

Saisir l'appartenance de caste à l'aide de questions ouvertes  
*retour critique sur une opération de quantification*

Mathieu Ferry (UVSQ/Printemps)

PUD Lille

Le 12 juin 2025

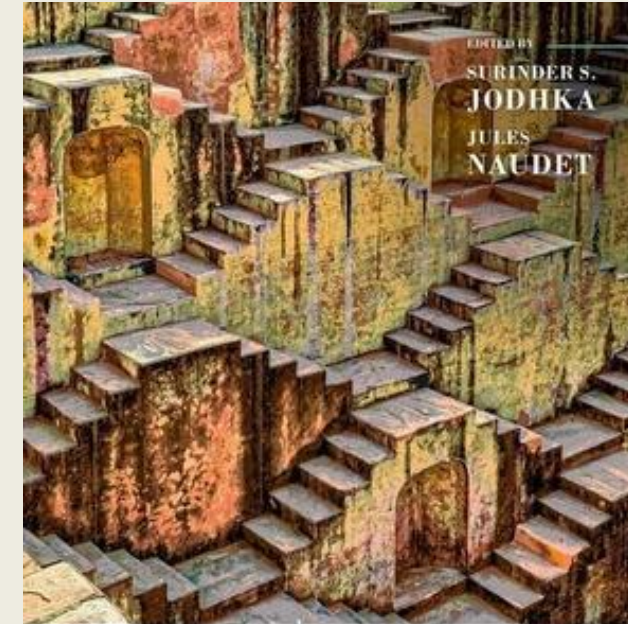
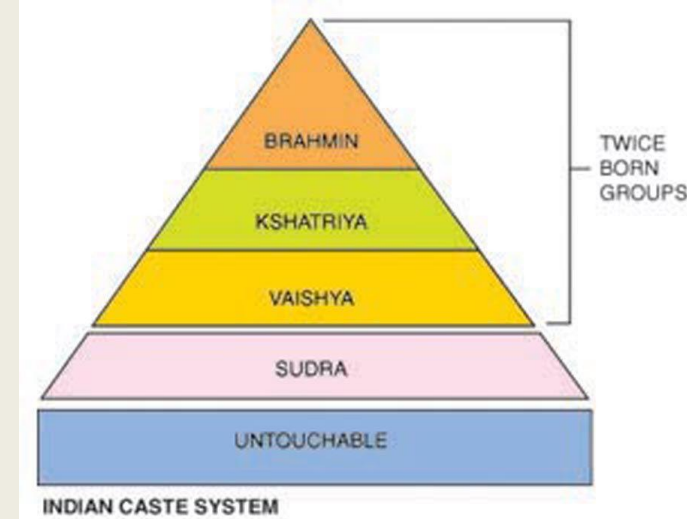
# Introduction

- Quantification de la caste, à partir de réponses à des questions ouvertes
  - Dans une enquête représentative des ménages indiens (*Indian Human Development Survey*), N = 42,000 ménages sur toute l'Inde, et N = 3,000 ménages hindous en Uttar Pradesh (région étudiée du nord de l'Inde)
  - Et dans le *National Family Health Survey* (N=600,000 ménages, dont 60,000 ménages hindous en Uttar Pradesh)
- Dans le cadre de ma recherche doctorale sur les styles de vie alimentaires et le rapport à la viande et au végétarisme
  - En fonction de la religion, du genre, de la classe... et de la caste !
  - Une enquête « mixte » : analyse secondaire d'enquêtes statistiques et entretiens

La catégorie sociale de la caste

# Mais qu'est-ce que la caste ?

- Trois niveaux de définition :
  - La **varna**, une représentation idéale (brahmanique) du monde social (les « deux fois nés » versus les basses castes + intouchables)
  - La **jati**, réalité sociale de la caste : plus de 4,5M d'identifications de jatis différentes dans le *Socio-Economic Caste Census* de 2011
    - Groupe héréditaire (naissance)
    - Groupe endogame (mariage)
    - Groupe occupationnel traditionnel (profession)
    - « inégalité graduelle »
  - Des **catégories administratives**
    - Au fondement de « politiques de réservation » dans le secteur public (enseignement supérieur et emploi)
    - = forme de discrimination positive



≡ The Oxford Handbook of  
**CASTE**

# La caste, des luttes de classement

- Fortes variations régionales
- Des catégories fluctuantes suivant les interlocuteur·ices : un « kaléidoscope des identités » (Headley, 2013)
- Des stratégies de (re-)catégorisation variées
- <https://www.youtube.com/watch?v=r4N98tQsp64>

*Ayaan: What about the temple? Aren't Dalits allowed to enter?*

*Sub officer: Right, sir.*

*Ayaan: Who beat them up?*

*Sub officer: Boys from Mahantji's ashram.*

*Ayaan: Who is Mahantji?*

*Ayaan's assistant: Sir, Mahantji is Mahantji.*

*Ayaan: He's upper caste?*

*Assistant: A Brahman, sir.*

*Ayaan: And those young men?*

*Assistant: They are Scheduled castes.*

*Ayaan: What caste is Brahmandatt?*

*Assistant: Thakur, sir.*

*Ayaan: So are Thakurs considered to be Brahmins, Kshatriyas or Vaishyas?*

*Constable: They are Kshatriyas, sir.*

*Ayaan: What are you?*

*Constable: I am a Jat. Now called OBC, Other Backward Class.*

*Sub officer: So why are you still protesting?*

*Constable: I'm not, sir. The Haryana Jats are protesting. They're still considered Jats, sir.*

*Ayaan: Jatav, aren't you a Dalit?*

*Sub officer: Yes, sir.*

*Ayaan: Same caste as the injured boys?*

*Sub officer: No, sir. I'm a Chamar. They were Pasis. We're much higher than them. We have nothing to do with them, we don't eat the food they touch.*

*Ayaan: And I can't eat the food you touch?*

*Sub officer: Right, sir.*

*Ayaan: Right... And what are you? [To his assistant]*

*Assistant: A Kayasth, sir.*

*Ayaan: What's a Kayasth?*

*Assistant: We're different. We don't fall into the four main castes.*

*Ayaan: And me, what am I?*

*Assistant: A Brahmin, sir.*

*Ayaan: So am I Mahantji's equal?*

*Assistant: No no, sir. He is a Kanyakuja Brahmin, a top Brahmin. You're a Saryuparin.*

*Ayaan: What the fuck is going on here?*

# La mesure de la caste, une histoire coloniale

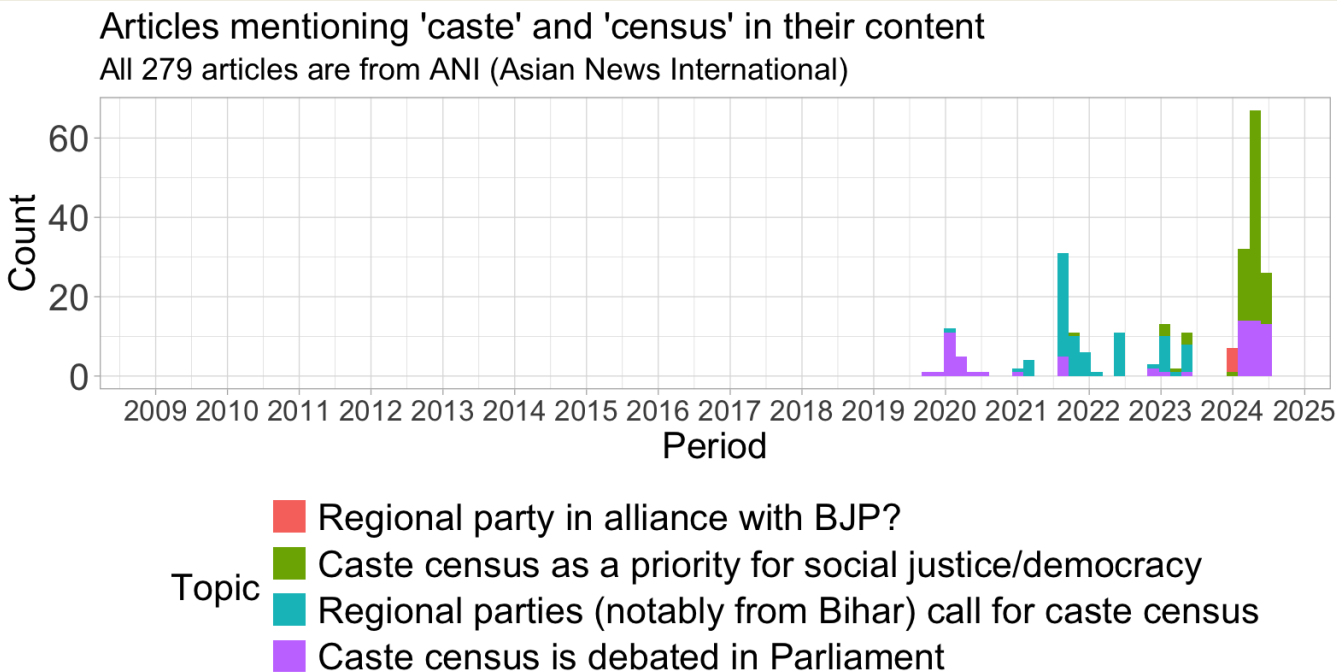
# Performativité statistique sous l'ère coloniale

- Les premières mesures statistiques de la caste : les recensements coloniaux (1871 à 1931)
  - Une administration statistique (+anthropologie coloniale) qui rigidifie et hiérarchise des catégories sociales
    - Un travail marqué par une théorie des races et un paradigme évolutionniste des administrateurs européens
    - Renforcé par le rôle des informateurs « indigènes », issus de milieux sociaux lettrés, très largement de hautes castes
    - => une « compréhension partagée » entre pensée évolutionniste et pensée brahmanique
    - => une vision hiérarchique et religieuse de la caste
- Un legs à ne pas négliger jusqu'à nous :
  - Chez Max Weber, Louis Dumont...

Compter la caste : un objectif qui s'est imposé dans le débat public

# Mesurer la caste par « justice sociale »

- Mesurer la caste : une demande politique des partis régionaux dominés par des castes « moyennes »
  - Se compter pour demander des mesures de justice sociale
  - Une demande satisfaite par le poids politique grandissant de ces partis



Source : Europresse (articles from ANI available since 2008 upto August 2024)

## INTENT TO CONDUCT CENSUS TO BE NOTIFIED MID-JUNE

# Long wait ends, Census by March 2027, will include caste count too

Umbrella OBC group unlikely; women's quota, delimitation to be linked to Census

### Why this reopens debate on delimitation, timeline

**LIZ MATHEW**  
NEW DELHI, JUNE 4

THE GOVERNMENT on Wednesday said the process of data collection for the Census, along with caste enumeration, would commence early next year to give a snapshot of the country's population as on March 1, 2027.

The last Census provided population data as on March 1, 2011. The next round of Census was due in 2020-2021, but was postponed due to the Covid pandemic outbreak.

The exercise this year will

**CONTINUED ON PAGE 2**

**MONSOON SESSION FROM JULY 21 PAGE 10**

**DEEPTIMAN TIWARY & MANOJ CG**  
NEW DELHI, JUNE 4

**EXPLAINED** THE ANNOUNCEMENT Wednesday of the decision to conduct Population Census-2027 in two phases

along with enumeration of castes has reopened the debate on delimitation, especially in the southern states, with Tamil Nadu Chief Minister M K Stalin seeking an answer from the Centre and accusing it of delaying the headcount to reduce the state's parliamentary representation.

**CONTINUED ON PAGE 14**

En attendant 2027...

# Les questions ouvertes sur la caste dans l'IHDS

1.12 Which caste/jati and sub caste/sub jati do you belong to?

1.12a Jati  ID12aNm

1.12b Sub-jati [IF ANY]  ID12bNm

1.13 Is this Brahmin, General/Forward OBC, SC, ST or Others?  ID13

Brahmin=1 OBC=3 ST=5  
 Forward/General (except Brahmin)=2 SC=4 Others=6

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
...	...

- Parmi les ménages hindous de l'Uttar Pradesh (N=2,958) :
  - 755 énoncés de jatis différents
  - 878 énoncés de « sous-jatis » différents
  - 1603 combinaisons distinctes de jatis / sous-jatis
- => Construction d'une classification à deux niveaux :
  1. La jati comme « auto-identification » (approche inductive)
  2. La catégorie de caste comme « localisation sociale » (approche déductive)

# Cinq grandes étapes de la standardisation

1. Uniformisation orthographique des énoncés de jatis : procédure semi-automatique
2. Classification automatique des ménages / castes suivant les co-occurrences de jatis et sous-jatis
3. Revue de littérature (travaux historiques + ethnographiques) pour saisir les « principes structurants » réputationnels des jatis et classement manuel des ménages
4. Tests (et corrections) de la nomenclature :
  1. Comparaison classification algorithmique et nomenclature manuelle
  2. Comparaison entre nomenclature et catégories administratives
  3. Structure sociale de la nomenclature et homogénéité interne des catégories
  4. Confrontation de la nomenclature aux enquêté·es dans les entretiens
5. De N=3,000 ménages à N=60,000 ménages...

# 1. Uniformisation orthographique

- Les jatis / sous-jatis comme des chaînes de caractères
  - Calcul des distances et de la similarité des séquences de caractères
  - Plusieurs distances possibles, principalement « distance de Levenshtein »
    - Nombre minimum de caractères à supprimer, insérer ou remplacer pour transformer une chaîne en une autre
- Logiciel **OpenRefine**
- Réduction d'1/3 du nombre de combinaisons distinctes jatis/sous-jatis

**Cluster & Edit column "ID12ANM"**

This feature helps you find groups of different cell values that might be alternative representations of the same thing. For example, the two strings "New York" and "new york" are very likely to refer to the same concept and just have capitalization differences, and "Gödel" and "Godel" probably refer to the same person. Find out more ...

Method: **nearest neighbor** Distance Function: **levenshtein** Radius: **1.0** Block Chars: **6** **64 clusters found**

Cluster Size	Row Count	Values in Cluster	Merge?	New Cell Value
4	178	<ul style="list-style-type: none"><li>• CHAMAR (173 rows)</li><li>• CHAMARA (3 rows)</li><li>• CHAMARE (1 rows)</li><li>• DCHAMAR (1 rows)</li></ul>	<input type="checkbox"/>	CHAMAR
3	5	<ul style="list-style-type: none"><li>• BRHAMAN (3 rows)</li><li>• BRRHAMAN (1 rows)</li><li>• BARHAMAN (1 rows)</li></ul>	<input type="checkbox"/>	BRHAMAN
3	7	<ul style="list-style-type: none"><li>• KAYASTH (3 rows)</li><li>• KAYASTA (2 rows)</li><li>• KAYASTHA (2 rows)</li></ul>	<input type="checkbox"/>	KAYASTH
3	11	<ul style="list-style-type: none"><li>• CHHATRIYA (6 rows)</li><li>• CHATRIYA (3 rows)</li><li>• SHATRIYA (2 rows)</li></ul>	<input type="checkbox"/>	CHHATRIYA
3	5	<ul style="list-style-type: none"><li>• RAIDASH (3 rows)</li><li>• RAIDASS (1 rows)</li><li>• RAIDAS (1 rows)</li></ul>	<input type="checkbox"/>	RAIDASH
3	6	<ul style="list-style-type: none"><li>• KHETRIYA (3 rows)</li><li>• KSHETRIYA (2 rows)</li><li>• KIVETRIYA (1 rows)</li></ul>	<input type="checkbox"/>	KHETRIYA

Control buttons: **Select All** **Unselect All** **Export Clusters** **Merge Selected & Re-Cluster** **Merge Selected & Close** **Close**

Right-side histograms:

- # Choices in Cluster: Range 2 — 4
- # Rows in Cluster: Range 0 — 180
- Average Length of Choices: Range 6.5 — 11.5
- Length Variance of Choices: Range 0 — 0.8170000000000001



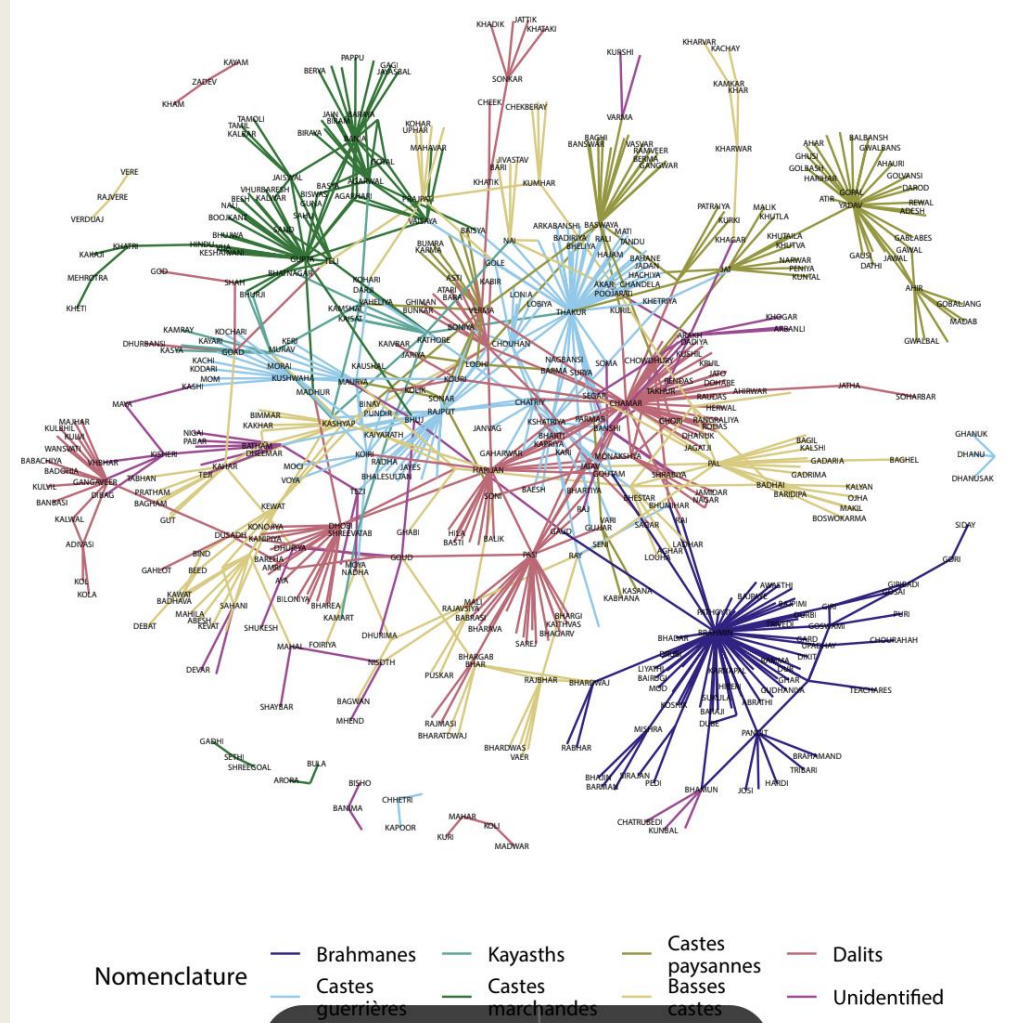
# 3. Distinctions réputationnelles des jatis

Category	Description	Varna status	Official category	Composition (among Hindus)	Self-identifications	Count
Dalit	Former untouchable caste groups	Rather unambiguous: perceived and self-acknowledged as untouchable, possible counter-cultural claim as Dalit or Harijan to denounce their stigmatization in the caste social order	Scheduled castes (93 percent), a few as Scheduled Tribes (4 percent)	29.2	CHAMAR	233
					HARIJAN	135
					JATAV	78
					PASI	76
					DHOBI	41
					KORI	32
					KHATIK	20
					MUSHAR	19
					GAUTAM	17
					GAUD	15
					VALMIKI	15
Low	Former craftspeople and agricultural labourer caste groups	Perceived but usually not claimed as Shudra	Largely counted as Other Backward Class (86 percent)	20.2	RAJBHAR	38
					PAL	29
					BHAR	27
					KUMHAR	26
					GADARIA	24
					PRAJPATI	24
					NISHAD	23
					VISHWAKARMA	21
					KAHAR	19
					RAJVERE	19
					KEVAT	18
					KASHYAP	17
					BIND	16
					SINGH	16
Agricultural	Small landowner caste groups	Disputed: self-identification to the Kshatriya or Vaishya, may be considered as Shudra	Largely counted as Other Backward Class (91 percent) or Forward (7 percent)	19.4	YADAV	174
					AHIR	114
					KURMI	46
					JAT	40
					VERMA	23
					PATEL	22
					GUJJAR	18
Merchant	Sanskritized and business caste groups	Rather unambiguously Vaishya	Either Other Backward Class (71 percent) or Forward (27 percent)	5.5	GUPTA	28
					TELI	24
					VAISHYA	17
					BANIA	15
Lordly	Former landowner caste groups	Rather unambiguously Kshatriya	Either Other Backward Class (49 percent) or Forward (46 percent)	13.3	RAJPUT	50
					KSHATRIYA	43
					LODHI	37
					THAKUR	35
					CHOUHAN	33
					KOIRI	20
					KUSHWAHA	18
MORYA	16					
Brahman	Dominant in the Hindu social order	Rather unambiguously Brahman	Largely counted as Forward (96 percent)	12.1	BRAHMIN	303

# 4.1. Test : nomenclature vs structure des énoncés

- La classification dans la structure des co-énonciations des jatis et sous-jatis
  - Des catégories relativement bien distinctes dans la structure des co-énonciations...
  - ... même si de nombreux « ponts » entre les auto-identifications

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



## 4.2. Nomenclature et catégories administratives

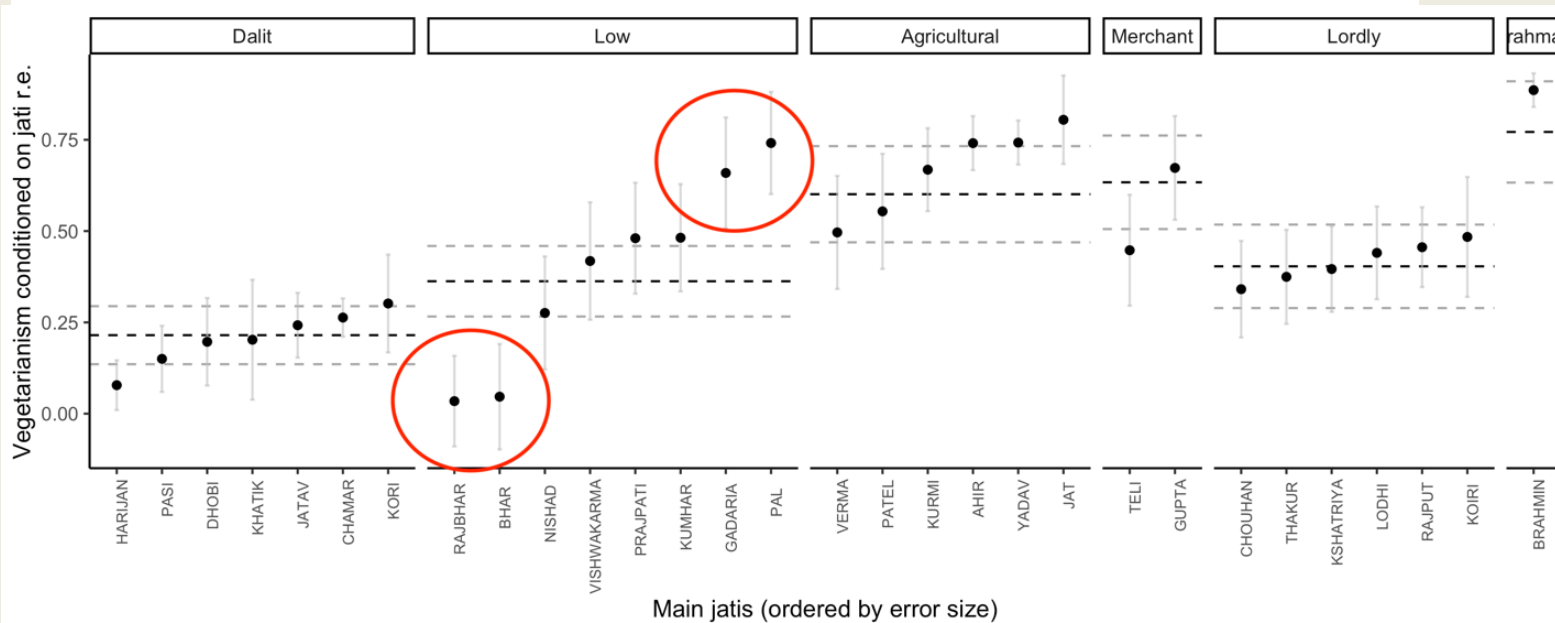
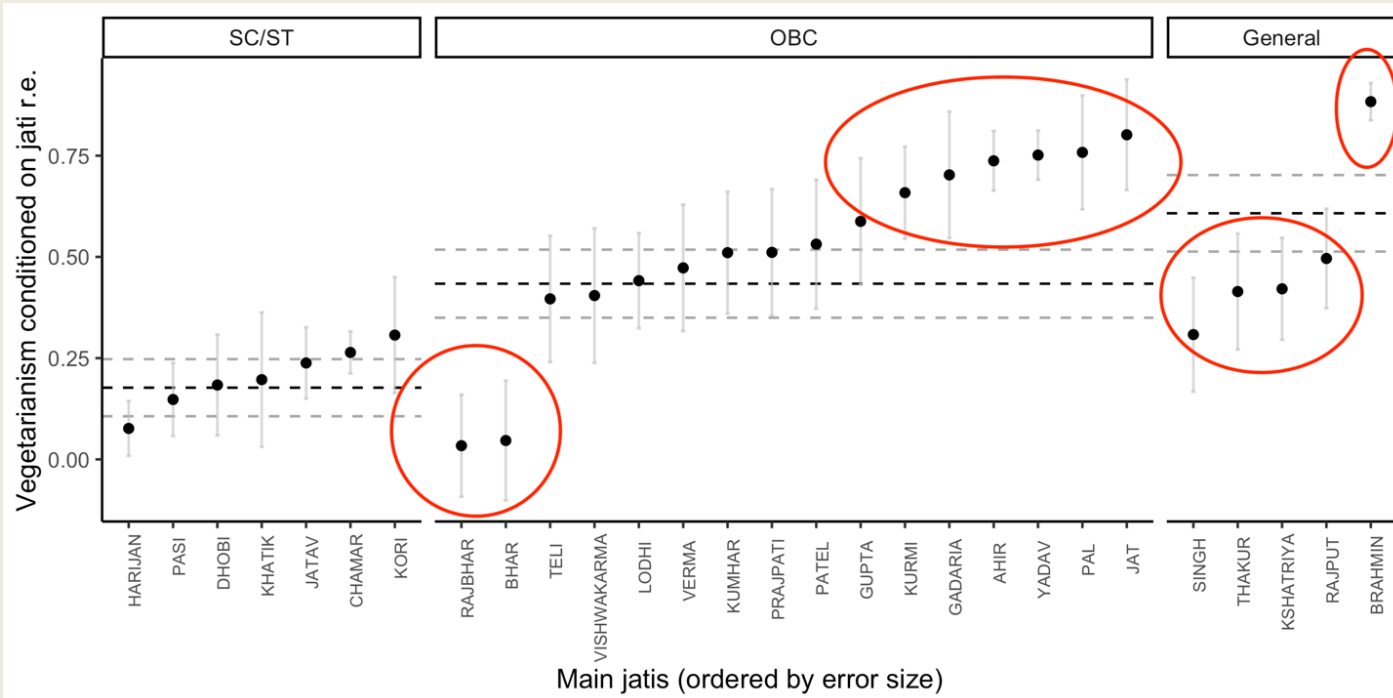
- Les catégories administratives comme un « gold standard test » ?

1.13 Is this Brahmin, General/Forward OBC, SC, ST or Others?  Brahmin=1  OBC=3  ST=5  ID13  
 Forward/General (except Brahmin)=2  SC=4  Others=6

- Sensibilité :
  - % identifiés comme Brahmanes (ou Dalit/SC) par les jatis aussi identifiés comme tels par variable ID13
- Spécificité :
  - % NON identifiés comme Brahmanes (ou Dalit/SC) par les jatis aussi NON identifiés comme tels par ID13

	Brahmanes	Dalit-SC
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

# 4.3. Nomenclature et homogénéité interne



- Sur différentes dimensions clefs testables de la nomenclature de caste (végétarisme, capital foncier, niveau de diplôme) :
  - Hétérogénéité interne de la nomenclature de caste plus faible que les catégories administratives
- Pouvoir prédictif de la nomenclature de caste plus élevé pour étudier la stratification sociale

## 4.4. Statistique et représentation « pratique »

- Environ 70 entretiens dans 6 districts à travers l'Uttar Pradesh
- Quelle auto-identification à la caste et rapport à la nomenclature ?

### Group practices

#### \*Religion & Caste

*I would to talk a bit about your religious and caste (jati) identity if you will. How do you identify yourself, in terms of religion? In terms of community/caste/jati?*

*Is this community classified as backward? (If yes, precise; SC, ST, OBC).*

*If Hindu, among these categories, which one does **your community** fall in:*

*-Brahman*

*-Kayasth*

*-Lordly castes, Rajput*

*-Merchant caste, Baniya*

*-Agricultural caste, Kisan*

*-Lower castes*

*-Dalits*

*Among the communities listed above, which are the ones that consume meat, buff/beef meat? (You can express it in %, e.g. "10% of Brahmans eat meat").*

- Des catégories qui « font sens » du point de vue des frontières et logiques de regroupement
- Les qualificatifs des catégories étaient plus contestés
  - « Dalit », terme repris dans la littérature académique, finalement peu mobilisé
  - « Agricultural », trop occupationnel
  - « Low », usage de « backward »

## 5. De 3,000 à 60,000 ménages

- Au cours de la thèse :
  - Accès aux données de réponse ouverte à l'appartenance de caste dans le *National Family Health Survey (2015-2016)*
  - Intérêt :
    - Plus grande précision statistique
    - Représentativité géographique fine (échelle des districts = 70 en Uttar Pradesh)
- « Nettoyage » :
  - Par uniformisation orthographique
  - Transformation des auto-identifications de jatis dans IHDS & NFHS par algorithme « soundex » [Yadav ; Yaadav ; Yadava ; ...] = « Y310 »
- Détermination probabilistique de la catégorie de caste (nomenclature) dans NFHS grâce aux « soundex » correspondants à chaque jati dans IHDS

Bilan

# Retour critique

- Environ 2-3 % de ménages : jati / catégorie de caste non assignée
- Un effet enquêteur dans les enquêtes ?
  - « Thakur, ye log apne aap ko Thakur bolate hai lekin OBC me ate »
- Un travail fastidieux de longue haleine...
  - Qui permet quand même d'apprendre sur son objet
  - Et de réfléchir à ce qu'est une « catégorie sociale »
- Fallait-il créer une « nomenclature réputationnelle » de catégories de castes en 6-7 catégories ?
  - Une nomenclature « pragmatique » vue la faiblesse de l'échantillon dans l'IHDS (N=3,000 ménages)
  - Moins nécessaire avec une taille d'échantillon aussi élevée que le NFHS (N=60,000 ménages)



[mathieu.ferry@uvsq.fr](mailto:mathieu.ferry@uvsq.fr)