

## Outline

- ▶ Social exclusion is often associated with lower educational outcomes
- ▶ effects of social exclusion and segregation not obvious
- ▶ We ask:  
What is the impact of social exclusion – in particular, linguistic distance between mother tongue and medium of instruction, on Scheduled Tribe (ST) education?
- ▶ Linguistic distance lowers educational outcomes in 2001
- ▶ Greater distance is associated with greater second language acquisition, which has a positive impact on educational outcomes
- ▶ Controlling for second language acquisition, suggests stronger negative impact of linguistic distance on education

## Background: Scheduled Tribes

- ▶ over 104 million people (8.6% of India)
  - ▶ historically disadvantaged due to geographical isolation / segregation
  - ▶ worst socio-economic outcomes among groups mandated for affirmative action
- | Category | Completion Rate (%) |         |        |           |          |
|----------|---------------------|---------|--------|-----------|----------|
|          | Literacy            | Primary | Middle | Secondary | Graduate |
| Total    | 64.84               | 50.10   | 33.74  | 23.88     | 6.67     |
| SC       | 54.69               | 38.29   | 22.63  | 13.45     | 2.66     |
| ST       | 47.10               | 29.31   | 16.66  | 9.76      | 1.76     |
- ▶ sparsely studied
  - ▶ little understanding of *within*-ST inequalities

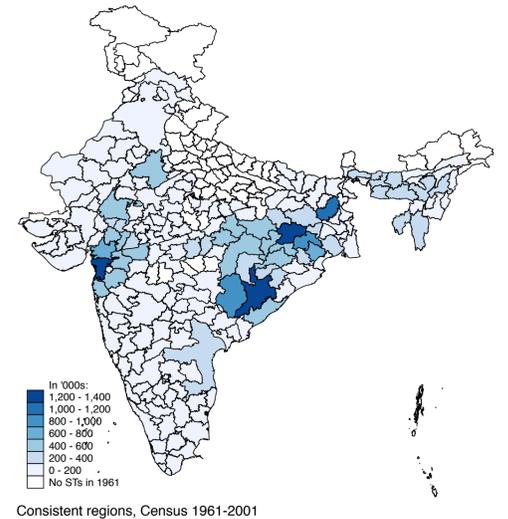


Figure 1: ST Population Distribution, 1961

## Data

District-level census data on

- ▶ for each ST in 1961, number of speakers of each mother tongue, and for each such linguistic group, number who spoke an additional language
- ▶ Literacy rates for individual STs in 1961
- ▶ Educational outcomes for individual STs in 2001
- ▶ 258 tribe groups, 527 mother tongues, 204 districts
- ▶ median
  - 5 districts per tribe
  - 8 tribes per district
  - 9 mother tongues per district
  - 6 mother tongues per tribe
- ▶ School-level information on medium of instruction from District Information System for Education (DISE), 2013-14
- ▶ Ethnologue language tree for linguistic distance

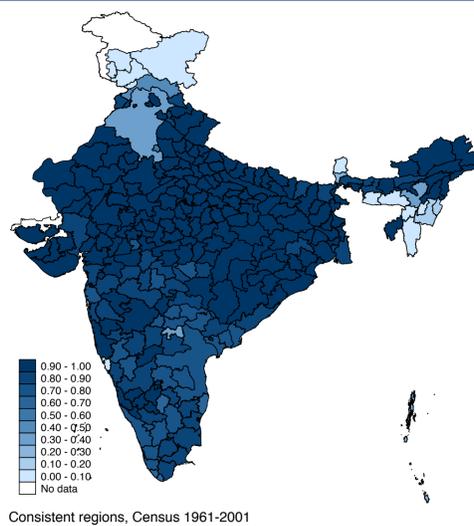


Figure 2: Fraction of schools using dominant language as medium of instruction (DISE, 2013-14)

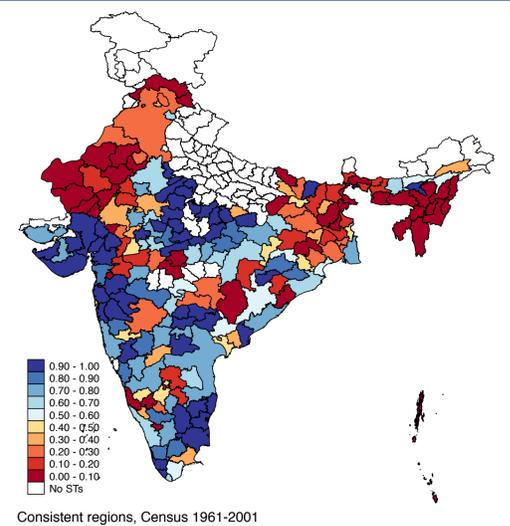
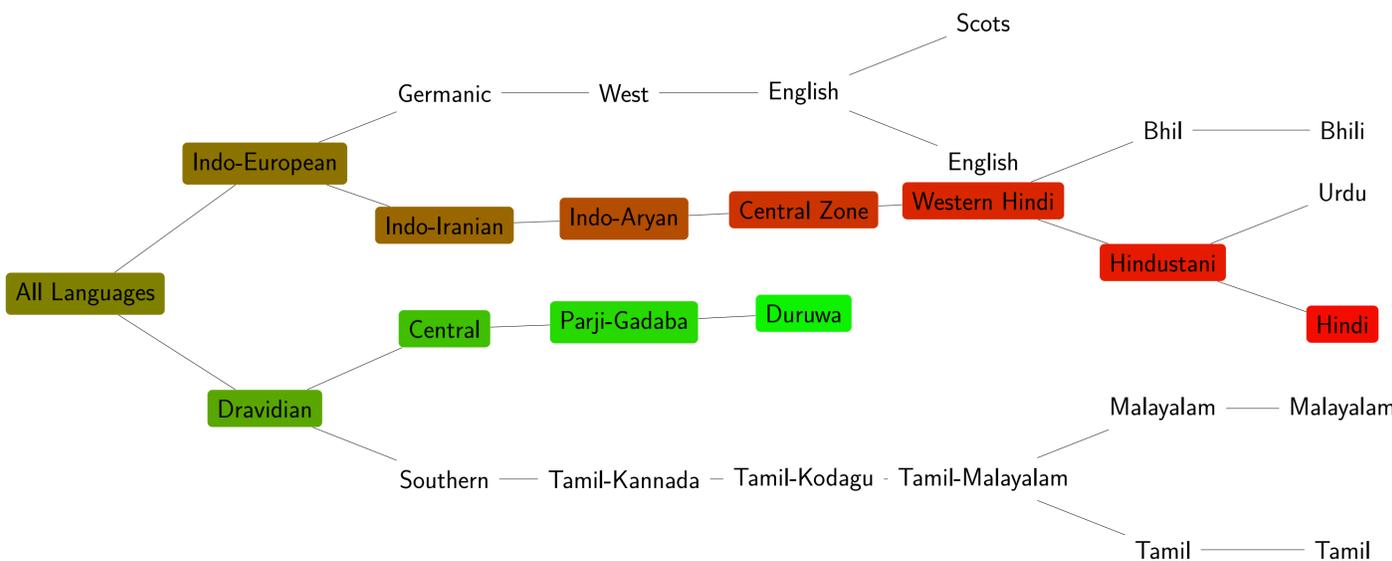


Figure 3: Fraction of ST Population with dominant language as mother tongue, 1961

## Ethnologue tree example



## Linguistic Distance

- ▶ measure distance between two languages as the number of nodes traversed in going from one to the other in the Ethnologue language tree distance from
  - ▶ Duruwa to Hindi = 11
  - ▶ Hindi to English = 10
- ▶ For each linguistic group in a tribe in a district, calculate linguistic distance to
  - ▶ dominant language spoken in the state [0 to 15; mean 6.05; stdev 4.98]
  - ▶ modal medium of instruction in the district, as per DISE 2013-14 [0 to 15; mean 5.71; stdev 4.95]
  - ▶ all media of instruction in the district, weighted by ST enrolment [0 to 14.9; mean 6.06; stdev 4.68]

## Estimation

Educational outcomes in 2001:

$$y_{ti}^{2001} = \beta_0 + \beta_1 \sum_l L_{lti} * \frac{Pop_{lti}}{ST\_Pop_i} + \beta_2 Literacy_{ti}^{1961} + \text{tribe}_t + \text{district}_i + \epsilon_i \quad (1)$$

Supplementary language acquisition:

$$\phi_{ti}^{1961} = \delta_0 + \delta_1 \sum_l L_{lti} * \frac{Pop_{lti}}{ST\_Pop_i} + \text{tribe}_t + \text{district}_i + \epsilon_i \quad (2)$$

Controlling for supplementary language acquisition:

$$y_{ti}^{2001} = \gamma_0 + \gamma_1 \phi_{ti}^{1961} * \sum_l L_{lti} * \frac{Pop_{lti}}{ST\_Pop_i} + \gamma_2 \sum_l L_{lti} * \frac{Pop_{lti}}{ST\_Pop_i} + \gamma_3 \phi_{ti}^{1961} + \gamma_4 Literacy_{ti}^{1961} + \text{tribe}_t + \text{district}_i + \epsilon_i \quad (3)$$

where  $L_{lti}$  is the linguistic distance from a mother tongue  $l$  spoken by tribe  $t$ ;  $Literacy_{ti}^{1961}$  is the literacy rate in 1961 for tribe  $t$  in district  $i$ ;  $\phi_{ti}^{1961}$  is the fraction of tribe  $t$  in district  $i$  that speaks any secondary language; and  $\text{district}_i$  and  $\text{tribe}_t$  are the respective district and tribe fixed effects.

## Results

Educational outcomes in 2001:

	Literacy	Primary	Middle	Matric	Graduate
<b>A. Dominant state language</b>					
Distance	-0.036*** (0.000)	-0.035*** (0.000)	-0.032*** (0.000)	-0.032*** (0.000)	-0.015*** (0.000)
Literacy 1961	0.016 (0.301)	0.024 (0.154)	0.024 (0.133)	0.021 (0.120)	0.007 (0.176)
Constant	0.006 (0.960)	-0.125 (0.395)	-0.358*** (0.005)	-0.336*** (0.004)	-0.125*** (0.007)
<b>B. District modal medium of instruction</b>					
Distance	-0.036*** (0.000)	-0.034*** (0.000)	-0.029*** (0.000)	-0.030*** (0.000)	-0.014*** (0.001)
<b>C. All district media of instruction, weighted by ST enrolment</b>					
Distance	-0.039*** (0.000)	-0.036*** (0.000)	-0.033*** (0.000)	-0.033*** (0.000)	-0.016*** (0.000)
Observations	2,038	2,038	2,038	2,038	2,038

Note: Tribe and District fixed effects included. Heteroskedasticity-robust standard errors in parentheses.

## Results (contd.)

Subsidiary language acquisition:

	Dominant	Modal	All
Distance	0.103*** (0.000)	0.109*** (0.000)	0.108*** (0.000)
Constant	-0.059 (0.673)	-0.054 (0.701)	-0.080 (0.578)
Observations	2,038	2,038	2,038

Controlling for subsidiary language acquisition:

	Literacy	Primary	Middle	Matric	Graduate
<b>Dominant state language</b>					
Distance	-0.042*** (0.000)	-0.040*** (0.000)	-0.042*** (0.000)	-0.041*** (0.000)	-0.021*** (0.000)
<b>Subsidiary</b>					
Distance	0.035 (0.398)	0.076 (0.107)	0.118** (0.017)	0.116** (0.015)	0.049 (0.105)
Distance × Subsidiary	0.010 (0.730)	-0.012 (0.684)	-0.010 (0.743)	-0.009 (0.751)	0.005 (0.782)
Literacy 1961	0.015 (0.325)	0.022 (0.168)	0.020 (0.151)	0.017 (0.139)	0.005 (0.228)
Constant	0.008 (0.946)	-0.122 (0.411)	-0.353*** (0.006)	-0.331*** (0.005)	-0.122*** (0.007)
Observations	2,038	2,038	2,038	2,038	2,038

Note: Tribe and District fixed effects included. Heteroskedasticity-robust standard errors in parentheses.