

LINGUISTIC DIVERSITY, OFFICIAL LANGUAGE CHOICE AND NATION BUILDING: THEORY AND EVIDENCE

David D. Laitin (Stanford University)
and
Rajesh Ramachandran (Goethe University)

The International Political Economy Society Meeting
Stanford University

November 13-14th, 2015

- ▶ Striking development of post-WW-II era is the birth of a number of failing states.
 - ▶ Underlying roots of ineffective states found to lie “in the absence of common interests reinforced by noncohesive institutions” (Besley and Persson, 2010, 2011b,a)
- ▶ A large body of literature attributes ELF to be an important factor underlying “absence of common interests” (Alesina and Ferrara 2005, Easterly and Levine 1997, La Porta et al. 1999)
 - ▶ However little consensus on the specific mechanisms through which ELF operates to affect outcomes.

▶ qtrs

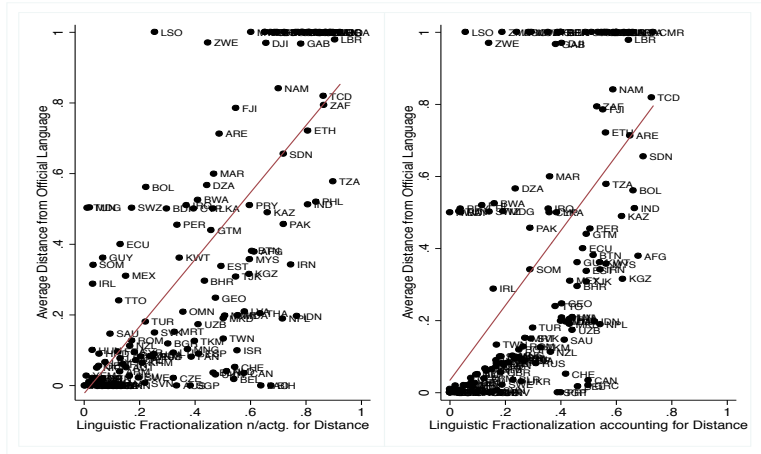
Our Paper

- ▶ Provide a theoretical framework and empirical evidence to analyze the channel through which linguistic diversity (LD) operates.
- ▶ Linguistically diversity amplifies the problem of coordinating on the choice of an indigenous language as official, and increases the probability of maintaining the colonial language.
- ▶ Choice of a colonial language that is not spoken indigenously and is very distant from the local languages negatively affects the levels of human capital in society.
- ▶ Our main contention - negative effects attributed to linguistic diversity in the cross-country literature are primarily mediated through the channel of language policy.

Linguistic Fractionalization and Official Language Choice

- ▶ Assumption - cost of human capital formation increasing in the distance to the official language.
- ▶ Groups derive utility from *Absolute payoff* and *Relative standing*.
- ▶ An increase in linguistic diversity in our framework has two effects:
 - ▶ Reduces the payoff for group i from coordinating on language j as cost of human capital formation increases.
 - ▶ Makes the relative gap between group i and j higher - if relative ranking or fairness concerns are present.
- ▶ Consider various equilibrium selection concepts - Risk dominance, Payoff dominance, Security, Fairness or Focal Points - to show increase in LF \Rightarrow Greater probability of retaining colonial language.

Average distance from official language and Linguistic Fractionalization



mesr

table

Why distance from official language matters

- ▶ Provide outline; refer to Laitin and Ramachandran (2015) for evidence.
- ▶ Two main facets of socio-economic development that our theory links to official language choice are:
 - ▶ Human capital formation
 - ▶ Health
- ▶ Individuals are utility maximizers and choose human capital and preventive health behavior to maximize wellbeing.
- ▶ Two key assumptions confirmed by L&R (2015) underlie our theory:
 - ▶ Higher the distance from the official language higher the cost of human capital formation.
 - ▶ Lower the exposure to the official language higher the cost of human capital formation.

Revisiting the cross-country literature on diversity and development

- ▶ Alesina and Ferrera (2005) and Easterly and Levine (1997) effect of LD on GDP per capita.
 - ▶ Find strong negative effects of diversity on growth.

▶ table

Revisiting the cross-country literature on diversity and development

- ▶ Alesina and Ferrera (2005) and Easterly and Levine (1997) effect of LD on GDP per capita.

- ▶ Find strong negative effects of diversity on growth.

▶ table

- ▶ La Porta et. al (1999) find negative effects of diversity on quality of government

- ▶ Corruption and infant mortality rates.

▶ table

▶ table

Revisiting the cross-country literature on diversity and development

- ▶ Alesina and Ferrera (2005) and Easterly and Levine (1997) effect of LD on GDP per capita.

- ▶ Find strong negative effects of diversity on growth.

▶ table

- ▶ La Porta et. al (1999) find negative effects of diversity on quality of government

- ▶ Corruption and infant mortality rates.

▶ table

▶ table

- ▶ Alesina et. al (2001), Alesina et. al (2003), Desmet et. al (2009)

- ▶ Diversity reduces redistribution.

▶ table

Application to the paper Artificial States by Alesina et. al (2011)

- ▶ Alesina et. al construct two measures of state artificiality:
 - ▶ (i) Share of partitioned ethnicities; (ii) Straightness of land borders.
- ▶ Show that higher degree of artificiality is associated with poorer economic outcomes. [▶ quote](#)
- ▶ One immediate consequence of partitioning ethnicities is the rise in ELF.
 - ▶ Our theory shows that this should exacerbate coordination on official languages.
 - ▶ Replicate tables from Alesina et. al (2011) but additionally control for ADOL.
- ▶ Main findings:
 - ▶ (i) Effect of ADOL larger (ii) Magnitude on the coefficients of artificiality reduces to around half its size.

Conclusions

- ▶ Presented theoretical and empirical evidence on a channel through which linguistic fractionalization affects socio-economic development.
- ▶ Empirical evidence suggests ADOL is an important omitted variable; and empirically, at least, all negative effects seem to stem through the channel of official language choice.
- ▶ Explored applications of our theoretical framework to existing empirical studies.
- ▶ Made a first step in identifying a factor amenable to policy choices, which can help create cohesive and inclusive societies.

THANK YOU

Quotes on Fractionalization

- ▶ “Fragmented societies are often more prone to poor policy management and pose more politico-economic challenges than homogenous ones; it is easy to find rather voluminous evidence on this point.” (Alesina and La Ferrara 2005)
- ▶ Banerjee et al. (2005) go so far as to describe ELF as “one of the most powerful hypotheses in political economy”

▶ back

Measuring official language choice and its implications

- ▶ In order to conceptualize the notion of language distance the measure based on language trees is used.
- ▶ The distance between any two language i and j is defined as:

$$d_{ij} = 1 - \left(\frac{\# \text{ of common nodes between } i \text{ and } j}{\frac{1}{2}(\# \text{ of nodes for language } i + \# \text{ of nodes for language } j)} \right)^\lambda$$

- ▶ Example - Bawaen and Indonesian

- ▶ Both belong to the Austronesian Language Family.
- ▶ Share 3 common nodes; Bawaen - 5 nodes; Indonesian - 7 nodes

- ▶ $d_{ij} = 1 - \left(\frac{3}{6}\right)^\lambda$

- ▶ Example - Spanish and Indonesian

- ▶ Different language families - $d_{ij} = 1$.

▶ tree

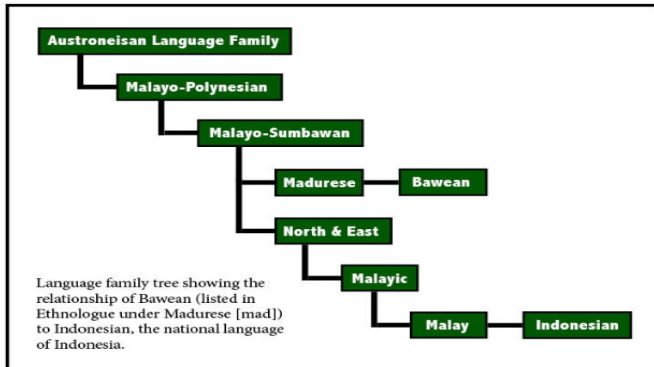
▶ back

Country level measure

- ▶ Consider all language groups comprising at least 1% of population (data from Fearon, 2003)
- ▶ Calculate distance from official language (d_{io}) for each linguistic group i in the country.
 - ▶ Official language is the language in which the first organic laws or constitution has been written.
 - ▶ Alternatively - language of secondary education and higher courts.
- ▶ The average distance from the official language for any country i is calculated as:

$$ADOL_i = \sum_{j=1}^n P_{ij} d_{jo}$$

▶ back



Language family tree showing the relationship of Bawean (listed in Ethnologue under Madurese [mad]) to Indonesian, the national language of Indonesia.

Figure: Family Tree

▶ back

The role of writing tradition

- ▶ The second important factor our theory highlights is the availability of a writing tradition.
- ▶ In the absence of a writing tradition states first have to invest to create a standardized script, orthography, and vocabulary.
- ▶ Two alternative interpretations:
 - ▶ Imposing a fixed cost.
 - ▶ Increasing the payoff uncertainty associated with coordinating on a language that has no writing script.
- ▶ Easy to show there exist fixed costs such that two societies with same levels of linguistic diversity but :
 - ▶ The polity with a writing script coordinates on the indigenous language.
 - ▶ The polity without a writing script sticks to the status quo.

Empirical evidence for the theoretical framework

	(1)	(2)	(3)	(4)	(5)
Dummy for whether country has a writing tradition	-0.612*** (0.0375) [-0.728]	-0.598*** (0.0408) [-0.711]	-0.613*** (0.0380) [-0.728]	-0.413*** (0.0740) [-0.490]	-0.358*** (0.0635) [-0.424]
Linguistic fractionalization accounting for distance	0.655*** (0.0752) [0.366]	0.667*** (0.0768) [0.373]	0.646*** (0.0779) [0.360]	0.615*** (0.0750) [0.343]	
Log GDP per capita at independence in 1990 US		-0.0186 (0.0148) [-0.0453]		0.0205 (0.0201) [0.0500]	0.0389* (0.0222) [0.0955]
Log Population in 1500 CE			0.00738 (0.00793) [0.0340]	0.00834 (0.00962) [0.0384]	0.00762 (0.00953) [0.0349]
Linguistic fractionalization n/actg. for distance					0.513*** (0.0670) [0.389]
Continent Dummies	No	No	No	Yes	Yes
Observations	131	131	130	130	126
R-squared	0.815	0.817	0.816	0.846	0.848

An instrumental variable approach

- ▶ Presence of a writing tradition is a strong predictor of ADOL.
 - ▶ However countries which possess writing traditions and those that do not, arguably differ on several unobservable characteristics.
- ▶ We draw from the work of Diamond (1998) and propose using distances from the sites of invention as an instrument for possessing a writing tradition.
 - ▶ Writing was independently invented in Mesopotamia (3200 BCE), Mesoamerica (600 BCE) and China (1200 BCE).
 - ▶ Diffused to rest of the world from these sites.
- ▶ To operationalize the measure:
 - ▶ Calculate the Great-Circle distance using the Haversine formula from each of the three sites.
 - ▶ Take the minimum of the distances as our measure of distance

IV Regressions

	(1)	(2)	(3)	(4)	(5)
Dummy for whether country has a writing tradition	-0.74*** (0.081) [-0.88]	-0.75*** (0.10) [-0.89]	-0.75*** (0.083) [-0.89]	-0.82** (0.41) [-0.97]	
Linguistic fractionalization accounting for distance	0.58*** (0.087) [0.32]	0.57*** (0.097) [0.32]	0.56*** (0.089) [0.31]	0.54*** (0.11) [0.30]	
Log GDP per capita at independence in 1990 US		0.0068 (0.024) [0.016]		0.034 (0.027) [0.083]	
Log Population in 1500 CE			0.0079 (0.0089) [0.036]	0.021 (0.016) [0.096]	
Continent Dummies	No	No	No	Yes	
Observations	131	131	130	130	
R-squared	0.795	0.792	0.793	0.785	

Distance from Site of Invention of Writing	-0.000090*** (0.000017) [-0.42]	-0.000075*** (0.000017) [-0.35]	-0.000089*** (0.000017) [-0.42]	-0.000027* (0.000014) [-0.12]
F-Stat	21.5	19.1	14.2	44.2

▶ back

Falsification tests - Regressing the distance from sites of invention

	(1) Avg. Prot. Against Expr. Rights	(2) Social Infrastructure	(3) Constraints on the Executive
Distance from Site of Invention of Writing	-1.8e-06 (7.4e-06) [-0.021]	-9.4e-06 (0.000011) [-0.080]	0.000060 (0.000080) [0.062]
P-Value	0.81	0.40	0.45
F-Stat	0.057	0.71	0.57

▶ back

Dependent Variable - Transfers & Subsidies as share of GDP (74-94)

	(1)	(2)	(3)	(4)
Linguistic fractionalization accounting for distance	-8.126*** (2.902) [-0.264]	1.158 (4.167) [0.0377]	0.116 (4.215) [0.00379]	1.529 (4.608) [0.0498]
Average Distance from Official Language		-9.255*** (2.560) [-0.503]	-8.230*** (2.569) [-0.447]	-10.06** (4.183) [-0.547]
Legal Origin - Dummies	No	No	Yes	Yes
Africa and Asia Dummy	No	No	No	Yes
Observations	68	68	68	68
R-squared	0.070	0.232	0.467	0.490

* $p < .10$; ** $p < .05$; *** $p < .01$. Robust SE's in parenthesis and standardized coefficients in square brackets.

Dependent Variable - Log GDP per capita

	(1)	(2)	(3)	(4)
Linguistic fractionalization accounting for distance	-1.362*** (0.501) [-0.233]	0.954* (0.566) [0.163]	0.901 (0.568) [0.154]	0.727 (0.508) [0.124]
Average Distance from Official Language		-2.271*** (0.266) [-0.691]	-2.403*** (0.293) [-0.731]	-1.548*** (0.489) [-0.471]
Legal Origin - Dummies	No	No	Yes	Yes
Africa and Asia Dummy	No	No	No	Yes
Observations	126	126	126	126
R-squared	0.054	0.375	0.432	0.498

* $p < .10$; ** $p < .05$; *** $p < .01$. Robust SE's in parenthesis and standardized coefficients in square brackets.

Dependent Variable - Corruption Score from ICRG

	(1)	(2)	(3)	(4)
Linguistic fractionalization accounting for distance	-1.773* (0.936) [-0.185]	0.00491 (1.208) [0.000512]	-0.348 (1.225) [-0.0363]	0.102 (1.346) [0.0107]
Average Distance from Official Language		-1.551** (0.656) [-0.292]	-1.316* (0.697) [-0.248]	-1.934* (1.085) [-0.364]
Legal Origin - Dummies	No	No	Yes	Yes
Africa and Asia Dummy	No	No	No	Yes
Observations	96	96	96	96
R-squared	0.034	0.085	0.183	0.208

* $p < .10$; ** $p < .05$; *** $p < .01$. Robust SE's in parenthesis and standardized coefficients in square brackets.

▶ back

Dependent Variable - Infant Mortality Rate in 2010

	(1)	(2)	(3)	(4)
Linguistic fractionalization accounting for distance	71.19*** (18.89) [0.328]	-22.74 (17.11) [-0.105]	-19.80 (17.14) [-0.0911]	-1.307 (17.14) [-0.00601]
Average Distance from Official Language		94.16*** (8.098) [0.769]	91.88*** (8.974) [0.750]	45.36*** (16.84) [0.370]
Legal Origin - Dummies	No	No	Yes	Yes
Africa and Asia Dummy	No	No	No	Yes
Observations	131	131	131	131
R-squared	0.107	0.512	0.521	0.582

* $p < .10$; ** $p < .05$; *** $p < .01$. Robust SE's in parenthesis and standardized coefficients in square brackets.

▶ back

Dependent Variable - Log Output per Worker

	(1)	(2)	(3)	(4)
Linguistic fractionalization accounting for distance	-1.545*** (0.391) [-0.332]	0.541 (0.420) [0.116]	0.486 (0.408) [0.105]	0.196 (0.366) [0.0423]
Average Distance from Official Language		-2.006*** (0.217) [-0.770]	-2.064*** (0.208) [-0.795]	-1.096*** (0.329) [-0.423]
Legal Origin - Dummies	No	No	Yes	Yes
Africa and Asia Dummy	No	No	No	Yes
Observations	94	94	93	93
R-squared	0.110	0.503	0.546	0.610

* $p < .10$; ** $p < .05$; *** $p < .01$. Robust SE's in parenthesis and standardized coefficients in square brackets.

- ▶ “When states represent people put together by outsiders, these peoples may find it more difficult to reach consensus on public goods delivery and the creation of institutions that facilitate economic development, compared to states that emerged in a homegrown way”.

▶ back

Dependent Variable - Log GDP per capita

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Average Distance from	-1.23***	-1.37***	-0.90*	-0.66	-1.73***	-1.34***	-1.69***	-1.38***	-1.07***
Official Language	(0.40)	(0.39)	(0.51)	(0.57)	(0.47)	(0.40)	(0.35)	(0.42)	(0.32)
	[-0.43]	[-0.48]	[-0.31]	[-0.23]	[-0.60]	[-0.47]	[-0.59]	[-0.48]	[-0.37]
First principal component	0.34***	0.25**	0.077	0.27**	0.19	0.25**	-0.052	0.24**	0.34***
	(0.10)	(0.11)	(0.14)	(0.12)	(0.12)	(0.11)	(0.15)	(0.12)	(0.086)
	[0.42]	[0.31]	[0.095]	[0.33]	[0.23]	[0.31]	[-0.064]	[0.30]	[0.43]
Second principal component	0.023	0.043	-0.078	0.021	0.011	0.041	-0.069	0.044	0.056
	(0.090)	(0.078)	(0.077)	(0.077)	(0.083)	(0.079)	(0.078)	(0.079)	(0.080)
	[0.022]	[0.042]	[-0.076]	[0.021]	[0.011]	[0.040]	[-0.067]	[0.043]	[0.055]
Climate, zone A (hot, rainy)		-0.51*	-0.28	-0.54**	-0.39	-0.52*	-0.36	-0.52*	-0.39
		(0.27)	(0.28)	(0.27)	(0.27)	(0.27)	(0.25)	(0.28)	(0.26)
		[-0.17]	[-0.093]	[-0.18]	[-0.13]	[-0.18]	[-0.12]	[-0.17]	[-0.13]
Africa			-1.80***	-0.59*					
			(0.46)	(0.33)					
			[-0.77]	[-0.25]					
Latin America			-1.49***		-0.41				
			(0.30)		(0.26)				
			[-0.55]		[-0.15]				
Asia and Oceania			-1.10**			0.15			
			(0.48)			(0.32)			
			[-0.24]			[0.032]			
Europe			-0.34				1.16***		
			(0.51)				(0.38)		
			[-0.12]				[0.40]		
Middle East			-1.17**					-0.056	
			(0.45)					(0.40)	
			[-0.20]					[-0.0097]	
North America									1.74***
									(0.28)
									[0.18]
Observations	71	71	71	71	71	71	71	71	71
R-squared	0.660	0.683	0.761	0.698	0.696	0.684	0.728	0.683	0.709