cruciaux pour le contrôle du territoire et de la population sous la domination de l'empire colonial (Appadurai, 1993). Si des techniques statistiques avaient déjà été introduites à partir d'expériences de recensement au Royaume-Uni, peu d'informations sociodémographiques y étaient recueillies (uniquement la religion en 1851), l'intérêt pour l'exotisme des groupes sociaux indiens étant beaucoup plus important (Guilmoto, 1998).

La caste est progressivement devenue la principale unité de description de la société indienne, une vision renforcée par l'anthropologie

officielle qui s'est développée parallèlement, sous l'égide de Herbert Risley, William Crooke et Denzil Ibbetson (Fuller, 2016, 2017). Le projet de description ethnographique des groupes de castes, *People of India*, renforçait *a priori* d'une société indienne traditionnelle, antithétique de la société européenne moderne.

### **b. La production du savoir colonial**

Selon Herbert Risley et, dans une certaine mesure, William Crooke, l'organisation sociale de la caste était fondée sur le paradigme évolutionniste de la théorie raciale, qu'ils cherchaient à soutenir à travers la collection de données anthropométriques. Cette conceptualisation était renforcée par une théorie de la « préséance sociale », reproduisant ainsi un modèle de compréhension sociale de la société britannique fondée sur le rang. Dans le même temps, l'administration statistique coloniale comptait également sur des informateurs indigènes, principalement des « bhadraloks » du Bengale, une classe lettrée composée de hautes castes brahmanes, Kayasths et Baidyas, qui a émergé durant l'Inde britannique (Fuller, 2017). La taxinomie coloniale se serait ainsi appuyée sur une vision socialement située de la caste, légitimant une vision hiérarchique du système de castes soutenu par les écrits sacrés, en particulier le *Rig Veda* (Appadurai, 1993).

L'historiographie coloniale est dès lors partagée sur le degré d'influence de ces acteurs dans la production du savoir colonial. La première position, qualifiée de « postcoloniale » par Fuller (2017), qui résume lui-même Wagoner (2003), est notamment celle de Nicholas Dirks (2001). Pour lui, la compréhension du système de castes en tant que hiérarchie sociale découle avant tout de la pensée européenne appliquée à la société indienne, où les informateurs indiens ne jouaient qu'un rôle passif dans la collecte de données. La deuxième position, qualifiée de « révisionniste » par rapport à la première, tient compte du fait que les informateurs ont joué un rôle actif dans la production du savoir colonial de la société indienne, qui serait donc le fruit à la fois de la pensée coloniale et autochtone. Cette position est plutôt celle développée par Susan Bayly (2001), qui note également que les valeurs brahmaniques sont devenues dominantes dans l'ordre moral de la société indienne au XVIII<sup>e</sup> siècle. L'émergence de cette configuration a été renforcée par la proximité des castes lettrées, et en particulier brahmanes, à l'appareil colonial. Synthétisant les deux positions, Fuller (2017) évoque une « compréhension partagée » entre la pensée évolutionniste européenne et la pensée brahmanique qui a

produit une connaissance de la caste organisée autour d'une vision hiérarchique de cette dernière.

Loin de développer l'idée que la caste est le produit d'un « fantasme » colonial (Bayly, 2001), cette historiographie permet de comprendre que la collecte et la compréhension des données reposent sur des paradigmes qui ont eu des effets sur la production du savoir colonial. Les anthropologues et les sociologues ont ensuite mobilisé ces sources de données même s'ils n'appartenaient pas au paradigme évolutionniste. C'est le cas, par exemple, de Max Weber, qui écrit à propos des documents d'Herbert Risley qu'ils « appartiennent à la meilleure littérature sociologique générale disponible » (Fuller, 2017, notre traduction), mais aussi de Louis Dumont (1974), qui développe un schéma de la caste fondé sur le modèle hiérarchique des varnas, où la dimension politique est moins importante que celle de la religion (Dirks, 2001).

### **c. Les difficultés classificatoires des identités de caste**

Si les questions sur l'appartenance de caste étaient ouvertes dans les recensements, les réponses étaient ensuite classées. Cette opération pose la question des limites de définition des groupes. D'une part, peu de castes possédaient une identité régionale et transrégionale et la prise en compte de la réalité locale de la caste n'est pas aisée. D'autre part, les frontières des castes ne sont pas fixes dans le temps et sont sujettes à des reconfigurations en fonction des relations sociales (Guilmoto, 1998). L'exercice de classification aurait alors institutionnalisé des identités en les consacrant comme traditionnelles, alors que leurs contours étaient auparavant fluides, changeants et ouverts. L'utilisation de la hiérarchie statutaire brahmanique de caste s'est en effet matérialisée dans l'ordre de classement des castes, comme par exemple dans la nomenclature utilisée dans la présidence de Bombay en 1872, où les Brahmanes étaient en position 1 et la sous-caste des Mahars (caste intouchable) était en position 147 (Appadurai, 1993).

Mais cette hiérarchie induite par les catégorisations de recensement a eu des effets performatifs sur le monde social (Desrosières, 2001). En effet, le recours à la classification des textes hindous dans le recensement a conduit à une mise en concurrence des castes pour leur place dans la nomenclature ainsi créée. Les jatis se sont recomposés et on a assisté à la formation de « méta-castes », qui consistent en un regroupement de jatis, de manière locale ou transrégionale, afin de revendiquer un statut de varna supérieur à celui de la nomenclature coloniale. Ces alliances de castes ont notamment débouché sur des

réclamations devant les tribunaux (Headley, 2013). Un exemple archétypique de ces alliances de castes est celui des Yadavs, qui regroupent les Ahirs, les Goallas et les Gopas, dans un but explicite de « sanskritization », c'est-à-dire de mobilité sociale collective par l'adoption de pratiques culturelles propres à des castes plus élevées dans la hiérarchie de varnas des castes (Michelutti, 2008). Loin d'être socialement neutre, le recensement a ainsi concrétisé administrativement et a institutionnalisé socialement la hiérarchie des varnas.

Cette tension entre le varna et la réalité anthropologique de la jati n'a pas disparu dans la période post-coloniale, mais un autre enjeu de classification s'opère. Pendant la période coloniale, les catégorisations de l'administration conduisaient à revendiquer un statut supérieur à celui accordé par l'administration, et à légitimer ce statut par une reconnaissance juridique. Dans la période contemporaine, la catégorisation des castes suivant la « politique des quotas » dans une logique de discrimination positive peut conduire à rechercher un statut « *backward* » reconnu par l'administration, et qui est particulièrement crucial pour l'accès aux établissements publics d'enseignement supérieur (Henry et Ferry, 2017). Les groupes de castes se sont dès lors mobilisés dès l'Indépendance pour la reconnaissance de leur statut de *Scheduled Castes* (SC) et de *Scheduled Tribes* (ST), et comme *Other Backward Classes* (OBC) dans les années 1980 (Lardinois, 1985). Pour poursuivre l'exemple des Yadavs, ils sont aujourd'hui reconnus comme des OBCs, un statut administratif « *backward* » étant compatible avec une revendication d'un statut de varna élevé. Ces mobilisations se poursuivent jusqu'à aujourd'hui, avec les récentes mobilisations des Marathas au Maharashtra, des Patels au Gujarat et des Jats en Haryana pour être reconnus comme OBCs (Deshpande et Ramachandran, 2017). Ironiquement, ces demandes sont justifiées par un retard socioéconomique, qui, pour être prouvé, nécessite la mobilisation de données statistiques de castes. Le rapport de la Commission Mandal (officiellement nommée la *Socially Backward Classes Commission*) de 1983 concluant à la nécessité de la création de la catégorie des *Other Backward Classes* avait ainsi mobilisé les données du recensement de 1931, où étaient recodées des questions ouvertes sur l'appartenance de caste.

Ce détour historique se justifie ici, non pas pour participer à la critique postcoloniale des institutions bureaucratiques coloniales, mais, plus humblement, pour éviter l'impasse du projet de quantification soutenu par « l'anthropologie coloniale ». Alors que de nombreux chercheurs

des sciences sociales participent à cette critique en rejetant la possibilité d'une connaissance statistique de la structure sociale des castes dans l'Inde contemporaine, nous nous éloignons de cette position. En fait, notre parti pris est de considérer que l'historiographie de l'appareil statistique colonial nous montre les écueils à éviter dans l'opération de quantification de la caste. Nous avons essayé ici d'esquisser les caractéristiques les plus problématiques, afin de ne pas les reproduire.

## 2. Un débat récurrent au sein des sciences sociales

Si les effets performatifs de la caste dans les recensements coloniaux sont reconnus dans l'ensemble des sciences sociales indiennes, certaines voix appellent tout de même à prendre en compte la caste dans les enquêtes statistiques actuelles. Cette position, à laquelle nous nous associons, vise à dépasser la critique postcoloniale de l'utilisation des catégories statistiques pour mesurer l'appartenance à une caste. En bref, l'objectif est de passer d'une pratique de non-catégorisation à une pratique de catégorisation soutenue par une quantification réflexive.

### a. Une complexité problématique

Le débat sur l'inclusion des castes dans les recensements a été relancé en Inde avec la création du *Socio Economic and Caste Census* de 2011 (SECC), suite à une proposition du *Registrar General of India* en 1998. Bien que des rapports de ce recensement aient été publiés, aucune donnée de caste autre que l'appartenance à une catégorie administrative des quotas (SC, ST, OBC et General pour les castes hors quota) n'est accessible publiquement<sup>2</sup>.

L'argument officiel avancé pour la non publication des résultats est le détournement possible de ces données à des fins discriminatoires. L'expérience de la mesure statistique de la religion en Inde suggère toutefois que cet exercice est également un moyen de surmonter les effets de la discrimination, tels que décrits par Deshpande et John (2010). D'autres raisons sont également avancées pour la non utilisation de catégories statistiques de castes. Celles-ci rappellent les débats sur l'utilisation des catégories ethniques dans le contexte français, qui soulèvent des problèmes techniques, mais aussi politiques et sociaux<sup>3</sup>.

2. Pour le recensement de 2001, nous connaissons cependant la composition de jati des *Scheduled Castes* et des *Scheduled Tribes*.

3. Nous faisons en particulier référence à la controverse telle que résumée par différentes positions dans le volume 49, n° 1 de la *Revue Française de Sociologie*, « L'usage des catégories ethniques en sociologie » (Felouzis, 2008).



D'un point de vue social, l'identité de caste renvoie à des affiliations multidimensionnelles, comme le rappelle Zoé Headley (2013), selon que le répondant préfère une réponse faisant référence à sa varna (qui n'est pas nécessairement acceptée par les autres castes), à sa jati (qui peuvent s'exprimer de plusieurs manières), ou à sa méta-caste (si elle existe), ces identités variant selon les régions et dans le temps, en fonction des relations sociales dans lesquelles les individus s'insèrent (Bayly, 2001).

Un argument souvent mis en avant par les détracteurs de la mesure des castes dans les recensements et, par extension, dans les grandes enquêtes par échantillonnages représentatifs, est que les personnes interrogées peuvent donner de fausses informations dans leur déclaration, dans une perspective utilitaire. Mais Satish Deshpande et Mary E. John (2010) écartent cet argument, en rappelant que le recensement est simplement une enquête publique obligatoire, indépendante du processus de réclamation des avantages de la politique des quotas, qui nécessite l'obtention d'un certificat de caste par l'administration publique. Cependant, l'argument est pertinent, car il est probable que des individus appartenant à des associations de castes mentionnent leur « méta-caste » plutôt que leur sous-caste (Yadav plutôt que Ahir par exemple). C'est donc aussi dans ce sens que le recensement, et toute enquête statistique référençant la caste, est un instrument sociologique pertinent pour étudier le monde social. Ils sont un moyen d'enregistrer les réponses que les individus veulent inclure dans les archives et les répondants mentionnent donc l'identité sociale dans laquelle ils se reconnaissent.

La réticence à prendre en compte la caste dans les enquêtes statistiques est finalement une critique de ce que les statistiques peuvent apporter à la compréhension de la société, notamment des processus de ségrégation et de discrimination, au-delà de la simple mise en évidence de corrélations statistiques avec des indicateurs d'inégalités sociales (Sundar, 2000). Sans entrer dans le débat, à la fois technique et épistémologique, concernant la distinction entre corrélation, causalité, discrimination et ségrégation, notons toutefois que nous devrions d'abord avoir des mesures d'association statistique entre la caste et d'autres variables pour discuter de la nature de leur relation. La méfiance à l'égard de l'analyse statistique apparaît ainsi liée à une position disciplinaire de la sociologie en Inde, proche de l'anthropologie, qui a peu investi dans les méthodes statistiques (Lardinois, 2007).

### **b. La pertinence de la quantification**

Les travaux existants en sciences sociales qui, par les méthodes quantitatives aussi bien que qualitatives, mettent en évidence des inégalités sociales fondées sur l'appartenance de caste rappellent toutefois la pertinence d'une prise en compte statistique de cette dernière (Deshpande, 2005). Parmi les travaux sur la caste mobilisant des catégories statistiques, il faut distinguer deux pôles : d'une part, ceux qui mobilisent les catégories administratives de la politique des quotas et, d'autre part, ceux qui utilisent des nomenclatures de caste plus précises, mais qui reposent sur des enquêtes géographiquement et socialement situées.

L'utilisation de catégories administratives (SC, ST, OBC) permet d'étudier les effets (positifs ou négatifs) de la « politique de quotas » au sein des établissements d'enseignement supérieur (Henry et Ferry, 2017) et de l'administration publique indienne (Benbabaali, 2008). Ces catégories peuvent également être mobilisées pour importer des questions et des méthodes d'étude autour de la discrimination et l'exclusion sociale issues des catégorisations raciales dans la sociologie anglo-saxonne (Thorat et Neuman, 2012), tout en tenant compte de la spécificité des modes d'exclusion dans la société indienne, en particulier ceux liés à la pratique de l'intouchabilité (Borooh, 2017). Ces catégories sont également utiles pour identifier les spécificités des conditions de vie des Dalits (correspondant approximativement à la catégorie SC), des Adivasis (catégorie correspondant aux ST) et des Brahmanes (avec les données du IHDS puisqu'une variable permet de les identifier grâce à une catégorie *ad hoc*). En revanche les OBCs ne correspondent pas à une identité de caste clairement identifiable. Ces catégories constituent donc une première étape pour étudier l'appartenance de caste, articulée avec celle de la religion et de la classe sociale, par exemple pour comprendre la structure sociale des modes de consommation (Ferry, Naudet et Roueff, 2018) ou les différents niveaux de fluidité sociale par caste dans la mobilité de classe intergénérationnelle (Vaid, 2018).

Cependant, une telle nomenclature devient inopérante pour rendre compte des descriptions plus fines de l'identité de caste. Il est en effet impossible de distinguer les hautes castes, les castes moyennes et les castes basses (sauf les castes intouchables et tribales), que l'on considère cette division du point de vue d'une hiérarchie rituelle ou socio-économique. Notamment, alors que le modèle de caste théorisé par Louis Dumont (1967) dans *Homo Hierarchicus* est critiqué pour

être trop orienté sur le statut rituel, ignorant ainsi le rôle de la caste dans l'appropriation des ressources, il reste difficile de tester des modèles théoriques concurrents (voir cependant Desai et Dubey, 2012, pour une tentative). Un modèle de caste basé sur une hiérarchie de statut à une dimension pourrait ne pas être approprié, ou du moins nécessiter des tests approfondis, dans la mesure où les hautes castes sont différenciées entre elles, en particulier celles qui placent les valeurs morales brahmaniques au premier plan par rapport aux valeurs morales concurrentes des Kshatriyas (Bayly, 2001). Par ailleurs, M. N. Srinivas (1952, 1959) a souligné le rôle des « castes dominantes » à l'échelle locale des villages, à savoir des castes de rang moyen ou supérieur rituellement, mais clés dans l'appropriation des ressources, résultant de leur domination agraire. Encore une fois, il est impossible d'objectiver cette réalité sociale en recourant à des enquêtes statistiques, à l'exception d'enquêtes locales (Himanshu, Jha et Rodgers, 2016)<sup>4</sup>. Les fractions de hautes castes sont également marquées par un accès différencié aux ressources sociales, ce que confirme l'analyse des trajectoires sociales permettant d'accéder à des positions de pouvoir économique. En effet, parmi les principaux dirigeants d'entreprise, les Brahmanes apparaissent plus proches de l'appareil étatique et doivent leur position de pouvoir à la possession d'un capital culturel institutionnalisé, le diplôme, tandis que les castes marchandes héritent d'empires économiques familiaux (Naudet, Allorant et Ferry, 2018). L'homophilie de caste apparaît également essentielle au développement de stratégies entrepreneuriales (Vissa, 2011), et est aussi très fortement présente dans le jeu politique indien, où l'appartenance à la caste détermine des « *vote-banks* ». L'analyse des appartenances de caste des élus indiens révèle ainsi l'affaiblissement du rôle des hautes castes en faveur des castes agricoles et des castes basses depuis les années 1950 (Jaffrelot et Kumar, 2012, Jaffrelot, 2010)<sup>5</sup>. L'importance des identités de caste dans la démocratie indienne exige alors la collecte de données complètes et précises sur les élites politiques indiennes, comme cela a déjà été fait dans le nord de l'Inde<sup>6</sup>.

4. Notons qu'une opérationnalisation statistique extrêmement prometteuse a été proposée par Iversen et al. (2010), en utilisant des données de l'IHDS, mais le codage des questions ouvertes, n'y est présenté que brièvement.

5. Pour un résumé des mouvements de castes agricoles dans l'Inde contemporaine, voir notamment Bayly (2001, chapitre 7).

6. Les données sont disponibles gratuitement auprès du Centre de données sociopolitiques : <https://cdsp.sciences-po.fr/fr/ressources-en-ligne/ressource/en.cdsp.ddi.NorthIndianMPs/>, et un projet combinant profondeur historique et une plus grande couverture géographique est également en cours : <http://www.sciencespo.fr/ceri/fr/content/sociologie-des-elus-nationaux-et-regionaux-du-raj-l-union-indienne-contemporaine>.

Cet aperçu non exhaustif montre l'importance des identités de caste dans l'étude de la structure sociale indienne, bien que les travaux de recherche utilisant des catégories de castes fines se limitent à des espaces limités sur le plan social et géographique. La quantification de la caste dans des catégories précises est souvent opérationnalisée grâce au codage manuel des noms de famille, malgré l'existence de stratégies (toutefois limitées) de « désidentification » de caste par des changements de noms (Copeman, 2015). Les informations sont parfois recoupées manuellement avec les biographies existantes pour confirmer les informations fournies par le nom. Une opération aussi fastidieuse, nécessitant beaucoup de temps et de ressources, ne peut toutefois être mobilisée sur des enquêtes statistiques à l'échelle de grandes zones géographiques avec des répondants anonymes.

### 3. Catégoriser la diversité des identités de caste

Les données d'enquête auprès des ménages utilisées ici ont pour but de fournir une image représentative de la répartition de la population de l'Uttar Pradesh. Une revue rapide de la littérature montre que les questions ouvertes de caste de l'Indian Human Development Survey (IHDS) et du National Family and Health Survey (NFHS) n'ont jusqu'à présent pas été recodées de manière systématique et nous présentons donc notre propre méthode de classification.

#### a. Le format des questions ouvertes

Dans l'IHDS-I (2004-2005), il existe une question ouverte sur l'appartenance de caste : « À quelle caste appartenez-vous ? » (« *Which caste do you belong to?* », ID12). Dans le NFHS-I de 1992-1993, les réponses à la même question ont été recodées en utilisant la classification (très détaillée) du recensement de 1931. Mais comme nous l'avons déjà noté, des critiques sur la pertinence de cette classification nous retiennent de la mobiliser. Dans le NFHS-II (1998-1999) et le NFHS-III (2005-2006), les réponses n'ont pas été recodées et sont laissées en format brut. Bien que l'enquête NFHS-IV (2015-2016) comprenne également une question sur l'appartenance de caste, les réponses à cette question n'ont pas encore été publiées. Contrairement aux enquêtes précédentes, cette dernière série d'enquêtes NFHS est représentative au niveau du district<sup>7</sup>, ce qui pourrait être la cause de la rétention de ces données, les identités de caste étant considérées comme sensibles

7. Un district est une division administrative géographique, plus localisée qu'un État. En 2011, le recensement dénombrait 640 districts en Inde.



(Deshpande et John, 2010). Notons que le NFHS pose ces questions aux individus (âgés de 15 à 59 ans) dans les ménages et non au répondant du ménage (souvent le chef de ménage) comme dans l'IHDS. Cela pourrait s'avérer utile à un stade ultérieur dans l'évaluation de la variabilité de l'énonciation des castes parmi les membres du ménage. Dans l'IHDS-II (2011-2012), deux questions ouvertes sont posées: « À quelle caste / jati et sous-caste / sous-jati appartenez-vous ? » (« *Which caste/jati and sub caste/sub jati do you belong to?* », ID12aNM et ID12bNM). Le tableau 1 présente un extrait des réponses dans la base de données. « Sous-caste » ou « sous-jati » sont ici des artefacts administratifs qui n'ont pas de sens réel, mais visent à préciser de manière plus précise la jati des répondants, qui l'utilisent parfois pour affirmer leur statut de varna. Le doublement de la question de l'appartenance à une caste témoigne de la complexité des identités de caste, que nous avons souligné. Étant donné que la complexité est souvent utilisée comme une des raisons pour ne pas compter la caste, nous partons de ces données pour la classer.

**Tableau 1** - Extrait des variables « jati » et « sub-jati » de la base IDHS-II

ID12ANM	ID12BNM
...	...
YADAV	AHIRI
YADAV	AHIRJABAL
CHAMAR	AHIRVAR
JATHA	AHIRWAR
CHAMAR	AHIRWAR
YADAV	AHRI
ARAKH	AKUR
DHOBI	AMRI
DHOBI	AMRI
ARAKH	ARBANLI
PANJABI	ARORA
KORI	ASTI
KORI	ATARI
BRAHMIN	AVASTHI
...	...

### ***b. Pistes méthodologiques dans la littérature***

Puisque les castes témoignent d'une forte variabilité régionale sur le sous-continent indien, Deshpande et John (2010) suggèrent que, dans l'hypothèse où l'appartenance de caste ferait l'objet d'une question

ouverte dans le recensement, la construction d'une nomenclature devrait partir d'une échelle géographique locale, idéalement du district. Leur étude statistique suggérée de la caste devrait donc être inductive au sens géographique. Le recensement étant exhaustif, il permettrait d'établir un tableau statistique extrêmement détaillé des identités de caste. Cette méthode est moins pertinente dans le cas d'enquêtes par échantillonnage puisqu'elles ne sont ni exhaustives ni représentatives au niveau des districts, comme dans l'IHDS-II, qui n'est que représentative à l'échelle du district. Cette variabilité régionale de la caste nous amène ici à nous concentrer sur un État indien en particulier. Sonalde Desai (2010), l'une des organisatrices de l'enquête IHDS, suggère un autre cadre d'analyse des castes, en supposant qu'une question serait intégrée au recensement de 2021. Elle suggère de s'inspirer des nomenclatures de professions emboîtées, en particulier celle utilisée en Inde, la National Classification of Occupations (NCO)<sup>8</sup>. Les regroupements entre castes se feraient à différents niveaux de la nomenclature, ces niveaux intégrant une dimension spatiale. Elle suggère de construire la nomenclature *a priori* en utilisant les données du recensement de 1931, les questions ouvertes de l'IHDS et les annonces matrimoniales, qui mentionnent souvent la caste. La création d'une telle nomenclature nécessiterait un travail de collaboration approfondi, à l'image de la commission nommée par le gouvernement indien pour recoder les 4 673 034 énonciations différentes de caste du SECC de 2011<sup>9</sup>.

L'utilisation actuelle des questions ouvertes de caste dans l'IHDS et le NFHS ne cherche pas à établir une nomenclature de caste. L'exercice parfois mené consiste à recoder des parties de l'échantillon afin d'examiner les conditions de vie de certaines castes en particulier, celles considérées comme archétypiques des castes incluses dans les catégories administratives des politiques de réservation par exemple (Srinivasan et Kumar, 1999, exercice réalisé à l'aide des données du NFHS). L'autre perspective, plus récente, consiste à examiner la situation socioéconomique de certaines castes qui revendiquent un statut « *backward* » et qui se mobilisent pour être officiellement considérées comme des OBCs. A. Deshpande et R. Ramchandran (2017) ont ainsi utilisé les données de l'IHDS pour étudier les Jats (en Haryana), les Patels (au Gujarat) et les Marathas (au Maharashtra). A. Kalaiyarasan (2016)

8. Cette nomenclature est très largement dérivée de l'International Standard Classification of Occupations (ISCO) de l'Organisation Internationale du Travail.

9. <http://indianexpress.com/article/india/india-others/cabinet-meeting-pa-nagariya-to-head-panel-to-classify-caste-census-data/>.

se concentre sur les Jats dans l'Haryana et A. A. Dongre (2017) étudie plus particulièrement les Marathas dans le Maharashtra.

Ces deux derniers articles présentent rapidement leur méthode de codage, qui consistait à sélectionner les ménages et les individus en fonction de leur appartenance déclarée à ces castes. Cela n'est cependant pas si évident, dans la mesure où plusieurs orthographes différentes pour une caste sont utilisées. Nous pouvons supposer que ces différentes orthographes ne signifient rien en elles-mêmes et ne sont que le résultat de l'opération de traduction et de transcription en alphabet latin. En effet, dans le cas des Jats, les différentes transcriptions sont: «JAAT», «JAHT», «JAT», «JAT SIKH», «JATH», «JATT», «JHAT» et «RON JAT». Dans le cas des Marathas étudiés par Ambish A. Dongre, son recodage est un peu plus complexe: « Les ménages qui signalent leur jati comme étant Maratha ont été définis comme des ménages Marathas. Leurs sous-jatis peuvent être soit Maratha, Patil, Kshatriya ou Rajput. Dans certains cas, aucune sous-caste n'est mentionnée. Les cas où la jati est Maratha mais les sous-jatis sont Agri, Kunbi ou Hatkar ne sont pas classés comme Maratha, mais comme OBC, conformément aux réglementations administratives »<sup>10</sup>.

Cela souligne la difficulté du codage. Tout d'abord, ce que Dongre considère comme une « sous-jati » dans le cas des Marathas correspond à l'affirmation d'un statut rituel lorsque les répondants se déclarent « Kshatriyas » (la deuxième varna la plus élevée) ou « Rajputs » (une jati revendiquant la varna Kshatriya), alors que les Marathas sont souvent considérés comme des « Shudras » (quatrième varna, de rang beaucoup moins élevé). Ces « sous-jatis » pourraient faire référence au rôle traditionnel militaire des Marathas, ou bien à une « rajputisation » ou à une « kshatriyisation » de leur appartenance de caste (Lardinois, 2005), un mécanisme social homologue à la « sanskritization ». Deuxièmement, nous observons que les énonciations de « jati » et de « sous-jati » correspondent plutôt à une gradation de l'affirmation de l'identité (les répondants sont d'abord Marathas, puis Kshatriyas), plutôt qu'à une logique de nomenclature emboîtée. Troisièmement, certains ménages revendiquent l'identité de caste Maratha, mais ils précisent également qu'ils appartiennent à un jati classée comme OBC, revendiquant à la fois l'identité Maratha et un statut « backward », ce qui est contre-intuitif puisque les Marathas ne sont pas classés

10. "The households who report their jati to be Maratha were defined as Maratha households. Their sub-jati can be either of Maratha, Patil, Kshatriya, Rajput. In some instances, no sub-caste is mentioned. Instances where jati is Maratha but sub-jati is Agri, Kunbi, Hatkar are not categorized as Maratha but as OBCs, in line with government rules.", notre traduction.

OBCs. Dans l'ensemble, les précisions de Dongre rappellent donc que les identités de caste sont à la fois fluides et multidimensionnelles.

### c. Stratégie empirique

Ces remarques sont extrêmement utiles et en même temps déconcertantes pour construire une classification. Les identités de caste énoncées dans les questions de « jati » et de « sous-jati » correspondent-elles à des régularités pouvant être mesurés statistiquement ? Ou bien ces énonciations de caste soulignent-elles simplement l'importance des identités de caste et la renforcent comme complexe, impossible à saisir dans une seule classification ?

#### i. Dépasser la variabilité orthographique

Le décompte des déclarations de caste dans l'IHDS-II montre qu'au niveau indien, 8 318 déclarations de « jati » et 9 163 déclarations de « sous-jati » sont présentes dans la base de données et que, lorsqu'elles sont combinées, 18 425 façons d'exprimer sa caste existent donc, alors que l'échantillon comprend 42 152 ménages. Lorsque l'on restreint l'échantillon aux hindous de l'Uttar Pradesh, qui comprend un échantillon de 2 958 ménages, 755 déclarations de « jati », 878 déclarations de « sous-jati » et 1 603 combinaisons différentes apparaissent dans les données.

Cependant, cette diversité apparente des identités de caste est surestimée par la variabilité orthographique des énoncés. Ces différences dans l'orthographe peuvent être liées à la transposition en alphabet latin des différentes langues du questionnaire (11 langues de passation au total), à des orthographes multiples du même terme selon la convention adoptée en alphabet latin et à des erreurs orthographiques lors de la transcription.

Après avoir nettoyé les chaînes de caractères de tous les signes « polluants » (tels que '[' ou '/'), nous procédons à une correction orthographique. Nous mobilisons des algorithmes de regroupement de chaînes basés sur la similarité des séquences de caractères. Ces algorithmes sont fondés sur la méthode des « proches voisins », avec l'algorithme « Fingerprint » et la distance Levenshtein, mobilisés dans *OpenRefine* (Verborgh et Wilde, 2013)<sup>11</sup>. Cette technique repose sur la similarité de deux chaînes de caractères, définie par le nombre minimum de caractères à supprimer, à insérer ou à remplacer pour transformer une

11. Le logiciel propose une liste de regroupements qui sont ensuite acceptés ou rejetés par l'utilisateur.

chaîne en une autre<sup>12</sup>.

Cette méthode permet de diminuer considérablement le nombre d'énonciations de caste distinctes dans la base de données. Le nombre de déclarations distinctes de « jati » chute à 425, le nombre de « sous-jati » à 589 et un total de 1 095 combinaisons de castes est alors présent dans les données. Cette étape réduit donc la variabilité des combinaisons de castes de près d'un tiers (32,5%).

## ii. Visualiser la complexité de la caste avec l'analyse de réseau

Ensuite, nous proposons de considérer les énoncés de « jati » et « sous-jati » comme une liste de liens (edgelist), ce qui est le point de départ pour construire un réseau, où les nœuds (les sommets) sont les déclarations de « jati » et de « sous-jati » ( $2 * 2\,928$  nœuds maximum), qui sont connectés par les ménages qui les ont co-énoncés (2928 liens maximum différentes). Les liens du réseau sont pondérés de telle sorte qu'une énonciation « jati »-« sub jati » est considérée plus importante si elle est plus fréquente dans la population<sup>13</sup>.

Théoriquement, si le système de castes formait un système parfaitement emboîté, nous devrions observer un réseau similaire au schéma de la Figure 1. Le réseau devrait former un ensemble de sous-graphes déconnectés. Chaque sous-graphe (dont nous ne prévoyons pas le nombre *a priori*) formerait un réseau en étoile, avec une jati au centre connectée à plusieurs jatis. Un exemple de réseau pourrait être le sous-graphique des Brahmanes, avec la « jati » des Brahmanes au centre et les « sous-jatis » Trivedi, Chaturvedi, Tiwari, etc. à la périphérie. On pourrait donc simplement considérer dans la classification les sous-graphes déconnectés comme unités, nommées d'après la « jati » centrale dans chaque sous-graphe.

En pratique cependant, la représentation empirique du réseau formé par les ménages hindous dans l'Uttar Pradesh (Figure 2) diffère de ce

12. Une seconde méthode basée sur un « algorithme phonétique », en transformant les chaînes de caractères en « chaînes de prononciation » a été testée. Dans cet algorithme, deux chaînes sont identiques si elles sont prononcées de la même manière. Nous nous inspirons ici de Raphael Susewind (2015) qui utilise une version modifiée de l'algorithme « Soundex », adapté aux langues vernaculaires indiennes par Santhosh Thottingal, pour identifier les noms de famille musulmans sur les listes électorales. Voir le site Web de Thottingal pour plus de précisions : <http://thottingal.in/blog/2009/07/26/indicsoundex/>. L'utilisation de cet algorithme n'étant pas aussi performante dans les tests de classification que l'algorithme des chaînes de caractères, en particulier du point de vue de la comparaison avec le « *gold standard* » (voir section suivante), nous préférons nous baser sur la première méthode.

13. Pour faciliter la lecture des diagrammes (Figures 2 et 4), la largeur des liens n'est pas tracée en fonction de la fréquence de déclaration, mais certaines sont nettement plus courantes et ce sont en fait les instructions les moins fréquentes qui relient les sous-réseaux en étoile.

schéma théorique. Plutôt qu'une accumulation de sous-graphes déconnectés, on observe un grand sous-graphe connecté avec plusieurs sous-graphes plus petits (paires de nœuds). Le nombre de nœuds de ce réseau total est égal à 819 (il y a donc 819 déclarations de caste différentes, que ce soit « jati » ou « sous-jati »), alors que le nombre de liaisons différentes entre deux nœuds est égal à 917 (le nombre de ménages déclarant une combinaison « jati » et « sous-jati » différente).

Cette projection empirique ne nous surprend pas compte tenu de la complexité des identités de caste. Mais en examinant attentivement le réseau, nous retrouvons partiellement la structure théorique de la figure 1. Par exemple, le nœud « Brahmin », central dans le sous-graphe le plus grand, est connecté à un ensemble de nœuds, y compris les « jatis » mentionnés précédemment. La structure du réseau en étoile, reflétant les identités de caste comme un ensemble de catégories imbriquées, n'est donc pas absente du réseau, mais la réalité montre encore une énonciation plus complexe.

Les critiques du comptage de la caste soutiennent que la caste est un phénomène social trop complexe pour être pris en compte dans les catégories statistiques et il est clair que si la structure d'énonciation des identités est bien plus compliquée qu'une simple tabulation statistique, le réseau permet de la prendre en compte.

## iii. La structure du réseau

Si la représentation des identités de caste dans un réseau rend justice à la complexité du système de castes, nous observons également que les relations entre les relations identitaires identifiées sont structurées, puisque les réseaux en étoile ne sont pas absents du diagramme.

Le réseau constitué ici est alors un artefact qui permet d'identifier la structure des co-identités de caste, en utilisant le réseau comme un « système d'interdépendances » (Lazega, 2014) des identités de castes. Nous cherchons donc ici à identifier les régularités de ce système, par l'identification de « blocs » de sous-réseaux. Notre approche de classification est inductive, en ce sens qu'elle cherche à mettre en évidence les régularités de l'association telles qu'elles sont énoncées dans la population, plutôt que d'imposer des groupements *ad hoc*. Une telle utilisation de l'analyse de réseau est similaire à l'algorithme d'identification ethnique de Mateos, Longley et O'Sullivan, (2011). Dans leur cas, le partage du même nom ou du même prénom forme les liens entre les individus. Cette méthode de classification des groupes ethniques,



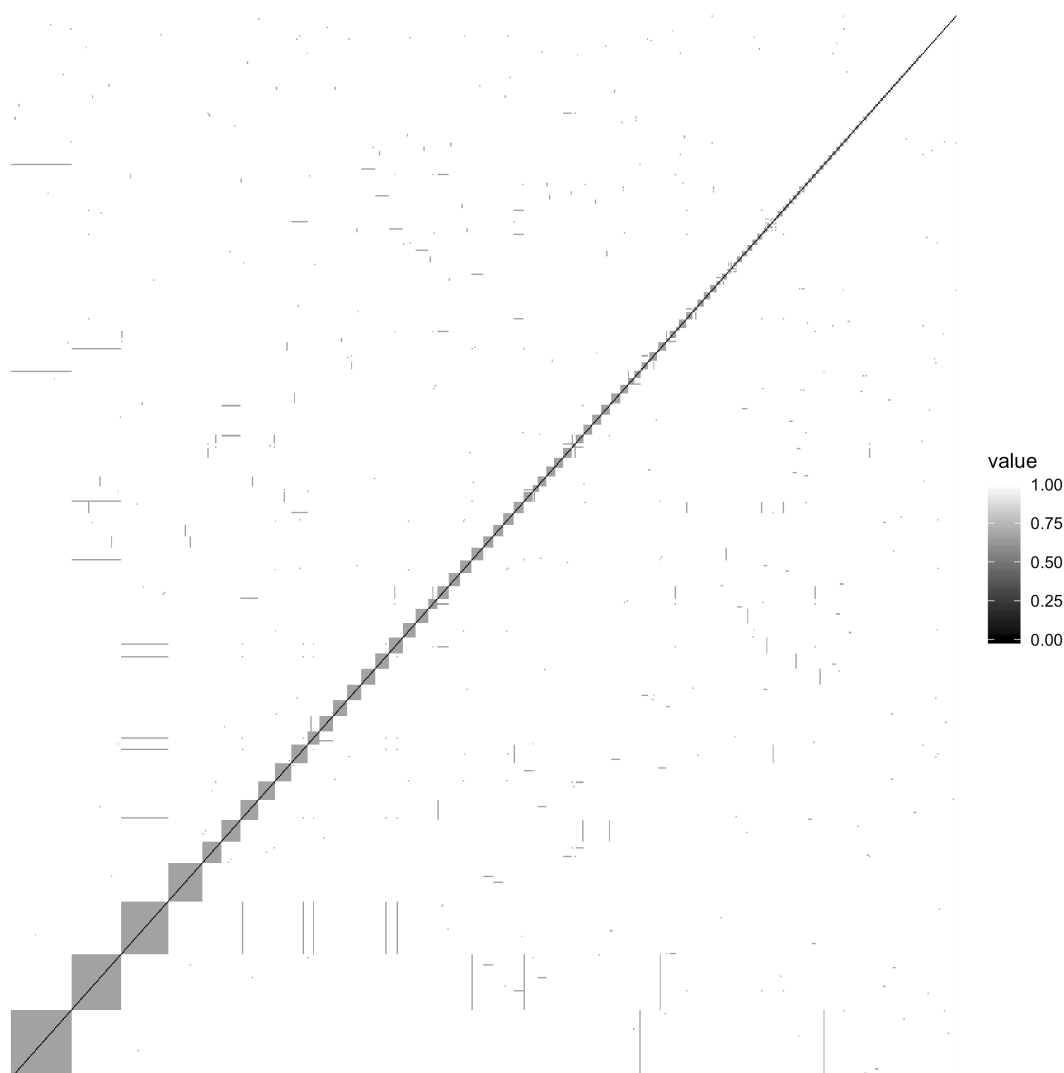
ou, dans notre cas, des groupes de castes, a l'avantage d'identifier la structure des régularités des énoncés plutôt que de faire correspondre une déclaration *a priori* à un groupe identifié par un répertoire existant, ce qui pourrait être une procédure alternative. De plus, nous n'avons pas de noms de famille ici, auquel cas nous aurions pu utiliser la méthode proposée par Vissa (2011) ou Susewind (2015) en utilisant un site Internet matrimonial identifiant le nom et la caste (le site matrimonial est alors utilisé comme répertoire des noms de caste).

Avant toute chose, on peut tester dans quelle mesure le réseau présente des « blocs » significatifs, c'est-à-dire tester si le réseau est structuré. L'inspection visuelle du réseau a tendance à répondre à cette question par l'affirmative (avec les sous-réseaux en étoile), et la représentation graphique de la matrice de dissimilarité ordonnée en fonction des ménages les plus proches va également dans ce sens

(figure 3). Cette matrice de dissimilarité est construite à partir du coefficient de Jaccard, qui calcule la distance entre deux ménages (liens), en fonction du partage des mêmes énoncés de caste (nœuds)<sup>14</sup>. Ce graphique est une évaluation visuelle de la tendance de regroupement (Visual Assessment of clustering Tendency, VAT, Bezdek et Hathaway, 2002). En noir, les ménages les plus similaires (la distance est égale à 0), et en blanc les ménages les plus éloignés les uns des autres (la distance est égale à 1). La diagonale (croisant des ménages identiques) est nécessairement en noir, le même ménage étant similaire à lui-même. Sur la représentation graphique, nous voyons des agglomérations similaires de ménages qui semblent émerger: ainsi, la structure du réseau de castes suggère que les ménages sont organisés autour de nœuds communs, mais que le nombre de « blocs » est relativement important.

14. Précisément, l'indice calcule le nombre de nœuds partagés entre deux liens.

**Figure 3** – Évaluation visuelle de la tendance de regroupement dans le réseau de caste



Dans la littérature de l'analyse de réseaux, de nombreuses méthodes existent pour classer et identifier les « blocs ». Nous voulons ici obtenir des « blocs » de ménages (qui sont les liens), de manière à chercher à construire des blocs à partir de similitudes entre les liens du réseau, plutôt qu'à partir de la similarité des nœuds. Ce sont en effet les ménages qui sont *in fine* classés dans des catégories de caste<sup>15</sup>.

Nous mobilisons ici une classification ascendante hiérarchique (avec la méthode de la moyenne) pour créer une partition du réseau en communautés statistiques. Le dendrogramme de la classification est coupé de telle sorte que la partition finale maximise la densité de partition (*density partition*)<sup>16</sup>, ce qui équivaut à une minimisation de la variance intra-communauté, en d'autres termes de la maximisation de l'homogénéité des communautés. À partir de cette partition, les ménages non classés (en particulier ceux qui n'étaient pas des liens dans le réseau puisqu'ils n'ont mentionné qu'un seul énoncé « jati », et pas de « sous-jati ») sont classés dans une des communautés avec laquelle ils partagent un énoncé de « jati » commun. Enfin, les communautés sont nommées par l'énoncé de caste du nœud le plus central de chaque communauté.

#### iv. Des communautés statistiques à la nomenclature

L'algorithme de classification automatique identifie 79 communautés statistiques (77 lorsque l'homonymie est prise en compte), présentés dans le tableau en annexe. Ce nombre élevé de communautés était attendu compte tenu de la VAT (Figure 3) et permet de rendre compte des identités de caste les plus saillantes autour desquelles les interrogés s'identifient. En regardant la structure du réseau des communautés, elles se révèlent avoir une forme de réseau en étoile. Les nœuds les plus centraux peuvent être identifiés à des jatis connues dans la littérature de sciences sociales et les énoncés de caste de chaque groupe sont associés de manière cohérente (par exemple, Brahmin est bien associé à Tiwari et Mishra qui forment des sous-castes de Brahmanes).

Nous choisissons ensuite de regrouper les communautés manuellement dans une nomenclature en sept catégories. Cette nomencla-

15. L'algorithme mobilise le package R linkcomm (Kalinka et Tomancak, 2011) qui peut être consulté à cette adresse : <https://cran.r-project.org/web/packages/linkcomm/vignettes/linkcomm.pdf>.

16. La densité du réseau (network density) est le nombre de connexions réelles dans le réseau divisé par le nombre de connexions potentielles. La densité de la partition (partition density) est la densité des liens au sein des groupes de la partition, normalisée par rapport au nombre maximum et minimum de liens possibles dans chaque groupe (Ahn, Bagrow et Lehmann, 2010).

ture ne prétend pas épuiser la variabilité des identités de caste parmi les hindous de l'Uttar Pradesh, mais au moins de mettre en évidence les groupes importants que les enquêtes statistiques ne peuvent pas prendre en compte jusqu'à présent. On peut soutenir que cette dernière étape s'écarte d'une approche purement inductive mais elle se justifie pour trois raisons. Premièrement, le nombre encore élevé de communautés empêche toute analyse statistique compte tenu de leur taille (par exemple, la communauté des Lohars ne représente que 0,06% de la population hindoue de l'Uttar Pradesh, et l'échantillon de l'enquête est relativement réduit (2 928 ménages)). Deuxièmement, une analyse statistique qui mobiliserait un si grand nombre de catégories entraverait considérablement la lisibilité des résultats. Troisièmement, une partition avec un nombre plus faible de communautés dans la classification ascendante hiérarchique serait plus inductive dans son approche que des regroupements manuels mais elle serait statistiquement peu justifiée dans la mesure où elle ne respecterait pas le critère d'homogénéité des communautés.

Le rôle de l'algorithme automatique est alors d'avoir permis de réduire la complexité des identités de caste, que nous complétons par une procédure de classification manuelle fondée sur les noms des communautés statistiques. Cette nomenclature est largement fondée sur la réputation des identités des jatis et s'appuie sur la littérature de sciences sociales qui étudie les castes. En particulier, la synthèse historique de la caste de Susan Bayly (2001) motive les regroupements proposés (tableau 2). Même si elle n'essaie pas de construire un schéma de caste, sa discussion historique fondée sur des travaux empiriques, rendant compte de l'évolution temporelle des identités de caste, motive notre schéma. Les catégories de caste sont construites selon quatre dimensions qui se recoupent : l'affirmation d'un statut rituel, les valeurs morales et la concurrence entre statuts, les positions économiques et de domination, la conscience et l'identité collective de caste.

Premièrement, la caste peut être identifiée à une opposition entre des groupes qui incarnent des valeurs pures par rapport à d'autres qui incarnent des valeurs impures. Alors que les Brahmanes incarnent des valeurs pures qu'ils affichent à travers leurs styles de vie (par exemple dans l'alimentation avec l'abstinence de consommation de viande), les castes intouchables sont considérées comme moralement particulièrement impures et polluées. Cette logique statutaire rituelle suit la quadripartition hiérarchique des varnas. Mais deuxièmement,



Susan Bayly rappelle également la concurrence historique entre les valeurs Brahmanes et les valeurs Kshatriyas dans la domination de l'ordre moral. Bien que les premières soient devenues dominantes au 18<sup>ème</sup> siècle, les secondes n'ont pas disparu et sont encore incarnées par certains groupes de castes, les castes nobles (correspondant aux castes Rajputs et affiliées, incarnant les valeurs traditionnelles nobles et conquérantes). Troisièmement, elle soutient que cette concurrence

entre valeurs morales est liée aux positions économiques et politiques des groupes de castes dans la société. La domination progressive des valeurs morales des castes Brahmanes correspond historiquement à leur rôle croissant dans les grandes administrations dirigeantes (l'empire Marathe, l'empire Moghol, les administrations coloniales et plus tard l'empire britannique). Cette position ascendante était liée à leurs professions traditionnelles en tant que prêtres et enseignants, ce qui

**Tableau 2** – Les sept catégories de la nomenclature de caste

Catégorie	Distribution (%)	Description
Brahmanes	11.95	Hautes castes considérées au sommet de la hiérarchie des castes dans le modèle théorique de l' <i>Homo Hierarchicus</i> , mais cette population n'a pas nécessairement incarné des valeurs morales dominantes à toutes les périodes. Il est toutefois indéniable que les Brahmanes sont devenus dominants dans l'ordre social des castes, en particulier à partir du 18 <sup>e</sup> siècle, lorsque les frontières de la pureté et de l'impureté ont été renforcées et accentuées.
Castes nobles	11.53	Hautes castes identifiées comme guerrières, nobles, propriétaires terriennes et centrées sur l'identité Rajput, elles incarnent les valeurs morales Kshatriya, en concurrence avec les valeurs Brahmanes.
Kayasths	1.23	Caste lettrée, composée de groupes davantage urbains, et qui ont historiquement déteu des positions de hauts fonctionnaires dans les administrations publiques. Cette caste est proche des valeurs morales Brahmanes, bien que ses membres revendiquent parfois le statut de varna Kshatriya.
Castes marchandes	5.6	Traditionnellement des communautés marchandes, dont l'identité centrale est le titre de « Baniya ». Ces hautes castes ont historiquement adopté des valeurs sanskritisées (Brahmanes), tout en participant à leur diffusion. Elles revendiquent un statut de varna Vaishya. Les Panjabis (migrants hindous du Panjab arrivés en Inde suite à la Partition indo-pakistanaise de 1947) sont ajoutés à cette catégorie étant donné les positions économiques traditionnellement similaires.
Castes agricoles	19	Communautés identifiées comme petits et moyens propriétaires terriens, qui se sont mobilisées à partir des années 1920 sous la figure du « kisan » (agriculteur) et revendiquent une varna supérieure à celle dans laquelle ils sont habituellement assignés (Shudras). Certaines communautés ont, à partir des années 1980, revendiqué (parfois avec succès) un statut « backward » dans les politiques de réservation.
Basses castes	20	Rassemblement de communautés historiquement constituées de petits artisans, de petits ouvriers agricoles ou de paysans sans terre. Cette catégorie regroupe les castes qui ont un statut rituel inférieur (considéré comme Shudra), mais qui sont néanmoins plus hétérogènes dans leurs professions, valeurs et modes de vie que les autres catégories de la nomenclature.
Dalit	28.26	Groupe de jatis considérés comme populations intouchables ou tribales (les tribus répertoriées, <i>Scheduled Tribes</i> ne représentent cependant que 0,1% de la population de l'Uttar Pradesh selon le recensement de 2011, ce qui justifie leur regroupement avec les Dalits). Composée de différentes jatis, la catégorie est unifiée par un statut rituel assigné inférieur et une conscience de caste autour de l'identité Dalit.

Note : Huit communautés statistiques représentant 2,43% des hindous de l'Uttar Pradesh n'ont pu être identifiées selon leurs noms (ces noms ont été recherchés dans la littérature, dans les répertoires coloniaux et dans un projet ethnographique de descriptions de caste, *People of India*). L'appellation des catégories s'inspire des distinctions de caste proposées par Bayly (2001).

faisait des Brahmanes un groupe davantage lettré par rapport au reste de la société. C'est également le cas des Kayasths qui occupaient des postes similaires. Dans le même temps, les castes nobles Rajputs ont pu conserver un certain pouvoir politique, grâce à la domination de petits royaumes et de domaines agricoles. Quatrièmement, la domination progressive des valeurs Brahmanes dans la société a renforcé la diffusion du système des varnas, son acceptation par l'ensemble de la société et l'auto-identification des jatis à des varnas. Cela correspond à une « substantialisation » de la caste pour accepter, affirmer et maintenir collectivement la pureté rituelle d'une jati et pour revendiquer un statut de varna (Dumont, 1967), comme on peut l'observer avec les castes marchandes (regroupées autour de l'identité Baniya) qui revendiquent un statut de haute caste Vaishya. Les mobilisations de castes qui ont émergé au 19<sup>ème</sup> siècle et se sont poursuivies jusque dans la période contemporaine ont aussi constitué un moyen d'affirmer un statut de varna, comme nous l'avons noté dans la première section avec la formation de « méta-castes », et ont accru la conscience de caste de groupes précédemment désagrégés. C'est le cas de groupes marginalisés, occupant des positions agricoles (mobilisées sous la figure du « kisan », qui signifie paysan en hindi) et habituellement considérés comme Shudras, qui ont affirmé leur conscience de caste par un regroupement identitaire et une mobilisation politique. Dans une certaine mesure, les castes intouchables ont également évolué dans cette direction avec l'émergence d'une identité Dalit, pour contrer leur ascription statutaire impure. Cette conscience de caste s'affirme donc soit dans la contestation du statut rituel en tant que tel, soit dans la revendication d'un statut rituel supérieur.

Cette prise en compte de quatre dimensions de l'identité de caste s'écarte clairement d'une vision de la caste en tant qu'entité traditionnelle assignée et reconnaît les processus historiques et sociaux de transformations des identités de caste. Par conséquent, notre catégorisation rompt avec le schéma de caste idéalisé d'*Homo Hierarchicus*, centré sur le statut rituel des textes sacrés, un reproche fait à Dumont (1974), notamment par Dirks (2001). Nous intégrons plutôt deux principes hiérarchiques, le statut rituel et le statut socio-économique, ainsi que leurs contestations. Nous sommes conscients que la définition conceptuelle de la caste a fait l'objet de débats très anciens (Lardinois, 1985). En suggérant ce schéma de caste, nous prenons une position théorique dans le débat, mais nous pensons qu'un schéma historiquement informé est la meilleure voie à suivre et l'objectif de la section suivante est d'évaluer

dans quelle mesure notre schéma saisit correctement nos dimensions conceptuelles.

#### 4. La pertinence de la nomenclature de caste

La nomenclature proposée est à notre connaissance la première qui vise à opérationnaliser la caste dans une enquête représentative de la population à une échelle géographique aussi large. Sa conceptualisation multidimensionnelle reflète différents aspects de la caste. Afin de renforcer sa pertinence, nous examinons la fécondité de notre schéma sous trois angles différents. Tout d'abord, nous souhaitons savoir si la nomenclature, bien que simplifiant la complexité des castes représentée dans le réseau, reflète bien les frontières sociales de l'auto-identification à la caste. Deuxièmement, si tel est le cas, nous visons à évaluer si la classification catégorise des identités de caste « réelles ». Troisièmement, nous souhaitons aborder la définition conceptuelle de la nomenclature en établissant des tests de validité préliminaires.

##### a. La structure du réseau et la nomenclature

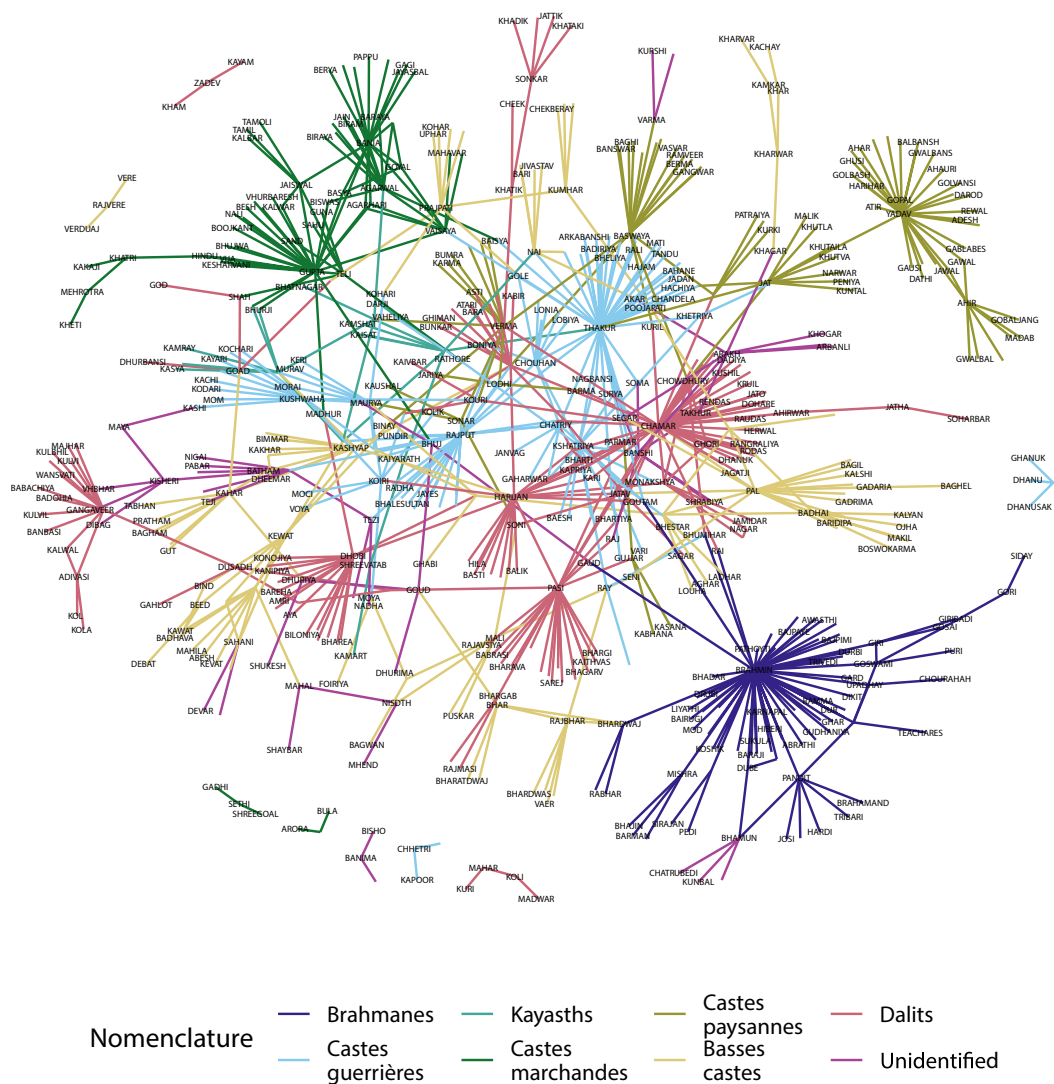
Visuellement, on peut inspecter le réseau de castes où les liens ont été colorés selon les catégories de la nomenclature (Figure 4), ainsi que le tableau récapitulatif des communautés statistiques (tableau en annexe). Les groupes de castes de la nomenclature sont proches, si on comprend par proximité les déclarations de caste partagées entre les communautés. Le groupe des Brahmanes (12% des hindous de l'Uttar Pradesh) est centré sur la communauté statistique « Brahmane ». En ce qui concerne les castes nobles (11,5%), elles sont marquées par la communauté statistique « Rajput », même si elle est moins centrale dans la catégorie. Les Kayasths (1,2%) ne regroupent que trois communautés statistiques, dont la communauté « Srivastav » constitue le nœud central. Parmi les castes marchandes (5,6%), la communauté « Gupta » est numériquement la plus importante et la grappe « Bania » est la plus centrale. Les castes agricoles (19%) sont caractérisées par une prédominance de la communauté « Yadav ». Au sein de la catégorie des basses castes (20%), qui regroupe des communautés statistiques plus dispersées dans le réseau, ainsi que des noms de communautés relativement éloignées culturellement, on remarque que les trois pôles de ce groupe de nomenclature sont les communautés « Kewat » « Rajbhar » et « Pal ». Enfin, parmi les Dalits (28,3%), qui regroupent des communautés dont les noms sont considérés comme intouchables ou tribaux,

la communauté « Chamar » constitue le groupe le plus important, suivi de « Pasi » et de « Dhobi ». Notons que le nom de ce groupe (« Dalit », qui signifie « homme brisé ») correspond à une convention adoptée dans la littérature, reprenant ainsi un terme popularisé par le leader Dalit Ambedkar (Jaffrelot et Naudet, 2013). Mais dans l’Uttar Pradesh, aucun des ménages enquêtés n’a utilisé ce terme pour identifier sa jati, ce qui est probablement lié à une « ethnicisation »<sup>17</sup> plus faible des basses castes du nord de l’Inde, pour lesquelles la mobilité ascendante collective a toujours été plus marquée par la « sanskritisation » (Jaffrelot, 2000a).

17. Le terme d’« ethnicisation » est utilisé par C. Jaffrelot en référence à la « substantialisation » de la caste de Louis Dumont, mais s’applique spécifiquement à des mouvements non brahmanes de basses castes.

Statistiquement, on peut se demander dans quelle mesure la nomenclature des castes reflète la complexité des frontières sociales des énonciations de caste saisies par la classification ascendante hiérarchique. Suivant cet algorithme, la partition de densité finale, mesure de l’homogénéité des communautés statistiques, était de 0,009022375. Si nous calculons la densité de partition en utilisant les groupes de nomenclature, il est clair que l’indice densité de partition diminuera car statistiquement la classification ascendante hiérarchique maximisait sa valeur par définition et parce que la nomenclature comprend une catégorie plus disparate que les autres en particulier, celle des castes basses (Figure 4). En effet, lors du calcul des densités de partition par catégorie de la nomenclature, les basses castes ont une

Figure 4 – Diagramme empirique du réseau de co-énonciation des « jatis » et « sous-jatis »



densité plus faible que les autres catégories. La densité de partition de la nomenclature est désormais de 0,003498481, ce qui signifie que notre nomenclature ne reflète que 39% de la densité de partition des communautés statistiques, soit 39% de l'homogénéité mesurée par la partition des communautés statistiques. Si ce niveau d'homogénéité peut paraître faible, nous pensons néanmoins que cette nomenclature a du sens car elle regroupe dans un même schéma différentes jatis considérées objectivement similaires (à partir des dimensions conceptuelles définies dans la section précédente) mais possiblement subjectivement différentes (en termes d'auto-identification).

En ce sens, notre nomenclature est clairement différente d'une classification où nous aurions choisi de couper le dendrogramme de la classification hiérarchique à un niveau plus élevé afin de réduire le nombre de communautés statistiques. Certes, cela conduirait à des communautés statistiquement plus cohérentes que nos catégories de nomenclature, mais aussi moins interprétables, car les communautés rassembleraient des ménages liés faiblement entre eux sur le réseau des castes par des stratégies d'énonciation relevant de processus de « rajputisation », de « kshatriyasation » ou de « sanskritization » (un répondant de basse caste énonçant une identité reflétant une appartenance statutaire de varna plus élevée que celle dans laquelle il est assigné, c'est-à-dire Brahmane, Kshatriya, ou Vaishya, par opposition à Shudra, comme dans l'exemple des Marathas développé à partir d'A. Dongre, section précédente). Ces stratégies de mobilité ascendante des castes développées par les castes inférieures (Lardinois, 2005), ici appliquées à l'auto-identification de la caste, auraient rassemblé les catégories en fonction de la culture légitime à laquelle les communautés de castes s'identifient et rendraient difficile l'étude de ces processus<sup>18</sup>.

### **b. Un proxy de « gold standard » pour la caste**

Dans la littérature mobilisant des algorithmes de classification automatiques ou manuels, en particulier pour identifier l'origine ethnique ou raciale des individus sur la base de leur nom, la pertinence de la méthode de classification est généralement évaluée par rapport à des indicateurs statistiques testant l'efficacité de la classification l'origine ethnique ou raciale « réelle » dans une population de référence. Ce

18. Au contraire, grâce à notre nomenclature, on pourra alors ultérieurement comparer les positions objectives et subjectives des castes en utilisant les catégories de la nomenclature, les communautés statistiques de castes et des indicateurs de styles de vie.

test s'appelle le « *gold standard test* » (Mateos, 2007). Nous n'avons pas ici une population-test sur laquelle effectuer l'évaluation de notre propre classification.

Cependant, l'IHDS-II permet d'évaluer l'efficacité de notre nomenclature pour deux catégories; les Brahmanes et les Dalits. Ces catégories sont en effet saisies dans une variable différente. Dans le questionnaire IHDS-II, l'enquêteur interroge les personnes après qu'elles aient déclarées leurs « jati » et « sous-jati », s'il s'agit de castes « Brahmanes, *General/Forward, Other Backward Classes, Scheduled Castes, Scheduled Tribes* ou autres » ( Question ID13). Ces groupes se réfèrent aux catégories réservées de la « politique des quotas », en ajoutant l'identification aux Brahmanes. Les OBCs constituant un composite de différentes castes, il est difficile de les comparer à notre nomenclature (bien que ce soit très certainement en grande partie une composition de castes agricoles et de basses castes). De même, les « *General/Forward* » (ceux qui ne sont dans aucune catégorie « backward ») correspondent *grosso modo* aux hautes castes (dans notre nomenclature, les castes nobles, les Kayasths et les castes marchandes), mais la correspondance entre catégories n'est pas nette. Nous procédons alors à une comparaison avec la catégorie « Brahmin » et « Scheduled Castes » (la comparaison est faite avec les Dalits de notre nomenclature pour cette dernière caste, les « Scheduled Tribes » y sont ajoutées bien qu'elles ne représentent qu'un faible nombre de ménages). Quatre indicateurs sont calculés pour comparer la nomenclature à ces deux catégories de castes « gold standard » (Tableau 3). La comparaison de nos résultats avec la revue des méthodes de classification de l'ethnicité réalisée par Mateos (2007) conforte notre nomenclature ; les indices ici sont équivalents ou supérieurs à ceux généralement obtenus.

La sensibilité évalue la proportion de ménages se déclarant Brahmanes (respectivement Dalits) dans la variable « ID13 » et correctement considérés comme Brahmanes (Dalits) dans la nomenclature. Cet indicateur est supérieur à 90%. La spécificité évalue la proportion de ménages qui ne se sont pas identifiés comme Brahmanes (Dalits) et qui ne sont pas classés comme tels dans la nomenclature. Elle est supérieure à 95%. La mesure de PPV (*Positive Predictive Value*, indice de valeurs positives prédites) dénote un taux de réussite légèrement inférieur pour les Brahmanes (moins de 90%, mais supérieur à 80%). Il calcule la proportion de ménages déclarés Brahmanes (Dalits) dans la variable « ID13 » parmi les Brahmanes (Dalits) de la nomenclature. Il est intéressant de noter que cet algorithme est plus efficace pour les Brahmanes que

pour les Dalits. Cela signifie que la nomenclature construite à partir des questions ouvertes d'auto-identification (« jati » et « sous-jati ») saisis une assertion de caste de statut élevé (Brahmane) alors qu'en question fermée, les mêmes répondants ne confirment pas ce statut élevé (car ils souhaitent par exemple également déclarer qu'ils appartiennent à une catégorie réservée). La variable d'auto-identification pourrait alors capturer une stratégie de « sanskritization » des basses castes. Enfin, la mesure de NPV (*Negative Predictive Value*, indice de valeurs négatives prédites), qui calcule la proportion de ménages déclarés non Brahmanes (non Dalits) dans la question « ID13 » et identifiés comme tels par la nomenclature, présente des taux de réussite élevés, supérieurs à 95%.

**Tableau 3** – Test « gold standard »

	<b>Brahmanes</b>	<b>Dalit-SC</b>
Taille de la catégorie (%)	11.9	28.3
Taille du « gold standard » (%)	11.8	29.6
Sensibilité	93.4	92.1
Spécificité	97.7	99.0
PPV	82.7	97.5
NPV	97.7	99.0

Ces résultats doivent être interprétés avec prudence. Il n'existe en effet pas d'identité de caste « réelle » ou « fausse » mesurée par l'enquête statistique. Les différences dans la comparaison peuvent bien sûr être dues à des erreurs dans l'algorithme de classification ou à des erreurs dans la collecte des données par les enquêteurs. Mais une partie des variations résulte aussi des significations divergentes entre la caste au sens de jati et de catégorie réservée, comme demandé dans la question « ID13 ». Il n'est pas impossible qu'une personne qui prétende appartenir à une jati classée dans une catégorie « *backward* » ne le sache pas ou ne souhaite pas le déclarer. Par ailleurs, ces catégories sont définies au niveau de l'État régional et, selon la position géographique du ménage dans l'Uttar Pradesh, sa déclaration peut également varier, car les listes de catégories réservées sont différentes dans les États voisins<sup>19</sup>. Les listes de catégories réservées de jatis varient également se-

19. Ainsi, un tiers des ménages de la communauté statistique des « Baniyas » s'identifient eux-mêmes comme OBCs, alors que cette jati ne l'est pas dans l'Uttar Pradesh, ce qui peut s'expliquer par le fait que dans le Bihar, les Baniyas sont répertoriées comme tels et les ménages de l'est de l'Uttar Pradesh peuvent y faire référence.

lon les vicissitudes du jeu politique. En effet, elles sont fortement liées aux mobilisations politiques des groupes de castes et à la politique des « *vote banks* » (Jaffrelot, 2000b), où des groupes de castes soutiennent un candidat en échange, notamment, de la promesse d'inclusion parmi les listes de catégories réservées (voir également Jaffrelot, 2005). Ce point est important car le principe de l'auto-déclaration d'appartenance à une catégorie réservée implique un degré d'incertitude qui devrait encourager les chercheurs des sciences sociales qui mobilisent ces catégories à faire preuve de prudence dans leur utilisation.

Ainsi, la nomenclature apparaît globalement cohérente avec les autres réponses de l'enquête, ce qui est très rassurant, même s'il est impossible de reproduire la même procédure avec toutes les catégories de la nomenclature. La correspondance entre les deux variables testées ici n'est pas parfaite car les questions posées reflètent également les différentes significations symboliques attribuées à la caste par les répondants, soit en tant que jati, soit en tant que catégorie réservée.

### c. Tests préliminaires de la validité du schéma

Enfin, nous nous penchons plus attentivement sur les conventions théoriques de la quantification de la caste et nous proposons de vérifier si les quatre dimensions conceptuelles se reflètent empiriquement dans les sept catégories statistiques de la nomenclature. Ce faisant, les tests effectués visent à vérifier si les catégories statistiques mesurent effectivement ce qu'elles sont censées mesurer (Heath et Martin, 2012). Nous nous inspirons ici des tests de validité menés sur des schémas de classes, en particulier sur le schéma « EGP » (du nom de ses concepteurs Erikson, Goldthorpe, Portocarero, voir Evans, 1992, Evans, 1996, Evans et Mills, 1998).

Les tests de validité par critères (*criterion validity tests*) exigent qu'un concept de la variable testée soit mesuré de différentes manières, de sorte que les variables alternatives soient fortement associées à la caste. Un tel test dépend fortement de la présence de théories bien établies entre la caste et les variables alternatives qui capturent les quatre dimensions conceptuelles. Il est également hautement souhaitable que ces variables alternatives soient théoriquement faiblement corrélées aux autres dimensions de la stratification sociale. Si ce n'est pas le cas, des techniques économétriques plus avancées telles que les régressions seraient nécessaires pour mener une analyse multivariée. Par exemple, si une variable alternative visant à saisir l'une des quatre dimensions théoriques de la caste était également



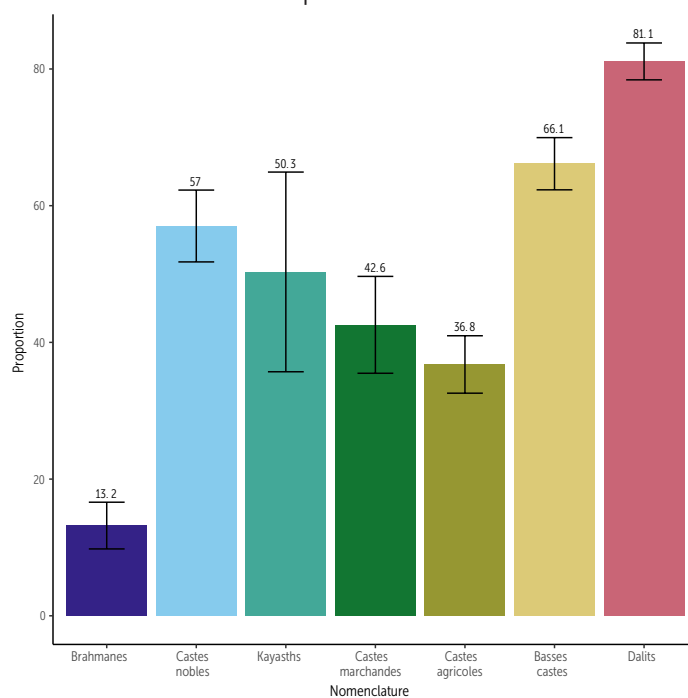
théoriquement liée à la classe sociale, il faudrait inclure un schéma de classe dans l'analyse. En effet, la caste n'épuise certainement pas la multidimensionnalité de la stratification sociale indienne, puisque, par exemple, l'étude de la mobilité intergénérationnelle prouve que les positions des classes sociales sont importantes dans la reproduction sociale (Vaid, 2018). Même si on peut soutenir que certaines des variables testées ici sont corrélées à la classe sociale, malheureusement, le manque de schémas de classes standardisés en Inde<sup>20</sup> est problématique pour mener une analyse multivariée. La disponibilité des variables collectées dans l'enquête limite également notre analyse. Ici, nous ne présentons donc qu'un test préliminaire de la validité de la nomenclature de caste.

Dans ces conditions, nous nous concentrons ici sur quatre indicateurs qui, selon nous, sont fortement corrélés à notre conceptualisation de la caste. L'assertion du rang rituel est étudiée en fonction d'un mode d'alimentation particulier, à savoir le végétarisme. *Homo Hierarchicus* (Dumont, 1967) offre une description claire des divisions alimentaires, les produits alimentaires étant hiérarchisés suivant une logique allant des produits « purs » - les aliments végétariens - aux produits « impurs » (ou les plus pollués) - les aliments non-végétariens. Cette idéologie hindoue forme une « orthopraxie alimentaire » où les pratiques alimentaires se conforment à une codification unie par un dharma, un ordre socio-cosmique, qui régule la vie sociale. La hiérarchie alimentaire fait également référence à la hiérarchie des castes. Ainsi, dans l'hindouisme, la nourriture est « le lien fondamental entre les hommes et les dieux » (Appadurai, 1981). Nous nous attendons donc à ce que la prévalence du végétarisme soit plus élevée parmi les brahmanes et les castes marchandes, puisque les premiers sont au sommet de la hiérarchie de la pureté rituelle et les secondes sont des castes hautement sanskritisées. Cela devrait également être le cas des castes agricoles, qui revendiquent un statut rituel élevé. Au contraire, les castes basses et les Dalits, parce qu'ils sont au bas de la hiérarchie rituelle, sont censés être plus souvent non-végétariens et ne pas respecter les règles rituelles. Mais l'indicateur du végétarisme permet également de tester la deuxième dimension de notre nomenclature, à savoir la concurrence entre différentes valeurs morales entre hautes castes. Étant donné

20. Voir Divya Vaid (2018) pour une exception. L'annexe A de son étude développe un test préliminaire de validité des critères sur son schéma de classes sociales mais les enquêtes mobilisées (les National Election Studies du Center for the Study of Developing Societies) limitent considérablement les possibilités de cet exercice.

consommation de viande est valorisée, nous nous attendons à ce que la prévalence du végétarisme soit plus faible parmi ces castes (Brockert, 2018). Dans l'IHDS-II, chaque répondant a dû répondre à la question suivante : « Y a-t-il quelqu'un dans votre ménage qui mange de la nourriture non végétarienne ? ». Ceux qui ont répondu par l'affirmative à cette question sont représentés graphiquement sur la figure 5 (dans cette section, les intervalles de confiance sont tracés au seuil de 5%).

**Figure 5** – Distribution de la prévalence du non-végétarisme par caste



Les résultats sont conformes à nos affirmations théoriques, de sorte que, à partir de cet indicateur, la nomenclature des castes quantifie de manière adéquate les deux premières dimensions de la caste.

Les trois indicateurs suivants testent la troisième dimension conceptuelle de la caste, à savoir les positions économiques et de domination liées à la caste. Théoriquement, on s'attend à ce que l'incidence de la propriété agraire soit plus élevée parmi les castes nobles et les castes agricoles, puisque ces deux catégories regroupent des castes traditionnellement propriétaires terriennes.

Nous nous attendons à ce que le niveau d'éducation, et en particulier, l'obtention de diplômes de l'enseignement supérieur, soit plus élevé parmi les Brahmanes et les Kayasths, étant donné qu'ils forment traditionnellement des castes lettrées. Enfin, les castes marchandes sont plus susceptibles de posséder une entreprise étant donné qu'elles sont traditionnellement des communautés marchandes. On peut soutenir



que ces positions économiques sont des professions normatives et le résultat de processus historiques et il peut donc sembler audacieux de considérer qu'elles reflètent des positions économiques actuelles. Les réformes agraires suite à l'Indépendance de l'Inde ont ainsi supposé affecté la propriété foncière des castes nobles (Hoerber Rudolph et Rudolph, 2011), ce qui a parallèlement favorisé les castes agricoles (Jeffery, Jeffery et Jeffrey, 2011). De plus, la mise en place de catégories réservées (SC, ST, OBC) vise précisément à réduire les écarts en matière d'éducation en mettant en place des quotas d'accès à l'enseignement supérieur pour les basses castes. On peut donc s'attendre à ce que l'avantage scolaire des Brahmanes et des Kayasths se soit réduit avec le temps. Enfin, la « pluri-activité » économique (Jodhka, 2018) est un processus en forte augmentation en Inde rurale, qui conduit en particulier à la croissance des petites entreprises familiales, et pas seulement parmi les castes marchandes. Mais l'étude de ces processus dynamiques nécessite des enquêtes adéquates sur différentes périodes, ce que nous ne mobilisons pas ici<sup>21</sup>.

Les indicateurs de la propriété agraire, de l'accès à l'enseignement supérieur et de la propriété d'une entreprise sont construits sous forme

21. La nomenclature des castes pourrait néanmoins être implémentée dans l'IHDS-I (2004-2005) et dans le Human Development Profile of India (1993-1994) pour étudier plus précisément ces processus.

de variables dichotomiques et les réponses affirmatives sont représentées dans les figures 6, 7 et 8. La variable d'accès à l'enseignement supérieur détermine si au moins un membre du ménage a étudié au-delà de l'enseignement secondaire. La variable relative à la propriété d'une entreprise mesure si le ménage tire un revenu d'une activité entrepre-

Figure 7 – Accès à l'enseignement supérieur par caste

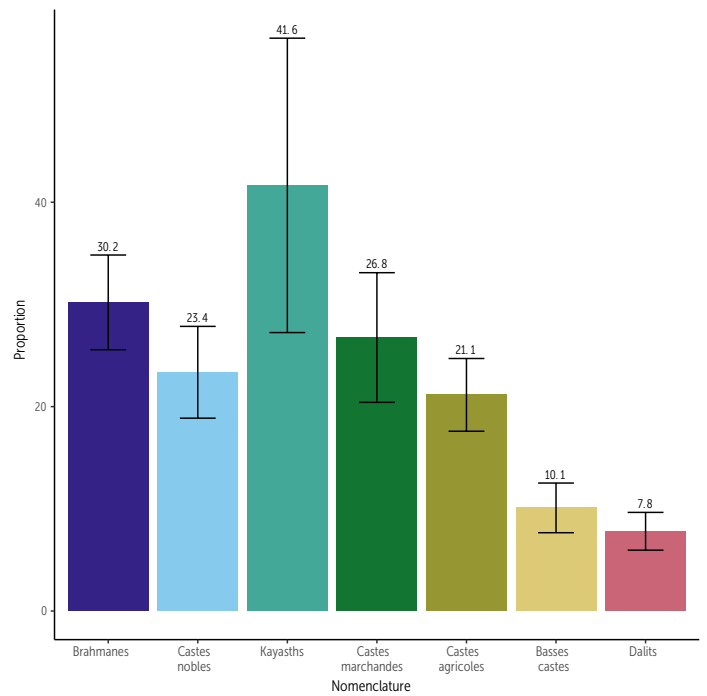


Figure 6 – Taux de propriétaires agraires par caste

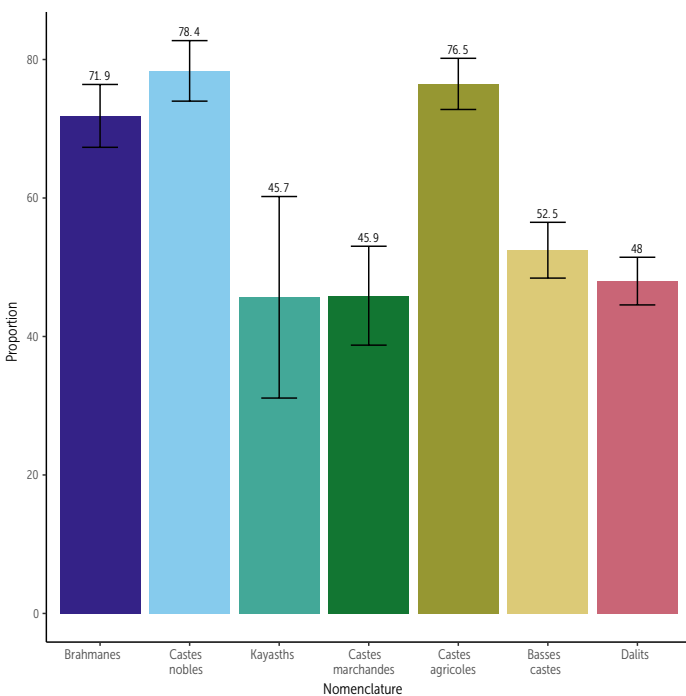
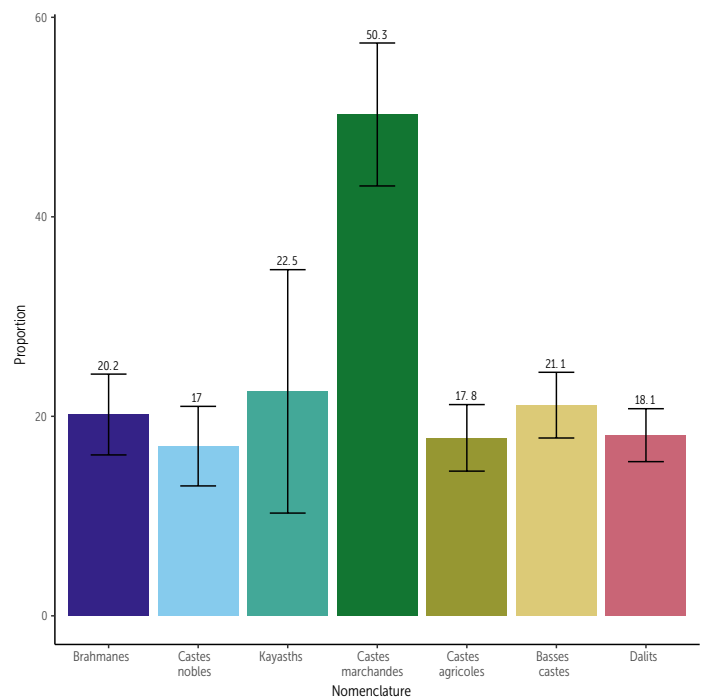


Figure 8 – Propriété d'une entreprise par caste



neuriale d'au moins un membre du ménage.

Les indicateurs sont conformes aux attentes théoriques, c'est-à-dire que les positions économiques actuelles reflètent dans une large mesure les positions économiques traditionnelles, mesurées par la propriété foncière, le niveau d'instruction et la propriété des entreprises. La seule exception provient des Brahmanes. Alors qu'ils sont certainement plus avantagés sur le plan éducatif que les autres castes, les castes marchandes ne sont pas loin derrière et les Brahmanes sont aussi plus souvent propriétaires fonciers qu'attendu (bien qu'ils soient toujours derrière les castes nobles et agricoles). Dans un État où 78 % de la population est comptabilisée comme rurale (contre 69% pour toute l'Inde, recensement de 2011), les ressources agraires des Brahmanes semblent renforcer leur position de haute caste, en les plaçant parmi les castes agraires dominantes de la région. Étant donné l'absence de variables appropriées concernant la dernière dimension conceptuelle - l'identité collective et la conscience de caste - nous ne la testons pas ici.

Ces tests montrent la pertinence de la nomenclature des castes construite à partir de questions ouvertes. Ils démontrent la pertinence d'utiliser une méthode combinant des classifications automatiques et manuelles sur ce type de données. Cette opérationnalisation de la caste ouvre donc la voie à des études statistiques de stratification sociale sur la société indienne.

## Conclusion

En retraçant à la fois l'histoire de la catégorie statistique, les débats sur les possibilités de quantification et les conventions posées par le codage, nous nous sommes efforcés ici de pallier les réticences à l'usage statistique de la caste dans les sciences sociales indiennes. Nous avons voulu montrer dans quelle mesure notre nomenclature de caste ne s'inscrivait pas dans la généalogie des recensements coloniaux, où l'instrument statistique venait renforcer les *a priori* théoriques des recenseurs européens, combinés à une vision socialement située des informateurs locaux. Le codage ne peut cependant prétendre à une neutralité historique et sociale, puisque cette classification a eu des effets performatifs sur les modes d'énonciation et de valorisation des identités de caste.

La représentation des castes dans une nomenclature statistique ne relève pas seulement d'un travail statistique et nous avons voulu ici aussi expliciter et prendre en compte les processus politiques qui conduisent à la formation d'une identité collective dans le monde social, ainsi que les processus cognitifs d'auto-identification dans une catégorie (Desrosières et Thévenot, 1988). Le développement de ces différents éléments tout au long du travail méthodologique ne permet pas seulement de justifier les « équivalences » proposées dans la nomenclature. Il montre aussi l'importance de la catégorie de la caste pour les interrogés, pour qui l'énonciation d'une identité de caste varie en fonction des stratégies de mobilité sociale dont ils font preuve et du sentiment d'appartenance à un groupe social plus ou moins institutionnalisé (par des associations de caste par exemple). Ainsi, notant que l'identité de caste reste souvent un point aveugle des études quantitatives de la société indienne, nous avons ici mis en place une méthodologie permettant de créer une nomenclature de caste, mais nous aussi avons cherché à montrer l'importance de l'étude des modes d'énonciation de l'appartenance de caste.

Bien qu'aucune classification réalisée à partir d'une enquête menée uniquement auprès de quelques milliers de ménages ne puisse prétendre à la perfection, la combinaison de méthodes automatiques et manuelles permet d'obtenir des regroupements précis. La possibilité de mobiliser des questions ouvertes d'identité de caste a curieusement

été jusqu'ici largement ignorée dans les sciences sociales de l'Inde. La non-utilisation de cette ressource est d'autant plus surprenante que les débats sur la saillance de la caste dans la société indienne se réduisent à des « indices », conduisant parfois à conclure à une diminution de la saillance de la caste dans la société indienne contemporaine (Béteille, 2012).

Ces questions ouvertes sont, à notre avis, une ressource unique pour étudier l'importance de la caste dans la structuration des modes de vie. Si la caste concerne de plus en plus une forme de séparation culturelle, ou d'« ethnicisation », plutôt qu'un statut hiérarchique (Fuller, 1996), il est urgent de saisir ces réalités changeantes à un niveau synthétique plus large que celui auquel permet d'accéder l'enquête ethnographique. De plus, si le statut de caste est aujourd'hui moins important que la classe sociale, comme le suggèrent certaines études qualitatives (Dolphijn, 2006), la recherche quantitative est indubitablement nécessaire pour évaluer l'importance de dimensions de stratification sociale concurrentes à la caste. De toute évidence, la quantification ne signifie pas abandonner les études ethnographiques, qui s'avèrent essentielles dans la quantification, comme le montre ce travail, mais la recherche quantitative pourrait apporter des indications théoriques supplémentaires sur la réalité sociale de la caste dans l'Inde contemporaine. Comme nous l'avons déjà noté, les pratiques alimentaires forment des styles de vie structurés par la caste et des travaux de recherche supplémentaires de notre côté seront axés sur des indicateurs de ce domaine culturel. La nomenclature développée ici est donc une étape nécessaire pour étudier la stratification sociale des pratiques alimentaires en Inde.

Les résultats ultérieurs issus d'une mobilisation de cette nomenclature devront être interprétés à la lumière de la définition théorique de la caste que nous avons fournie ici. Comme la caste est une catégorie théoriquement contestée, nous nous attendons à ce que les caractérisations conceptuelles mobilisées dans cet article alimentent des commentaires critiques. En particulier, nous ne comptons pas uniquement sur le statut rituel pour construire une nomenclature hiérarchique unidimensionnelle de la caste, telle que théoriquement envisagée par Dumont (1967). Au contraire, la nomenclature tente de prendre en compte les réalités historiques changeantes de la caste. Mais en expliquant les conventions de codage, la nomenclature quantifie un concept explicite, plutôt que de nommer caste un proxy statistique tel

que les catégories réservées de la politique des quotas. Un long chemin reste cependant à parcourir pour que les sciences sociales mobilisent des schémas de caste plus précis et plus standardisés et la présente étude ne constitue au mieux qu'une étape dans cette direction.

## References

- AHMAD I., 1973, Caste and social stratification among the Muslims, Manohar Book Service; [distributed in U.S.A.: South Asia Books, Columbia, Mo.], 302 p.
- AHNY.-Y., BAGROW J.P., LEHMANN S., 2010, "Link communities reveal multiscale complexity in networks", *Nature*, 466, 7307, p. 761-764. [<https://doi.org/10.1038/nature09182>]
- APPADURAI A., 1981, "Gastro-Politics in Hindu South Asia", *American Ethnologist*, 8, 3, p. 494-511. [<https://doi.org/10.1525/ae.1981.8.3.02a00050>]
- APPADURAI A., 1993, "Number in the Colonial imagination", dans BRECKENRIDGE C.A., VEER P. VAN DER (dirs.), *Orientalism and the postcolonial predicament: perspectives on South Asia*, Philadelphia, University of Pennsylvania Press.
- BAYLY S., 2001, *Caste, Society and Politics in India from the Eighteenth Century to the Modern Age*, Cambridge University Press, 448 p.
- BENBABAALI D., 2008, "Questioning the Role of the Indian Administrative Service in National Integration", *South Asia Multidisciplinary Academic Journal*. [<http://dx.doi.org/10.4000/samaj.633>]
- BÉTEILLE A., 2012, "The peculiar tenacity of caste", *Economic and Political Weekly*, 47, 13, p. 41-42. [[Lien Alt](#)]
- BEZDEK J.C., HATHAWAY R.J., 2002, "VAT: a tool for visual assessment of (cluster) tendency", *Conference Neural Networks, Proceedings of the 2002 International Joint Conference*, p. 2225-2230, vol. 3. [<https://doi.org/10.1109/IJCNN.2002.1007487>]
- BOROOAH V.K., 2017, "Caste and Regional Influences on the Practice of 'Untouchability' in India: The Practice of 'Untouchability' in India", *Development and Change*, 48, 4, p. 746-774. [<https://doi.org/10.1111/dech.12311>]
- BRUCKERT M., 2018, *La chair, les hommes et les dieux: la viande en Inde*, CNRS Editions.
- COPEMAN J., 2015, "Secularism's Names: Commitment to Confusion and the Pedagogy of the Name", *South Asia Multidisciplinary Academic Journal*, 12. [<http://dx.doi.org/10.1177/1525822X14564275>]
- DESAI S., 2010, "Caste and Census: A Forward Looking Strategy", *Economic and Political Weekly*, 45, 29, p. 10-13. [[Lien Alt](#)]
- DESAI S., DUBEY A., 2012, "Caste in 21st Century India: Competing Narratives", *Economic and political weekly*, 46, p. 40-49. [[Preprint](#)]
- DESAI S., VANNEMAN R., 2018, *India Human Development Survey-II (IHDS-II)*, 2011-12, Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2018-08-08. [<https://doi.org/10.3886/ICPSR36151.v6>]
- DESHPANDE A., RAMACHANDRAN R., 2017, "Dominant or Backward? Political Economy of Demand for Quotas by Jats, Patels, and Marathas", *Economic and Political Weekly*, 52, 19, p. 81-92. [[Working Paper](#)]
- DESHPANDE S., 2005, « Castes et inégalités sociales dans l'Inde contemporaine », *Actes de la recherche en sciences sociales*, n° 160, 5, p. 98-116. [<https://doi.org/10.3917/arss.160.0098>]
- DESHPANDE S., JOHN M.E., 2010, « Le déni de la caste en Inde », *La Vie des Idées*. [[Lien](#)]
- DESROSIERES A., 1993, *La politique des grand nombres: histoire de la raison statistique*, Paris, La Découverte.
- DESROSIERES A., 2001, « Entre réalisme métrologique et conventions d'équivalence: les ambiguïtés de la sociologie quantitative », *Genèses*, n° 43, 2, p. 112-127. [<https://doi.org/10.3917/gen.043.0112>]
- DESROSIERES A., DIDIER E., 2014, *Prouver et gouverner: une analyse politique des statistiques publiques*, Paris, La Découverte, 284 p.
- DESROSIERES A., THEVENOT L., 1988, *Les catégories socioprofessionnelles*, 5<sup>e</sup> éd, Paris, La Découverte (Collection Repères), 122 p.
- DIRKS N.B., 2001, *Castes of mind: colonialism and the making of modern India*, Princeton, N.J, Princeton University Press, 372 p.
- DOLPHIJN R., 2006, "Capitalism on a Plate: The Politics of Meat Eating in Bangalore, India", *Gastronomica*, 6, p. 52-59. [[Lien Alt](#)]
- DONGRE A.A., 2017, "Is Maratha Demand for Reservation Tenable?: Evidence from India Human Development Survey", *SSRN Scholarly Paper*, ID 2968146, Rochester, NY, Social Science Research Network.

- DUMONT L., 1967, *Homo hierarchicus: Essai sur le système des castes*, Editions Gallimard, 714 p.
- EVANS G., 1992, "Testing the Validity of the Goldthorpe Class Schema", *European Sociological Review*, 8, 3, p. 211-232. [<https://doi.org/10.3917/gen.043.0112>]
- EVANS G., 1996, "Putting Men and Women into Classes: An Assessment of the Cross-Sex Validity of the Gold Thorpe Class Schema", *Sociology*, 30, 2, p. 209-234. [<https://doi.org/10.1177/2F0038038596030002002>]
- EVANS G., MILLS C., 1998, "A Latent Class Analysis of the Criterion-Related and Construct Validity of the Goldthorpe Class Schema", *European Sociological Review*, 14, 1, p. 87-106. [<https://doi.org/10.1093/oxfordjournals.esr.a018229>]
- FELOUZIS G., 2008, « L'usage des catégories ethniques en sociologie », *Revue française de sociologie*, 49, 1, p. 127-132. [<https://doi.org/10.3917/rfs.491.0127>]
- FERRY M., NAUDET J., ROUEFF O., 2018, "Seeking the Indian social space. A multidimensional portrait of the stratifications of Indian society", *South Asia Multidisciplinary Academic Journal*. [<http://dx.doi.org/10.4000/samaj.4462>]
- FULLER, C.J. (dir.), 1996, *Caste today*, Delhi ; New York, Oxford University Press (SOAS studies on South Asia : understandings and perspectives), 295 p.
- FULLER C.J., 2016, "Anthropologists and Viceroy: Colonial knowledge and policy making in India", *Modern Asian Studies*, 50, 1, p. 217-258. [<https://doi.org/10.1017/S0026749X15000037>]
- FULLER C.J., 2017, "Ethnographic inquiry in colonial India: Herbert Risley, William Crooke, and the study of tribes and castes: Ethnographic inquiry in colonial India", *Journal of the Royal Anthropological Institute*, 23, 3, p. 603-621. [<https://doi.org/10.1111/1467-9655.12654>]
- GUILMOTO C.Z., 1998, « Le texte statistique colonial [À propos des classifications sociales dans l'Inde britannique] », *Histoire & Mesure*, 13, 1, p. 39-57. [[Lien](#)]
- HEADLEY Z., 2013, « Nommer la caste. Ordre social et catégorie identitaire en Inde contemporaine », *La Vie des Idées*. [[Lien](#)]
- HEATH A., MARTIN J., 2012, "Why Are There so Few Formal Measuring Instruments in Social and Political Research?", dans LY-  
BERG L., BIEMER P., COLLINS M., DE LEEUW E., DIPPO C., SCHWARZ N., TREWIN D. (dirs.), *Survey Measurement and Process Quality*, Hoboken, NJ, USA, John Wiley & Sons, Inc., p. 71-86.
- HENRY O., FERRY M., 2017, "When Cracking the JEE is not Enough", *South Asia Multidisciplinary Academic Journal*, 15, traduit par GEORGE R. [<http://dx.doi.org/10.4000/samaj.4291>]
- HIMANSHU, JHA, P., RODGERS, G. (dirs.), 2016, *The Changing Village in India: Insights from Longitudinal Research*, Oxford, New York, Oxford University Press, 592 p.
- HOEBER RUDOLPH S., RUDOLPH L.I., 2011, "From Landed Class to Middle Class: Rajput Adaptation in Rajasthan", dans BAVISKAR A., RAY R. (dirs.), *Elite and Everyman: The Cultural Politics of the Indian middle Class*, New Delhi (Inde), Routledge, Taylor & Francis Group.
- IVERSEN V., KALWIJ A., VERSCHOOR A., DUBEY A., 2010, "Caste dominance and economic performance in rural India", *Indian Statistical Institute, Planning Unit, New Delhi Discussion Papers*, 10-01, Indian Statistical Institute, New Delhi, India.
- JAFFRELOT C., 2000a, "Sanskritization vs. Ethnicization in India: Changing Identities and Caste Politics before Mandal", *Asian Survey*, 40, 5, p. 756-766. [[JSTOR](#)] [[Lien Alt](#)]
- JAFFRELOT C., 2000b, "The Rise of the Other Backward Classes in the Hindi Belt", *The Journal of Asian Studies*, 59, 1, p. 86-108. [[JSTOR](#)] [[Lien Alt](#)]
- JAFFRELOT C., 2005, *Inde, la démocratie par la caste: histoire d'une mutation socio-politique (1885-2005)*, Paris, Fayard (L'espace du politique), 593 p.
- JAFFRELOT C., 2010, *Religion, Caste, and Politics in India*, Primus Books, 835 p.
- JAFFRELOT C., KUMAR S., 2012, *Rise of the Plebeians?: The Changing Face of the Indian Legislative Assemblies*, Routledge, 531 p.
- JAFFRELOT C., NAUDET J., 2013, *Justifier l'ordre social caste, race, classe et genre*, Paris, Presses universitaires de France.
- JEFFERY R., JEFFERY P., JEFFREY C., 2011, "Are Rich Rural Jats Middle Class?", dans BAVISKAR A., RAY R. (dirs.), *Elite and Everyman: The Cultural Politics of the Indian middle Class*, New Delhi (Inde), Routledge, Taylor & Francis Group.
- JODHKA S., 2018, "Rural Change in Times of "Distress" ", *Economic and Political Weekly*, 53, 26-27, p. 5-7. [[Lien](#)]

- KALAIYARASAN A., 2016, "Mapping the Discourse from Domination to Deprivation: A Case Study of Jats", Draft, IGIDR, New Delhi (Inde), IGIDR. [\[Draft\]](#)
- KALINKA A.T., TOMANCAK P., 2011, "linkcomm: an R package for the generation, visualization, and analysis of link communities in networks of arbitrary size and type", *Bioinformatics*, 27, 14, p. 2011-2012. [\[https://dx.doi.org/10.1093%2Fbioinformatics%2Fbtr311\]](https://dx.doi.org/10.1093%2Fbioinformatics%2Fbtr311)
- LARDINOIS R., 1985, « Les Luites de Classement en Inde », *Actes de la recherche en sciences sociales*, 59, p. 78-84. [\[Persée\]](#)
- LARDINOIS R., 1996, « Rumeurs, résistances, rébellions : la mise en place des recensements dans l'Inde coloniale (XVIIIe-XXe siècles) », *Cahiers québécois de démographie*, 25, 1, p. 39. [\[https://doi.org/10.7202/010200ar\]](https://doi.org/10.7202/010200ar)
- LARDINOIS R., 2005, « Les usages politiques de la caste », *Actes de la recherche en sciences sociales*, n° 160, 5, p. 117-121. [\[https://doi.org/10.3917/ars.160.0117\]](https://doi.org/10.3917/ars.160.0117)
- LARDINOIS R., 2007, *L'invention de l'Inde. Entre ésotérisme et science*, Paris, CNRS Editions.
- LAZEGA E., 2014, *Réseaux sociaux et structures relationnelles*, Paris, Presses universitaires de France.
- MATEOS P., 2007, "A review of name-based ethnicity classification methods and their potential in population studies", *Population, Space and Place*, 13, 4, p. 243-263. [\[Lien\]](#)
- MATEOS P., LONGLEY P.A., O'SULLIVAN D., 2011, "Ethnicity and Population Structure in Personal Naming Networks", *PLOS ONE*, 6, 9, p. e22943. [\[https://doi.org/10.1371/journal.pone.0022943\]](https://doi.org/10.1371/journal.pone.0022943)
- MICHELUTTI L., 2008, " 'We are Kshatriyas but we behave like Vaishyas': Diet and Muscular Politics Among a Community of Yadavs in North India", *South Asia: Journal of South Asian Studies*, 31, 1, p. 76-95. [\[https://doi.org/10.1080/00856400701874726\]](https://doi.org/10.1080/00856400701874726)
- NAUDET J., ALLORANT A., FERRY M., 2018, "Heirs, corporate aristocrats and 'Meritocrats': the social space of top CEOs and Chairmen in India", *Socio-Economic Review*, 16, 2, p. 307-339. [\[https://doi.org/10.1093/ser/mwx035\]](https://doi.org/10.1093/ser/mwx035)
- ROBERTS N., 2016, *To Be Cared For: The Power of Conversion and Foreignness of Belonging in an Indian Slum*, New Delhi, Navayana, 286 p.
- SRINIVAS M.N., 1959, "The Dominant Caste in Rampura", *American Anthropologist*, 61, 1, p. 1-16. [\[Alt Link\]](#)
- SRINIVAS M.N., 1952, *Religion and Society among the Coorgs of South India*, Clarendon Press, Oxford.
- SRINIVASAN K., KUMAR S., 1999, "Economic and Caste Criteria in Definition of Backwardness", *Economic and Political Weekly*, 34, 42/43, p. 3052-3057. [\[JSTOR\]](#)
- SUNDAR N., 2000, "Caste as Census Category: Implications for Sociology", *Current Sociology*, 48, 3, p. 111-126. [\[https://doi.org/10.1177%2F0011392100048003008\]](https://doi.org/10.1177%2F0011392100048003008) [\[Lien Alt\]](#)
- SUSEWIND R., 2015, "What's in a Name? Probabilistic Inference of Religious Community from South Asian Names", *Field Methods*, 27, 4, p. 319-332. [\[https://doi.org/10.1177%2F1525822X14564275\]](https://doi.org/10.1177%2F1525822X14564275) [\[Preprint\]](#)
- THORAT, S., NEUMAN, K.S. (dirs.), 2012, *Blocked by Caste: Economic Discrimination in Modern India*, Oxford, New York, Oxford University Press, 400 p.
- VAID D., 2014, "Caste in Contemporary India: Flexibility and Persistence", *Annual Review of Sociology*, 40, 1, p. 391-410. [\[https://doi.org/10.1146/annurev-soc-071913-043303\]](https://doi.org/10.1146/annurev-soc-071913-043303)
- VAID D., 2018, *Uneven odds: social mobility in contemporary India*, First edition, New Delhi, Oxford University Press, 338 p.
- VERBORGH R., WILDE M.D., 2013, *Using OpenRefine*, Packt Publishing Ltd, 186 p.
- VISSA B., 2011, "A matching theory of entrepreneurs' tie formation intentions and initiation of economic exchange", *The Academy of Management Journal*, 54, 1, p. 137-158. [\[https://doi.org/10.5465/amj.2011.59215084\]](https://doi.org/10.5465/amj.2011.59215084) [\[Preprint\]](#)
- WAGONER P.B., 2003, "Precolonial Intellectuals and the Production of Colonial Knowledge", *Comparative Studies in Society and History*, 45, 4, p. 783-814. [\[JSTOR\]](#) [\[Lien Alt\]](#)

Seuls les liens avec doi sont officiels et pérennes.



## Annexes

Communautés statistiques de la classification ascendante hiérarchique regroupés par catégories de la nomenclature

Nomenclature	Communautés statistiques	Distribution (%)	Trois jatis les plus fréquemment énoncées		
Brahmanes	BRAHMIN	10.82	BRAHMIN	TIWARI	MISHRA
	JOGI	0.26	JOGI	UPADHAY	BRAHMIN
	PANDIT	0.20	PANDIT	BHAMUN	DUBE
	TIWARI	0.20	TIWARI	PEDI	SIRAJAN
	GOSWAMI	0.19	GOSWAMI	GIRI	GOSAI
	BHARDWAJ	0.16	BHARDWAJ	RAJBHATT	RABHAR
	MISHRA	0.12	MISHRA	BARMAN	BHAJIN
	<b>Total</b>	<b>11.95</b>			
Castes nobles	RAJPUT	3.74	RAJPUT	LODHI	SINGH
	KUSHWAHA	2.25	KUSHWAHA	MAURYA	KOIRI
	THAKUR	1.90	THAKUR	SINGH	ARKABANSHI
	CHOUHAN	1.76	CHOUHAN	LONIA	SINGH
	KSHATRIYA	0.89	KSHATRIYA	SINGH	CHATRIY
	KATHARIYA	0.25	KATHARIYA	DHANU	DHANUSAK
	CHHETRI	0.24	CHHETRI	SINGH	KAPOOR
	RAY	0.18	RAY	BHUMIHAR	KALGAR
	RATHORE	0.15	RATHORE	VAHELIYA	KSHATRIYA
	RAWAT	0.09	RAWAT	BHESTAR	MEKHTAR
	KHETRIYA	0.08	KHETRIYA	NIKUMBHA	SINGH
<b>Total</b>	<b>11.53</b>				
Kayasths	SRIVASTAV	0.89	SRIVASTAV	KAYASTH	LAL
	SAXENA	0.30	SAXENA	BHURJI	MURJI
	SHREEVATAB	0.05	SHREEVATAB	KAIYARATH	KAMART
	<b>Total</b>	<b>1.23</b>			
Castes marchandes	GUPTA	3.14	GUPTA	TELI	RATHORE
	BANIA	1.43	BANIA	AGARWAL	VAISAYA
	JAIWAL	0.73	JAIWAL	KALWAR	GUPTA
	KHATRI	0.16	KHATRI	MEHROTRA	KHETI
	SETHI	0.10	SETHI	GADHI	SHREEGOAL
	PANJABI	0.05	PANJABI	ARORA	BULA
	<b>Total</b>	<b>5.6</b>			
Castes agricoles	YADAV	9.13	YADAV	AHIR	GAWAL
	JAT	3.85	JAT	CHOWDHURY	SINGH
	KURMI	3.58	KURMI	PATEL	BERMA
	SONAR	1.10	SONAR	VERMA	SONI
	AHIR	0.55	AHIR	GAWAL	GWALBAL
	GUJJAR	0.28	GUJJAR	SINGH	KABHANA
	LODHI	0.26	LODHI	BARMA	JARIYA
	PATEL	0.24	PATEL	KUMAR	KURKI
	<b>Total</b>	<b>19</b>			

Nomenclature	Communautés statistiques	Distribution (%)	Trois jatis les plus fréquemment énoncées		
Basses castes	RAJVERE	5.02	SINGH	SHARMA	RAJVERE
	KEWAT	2.53	KEWAT	BIND	NISHAD
	RAJBHAR	2.35	RAJBHAR	BHAR	BHARDWAJ
	PAL	2.07	PAL	GADARIA	BAGHEL
	PRAJPATI	1.86	PRAJPATI	KUMHAR	KUMAR
	KAHAR	1.43	KAHAR	KASHYAP	DHEEMAR
	VISHWAKARMA	1.43	LOHAR	BADHAI	LADHAR
	NAI	1.03	NAI	SHARMA	SINGH
	NISHAD	0.85	NISHAD	BIND	MALLAH
	MALI	0.33	MALI	SAINI	MOYA
	BADHAI	0.28	BADHAI	SHARMA	JHA
	KHARWAR	0.24	KHARWAR	KAMKAR	KHAR
	KUMHAR	0.20	KUMHAR	CHEKBERAY	KASHYAP
	KANOJIA	0.11	KANOJIA	GHABI	MAURYA
	KASHYAP	0.11	KASHYAP	HARIJAN	KORI
	BHAR	0.09	BHAR	BHARATDWAJ	GOUD
	LOHAR	0.06	LOHAR	KISAN	LOHA
	<b>Total</b>	<b>20</b>			
Dalits	CHAMAR	13.03	CHAMAR	HARIJAN	JATAV
	PASI	4.21	PASI	BHARGAB	SAROJ
	DHOBI	3.21	DHOBI	KANOJIA	DIWAKAR
	HARIJAN	1.85	HARIJAN	JATAV	RAVIDAS
	KORI	1.43	KORI	JULAHA	KAMAL
	MUSHAR	1.18	MUSHAR	ADIVASI	KOL
	JATAV	1.01	JATAV	SINGH	JULAHA
	KHATIK	0.92	KHATIK	SONKAR	CHEEK
	VALMIKI	0.26	VALMIKI	JAMADAR	CHOWDHURY
	GOAD	0.25	GOAD	GOD	SHAH
	JATHA	0.23	JATHA	AHIRWAR	SOHARBAR
	DUSADH	0.19	DUSADH	PASWAN	GAHLOT
	KOLI	0.18	KOLI	MAHAR	KOL
	KOURI	0.13	KOURI	KAIVBAR	KULDEEP
	ZADEV	0.13	ZADEV	KAYAM	KHAM
	BHARGAB	0.04	BHARGAB	RAJMASI	YASI
	<b>Total</b>	<b>28.26</b>			

Nomenclature	Communautés statistiques	Distribution (%)	Trois jatis les plus fréquemment énoncées		
Non identifiés	SABAT	0.61	SABAT	KISHERI	TEZI
	BANSHI	0.56	ARAKH	BANSHI	KHAGAR
	UNIDENTIFIED	0.39			
	GOUD	0.30	GOUD	BHUJ	DHURIYA
	MAHAL	0.20	MAHAL	NISDTH	MHEND
	BHAMUN	0.14	BHAMUN	SHARMA	CHATRUBEDI
	BANIMA	0.08	BANIMA	BISHO	MADHISIMA
	SAHANI	0.08	SAHANI	DEVAR	KEBAR
	VARMA	0.07	VARMA	KURSHI	SAVRANKAR
	Total	2.43			





Contact auteur

[mathieukferry@gmail.com](mailto:mathieukferry@gmail.com)

## A propos de l'auteur

Mathieu Ferry est doctorant contractuel avec monitorat (ENS Paris Saclay) à l'Observatoire sociologique du changement et au Laboratoire de Sociologie Quantitative (GENES-CREST). Il mène depuis l'automne 2017 un travail de thèse sur la stratification sociale des pratiques alimentaires en Inde, sous la direction de Philippe Coulangeon (OSC) et de Christophe Jaffrelot (CERI). Il co-anime avec Joël Cabalion, Odile Henry, Jules Naudet et Olivier Roueff un séminaire sur la sociologie des inégalités en Inde à l'EHESS.

## Résumé

Cet article expose les difficultés rencontrées par le chercheur en sciences sociales lorsqu'il souhaite analyser quantitativement les identités sociales mesurées par des questions ouvertes dans des grandes enquêtes statistiques. La grande diversité apparente des réponses énoncées démontre la complexité de l'auto-identification mais ne retire en rien la pertinence de quantifier une catégorie sociale latente. Nous discutons de notre approche de la construction d'une nomenclature de caste à partir de questions ouvertes dans l'Indian Human Development Survey (2011-2012), en nous centrant sur les ménages hindous de l'Uttar Pradesh. Nous commençons par exposer les enjeux d'une telle quantification, en soulignant l'histoire coloniale à laquelle elle est fortement associée. Contrairement aux idées reçues, la caste est loin d'être une catégorie institutionnalisée incontestée et sa mesure statistique est fortement critiquée. Néanmoins, plusieurs arguments plaident en faveur de sa quantification. Nous décrivons notre algorithme de classification basé sur l'analyse de réseau, la classification ascendante hiérarchique et le recodage manuel. Nous suggérons ensuite d'évaluer la pertinence de notre classification sous trois angles différents dans ce travail préliminaire. Premièrement, les indicateurs d'homogénéité font apparaître des catégories de la nomenclature relativement homogènes. Deuxièmement, la comparaison de la nomenclature avec une variable « *gold standard* » permet d'évaluer son efficacité. Enfin, des tests de validité vérifient si les catégories de castes reflètent les dimensions conceptuelles du statut rituel et socioéconomique de la caste. Ce faisant, nous montrons que notre nomenclature en sept groupes de castes permet de rompre avec une vision hiérarchique unidimensionnelle à laquelle la structure sociale de caste est souvent associée.

## Mots clés

questions ouvertes, quantification, caste, structure sociale, algorithme de classification, analyse de réseau

## Ce texte est également disponible en version anglaise

### Version électronique, en ligne / On-line version

<http://www.sciencespo.fr/osc/fr/content/notes-documents-de-l-osc>

Également disponible sur SPIRE, l'archive ouverte de Sciences Po. <https://spire.sciencespo.fr/web/>

### Pour citer cet article

Mathieu Ferry, « Les liens de la caste: quantifier les identités sociales à l'aide de questions ouvertes », Sciences Po OSC Papers, n° 2019-1, mai 2019.

### Edition de ce numéro / Editorial board

Bernard Corminboeuf (valorisation de la recherche, Sciences Po-CNRS).  
[bernard.corminboeuf@sciencespo.fr](mailto:bernard.corminboeuf@sciencespo.fr)

Image de couverture : Chris JL, "Another morning for the lucky ones" (Mathura, Uttar Pradesh) - via Flickr (CC BY-NC-ND 2.0)

Observatoire Sociologique  
du Changement

27 rue Saint-Guillaume  
75337 Paris cedex 07  
01 45 49 54 50

<http://www.sciencespo.fr/osc/fr/>

Responsable de la publication :  
Mirna Safi

Responsable éditoriale :  
Agnès van Zanten



OSC 2019