

Language difference and regional authority

Sarah Shair-Rosenfield^a, Arjan H. Schakel^b, Sara Niedzwiecki^c, Gary Marks^{d,e}, Liesbet Hooghe^{d,e} and Sandra Chapman-Osterkatz^{d,*}

^aDepartment of Government, University of Essex, Colchester, UK; ^bDepartment of Comparative Political Science, University of Bergen, Bergen, Norway; ^cPolitics Department, University of California, Santa Cruz, USA; ^dDepartment of Political Science, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; ^eRobert Schuman Centre, European University Institute, Florence, Italy

ABSTRACT

This paper draws on research in geography, linguistics, and political science to explain the incidence of language regions and their effect on regional authority. It conjectures a chain of mechanisms beginning with the physical and political barriers to human interaction and culminating with contemporary patterns of regional authority. Using data on 1767 regions in 95 countries, it finds causal power in the claim that the linguistic distinctiveness of a region reflects the ratio of internal interaction to external interaction. Finally, the effects of a language region for regional authority depend decisively on the openness of the political regime.

KEYWORDS Multilevel governance; language; regional authority; measurement; regions; decentralization

When does a language region exist within a state, and what are the consequences for the territorial structure of authority? These questions are fundamental to research published in *Regional & Federal Studies* over the past thirty years. Yet they are extremely challenging for both empirical and theoretical reasons. In the first place, comparative data at the subnational level is sparse. This has limited our ability to exploit the inferential power of subnational comparison in explaining variation. Whereas the case-study literature has been attentive to variation among language communities within countries, broad comparative studies have had to take an aggregate national approach.

Second, the study of language difference is a theoretical as well as empirical challenge for it overarches an astonishing diversity of disciplines: geography, linguistics, political science, evolutionary biology, and physics. There

CONTACT Liesbet Hooghe hooghe@unc.edu Department of Political Science, UNC-Chapel Hill, Hamilton Hall, Chapel Hill, NC 27599, USA

*Author names are in reverse alphabetical order; all have contributed equally.

Supplemental data for this article can be accessed <https://doi.org/10.1080/13597566.2020.1831476>

appears to be a sharp trade-off between validity and generalizability. As one moves from the physical sciences to the social sciences, the focus of research shifts from mathematical models that abstract from time and place to research that compares language difference in particular social and political settings.

We seek to bridge these disciplinary islands by theorizing a chain of mechanisms that allows us to generalize about the settings that produce language difference. We take individual regions – defined as the most authoritative territorial jurisdictions within states – as our unit of analysis. Rather than treat regions as components of states, we treat states as composed of regions. This allows us to take a disaggregated view of the sources and consequences of language difference, and to theorize variation within – as well as among – states (Giraudy, Moncada, and Snyder 2019; Hooghe and Marks 2016; Jeffery and Schakel 2013).¹ A regional approach to language difference detects far greater variation than at the national level, and it allows the researcher to exploit the inferential benefits of comparison within countries, while holding national variables constant (Harbers, Bartman, and van Wingerden 2019; Hooghe et al. 2016: ch. 1).

We begin with the idea that barriers to human interaction shape the territorial pattern of language reproduction. Figure 1 shows that barriers to interaction produce language difference. The insight comes from Darwinian speciation and it motivates mathematical studies of language difference (Prochazka and Vogla 2017 for an overview). Here we pin down these models in empirically disconfirmable expectations about the effect of islands and peripheral regions, and we find that the spatial location of human settlement does indeed have a statistically significant and substantively large effect on language difference.

However, patterns of interaction among communities are humanly shaped as well as environmentally determined. Our prior is that a political regime

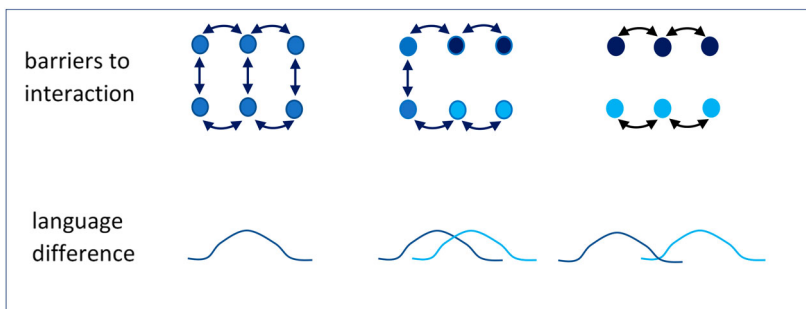


Figure 1. Barriers to interaction and language difference. Source: schematic representation of Darwinian speciation (Nosil, Harmon, and Seehausen 2009; Safran and Nosil 2012) applied to language difference.

raises the transaction cost of interaction beyond its borders and lowers that within. Hence the effect of physical geography in our model (Figure 2) is mediated by the effect of a state or empire in shaping patterns of human interaction by bringing communities within its rule into closer contact. The idea that regimes frame interaction and, hence, language difference has several testable implications. We find that a history of pre-colonial insulation strongly predicts the incidence of language regions. In addition, language difference is (a) enhanced when a region with a history of independent statehood is absorbed in a larger unit and (b) suppressed under inward migration.

Our final step is to probe the effect of language difference for regional authority. There is no monotonic association. Again, it depends on the political regime. Democracies provide space for the mobilization of demands for self-rule and incentivize rulers to compromise. Authoritarian regimes, by contrast, tend to clamp down on language regions because they fear that their opponents will exploit the opportunities presented by regional self-rule. So language difference is double-edged: in democratic regimes it enhances regional authority; in authoritarian regimes it diminishes regional authority.

Figure 2 summarizes our model as a series of questions for which we provide causal answers. How confident can one be that the independent

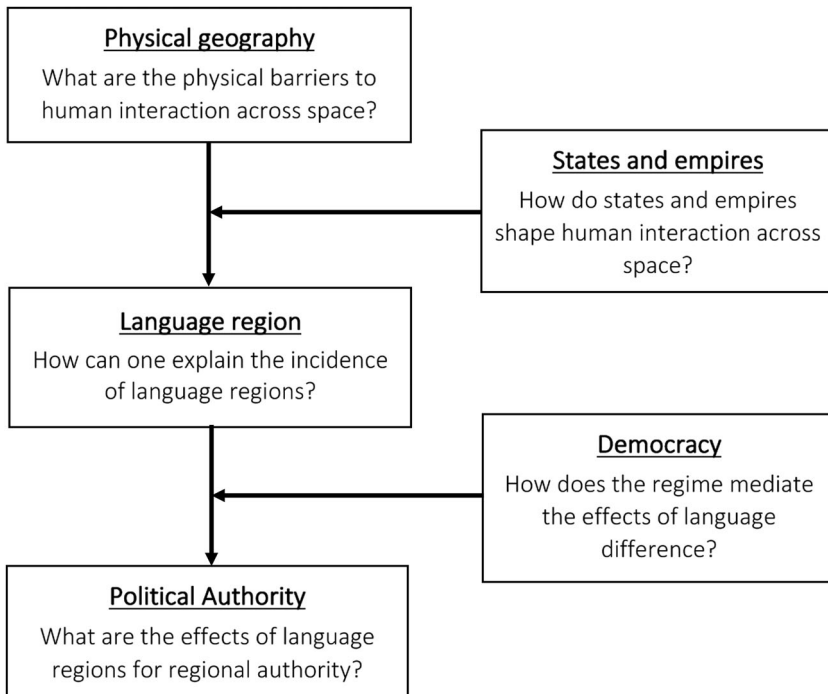


Figure 2. Causes and consequences of language difference. Source: authors' production.

variables we theorize are causally prior to the phenomena we wish to explain? Our response is to adopt a strategy of distal explanation. The existence of regions having a language of communication that differs from that of the national government is historically and, we suspect, causally prior to contemporary regional authority. To explain the incidence of language regions, we hypothesize the effect of political events in past centuries: colonial migration; a history of independent statehood; and multilevel governance prior to colonization. Fifty percent of the cases of independent statehood and multilevel governance that we conjecture as sources of contemporary language difference are dated prior to 1600AD, seventy percent are prior to 1800, and all, including colonial migration, refer to events before 1900.

We test our expectations using two datasets we have produced: the updated and expanded Regional Authority Index (RAI) and a new dataset estimating language difference in 1749 regions. The new release of the RAI broadens geographical coverage to 95 countries, extends the period from 1950 through 2018, and encompasses metropolitan regions. The conceptual and theoretical basis of the measure is developed in Hooghe, Marks, and Schakel (2010) and Hooghe et al. (2016). The dataset on language difference expands the Rokkan region dataset (Hooghe and Marks 2016) and includes region-level data on religion, prior state cores, prior multilevel governance, and geographical distance.

It is worth stressing that our analysis focuses on language difference while putting aside other features of culture and ethnicity. While language difference is often an important ingredient in ethno-cultural conflict, ethnicity and culture encompass a *mélange* of sometimes contradictory elements including race and religion (Brubaker 2013; Laitin 2000). Generalization about ethnicity can fall into the trap of essentialism by ignoring this multiplicity of elements. It is often argued, as if it were self-evident, that 'ethnic distinctions do not depend on the absence of mobility, contact and information' (Barth 1969, 9).² This may apply to religion and race, but we do not believe it applies to language. Our prior is that language is causally distinct from the remaining phenomena that are grouped under 'culture' and deserves separate analysis.

Next, we present a theoretical framework for explaining language difference and its effect on regional authority. We then describe the variables used in the analysis along with the new datasets before assessing the validity of our theory. We conclude by noting implications of this study for future research.

Theorizing language difference

Why do two distinct languages exist among the 1.4 million people who live in Trinidad & Tobago? Why is Papua New Guinea home to over 800 languages?

Why do just one in five regions in today's Peru have non-Spanish speaking majorities? Each of these cases involves context-specific features that draw the analyst into the histories of diverse peoples. However, our purpose here is to identify some general principles that help us explain why those living in a certain kind of region can be expected to speak a language that is different from the national core.

This project lies at the intersection of political science, geography, and linguistics, and we draw on each in framing a causal model. Despite their contrasting disciplinary commitments, researchers in each of these fields predict language difference by investigating patterns of interaction among language communities in the form of information exchange and migration. This yields a fundamental proposition that motivates our theory: The linguistic distinctiveness of a group depends on the ratio of internal interaction to external interaction.³ This claim can be unfolded in empirically testable statements that provide a fix on where one might find a language region.

Geography

To understand the effect of physical geography for language difference, it is useful to consider speciation, which as Charles Darwin (1882, 90) noted, is 'curiously parallel' to the formation of different languages.⁴ Evolutionary biologists find that geographical isolation plays a major role in speciation (Sobel et al. 2010). Perhaps the most famous example is Darwin's discovery that once finches from a common ancestor were established in the Galápagos islands, they became isolated from each other and mutated into distinct species.

Similarly, evolutionary models suggest that geographical isolation can be decisive for language extinction, invasion, coexistence, and diversity (Solé, Corominas-Murtra, and Fortuny 2010, 1647). Summarizing the results of computation models, Patriarca and Heinsalu (2009, 10) observe that 'geographical inhomogeneities ... have a relevant (even drastic) meaning for language spreading and competition. We have observed various examples where a language, which in the corresponding homogeneous model would disappear, actually survives.' A survey of research on language extinction concludes that 'the histories of countries where two languages coexist today generally involve split populations that lived without significant interaction, effectively in separate, monolingual societies' (Abrams and Strogatz 2003, 900).

Distance

Our first move is to draw on the simple and powerful idea that homogeneity is inversely related to distance as suggested by Tobler's first law of geography: everything is related to everything else; but near things are more

related than distant things. The implication for language difference is as follows:

H1: The greater the distance between a region and the national capital, the greater the likelihood that the population of that region will speak a distinctive language.

Island-region

Islands are the classic illustration of how insularity can produce endemic species and, by analogy, distinct languages (Pungetti 2012, 51–2). While many factors condition the effect of island insularity on language difference,⁵ we draw the expectation that where a state encompasses a region that is secluded on an island, this is propitious for language difference.

H2: A region that is isolated on an island is more likely to be a language region.

States and empires

The ratio of internal to external interaction has a political as well as a geographical logic. An overarching state or empire facilitates interaction among its population by providing common standards, laws, and collective goods, including defense. A state may also limit interaction between its population and that of neighbouring states by imposing border controls on migration and trade, or by maintaining idiosyncratic laws and institutions.

State building has motivated persistent efforts to impose national standards of communication. In modern states, language 'is an inescapable medium of public discourse, government, administration, law, courts, education, media and public signage. Public life can in principle be a-religious, but it cannot be a-linguistic' (Brubaker 2013, 6). 'Governments ... want to have some say in which language is used as the medium of instruction, or for keeping financial records for tax purposes, or for presenting appeals to overturn lower court decisions', a process that Laitin (1992, 6 & 9) calls language rationalization, the territorial specification of a common language for administration and rule.

In Northeast Asia, linguistic assimilation was fostered by states that endured for more than a millennium. In China, 'we have to see a state nationalism at work ... since very ancient times, operating primarily in a dynastic and elite mode but having the same effect of creating cultural homogeneities' (Reid 2010, 17). From early modern times, European states exerted deliberate, and at times coercive, pressure for linguistic homogeneity within their borders (Gellner 1983; Hobsbawm 1990; Hroch 1985; Liu 2015; Smith 2009). Even today one can see that national languages in Europe have been imposed on a continuum of dialects that still shade into each other.

Former state core

A state, i.e. a polity with explicitly articulated institutions and territory, may leave a legacy of linguistic differentiation long after it has ceased to exist. This is typically the case following dynastic union, when the smaller or weaker part could expect to gain formal recognition of its rights, language, and customs. The benchmark is the *fuero*, a charter granting substantial independence in respect of established legal traditions, negotiated for Basque and Catalan regions after dynastic merger with Castilian Spain. As a result, the potential for linguistic homogenization was severely curtailed even under a nationalizing regime.

However, prior statehood can sustain difference even when merger is coercive. It is one thing to conquer a state in war, but quite another to swallow it into the body politic. Sheer coercion is a blunt instrument for assimilating a people with a history of independence. In such cases it is not unusual for the defeated population to sustain its informal institutions, including its distinctive language (Marks 2012).

H3: A region that comprises the core of a prior state is more likely to be a language region.

Tribal governance and colonial settlement

The logic of interaction theory suggests that a clan or tribe that interacts with its neighbours only sporadically will tend to be linguistically distinctive. Under what circumstances might a tribal region leave a linguistic legacy that can be observed today? Our expectation is that this depends on European colonial settlement. Where European settlement was thin, indigenous languages could endure in bilingualism alongside an imperial language used chiefly by the local elite. Where European settlement was dense, it put a stamp on early institutions and imposed the colonial language in courts, education, administration, and security.⁶ If the local language survived, it was most likely to do so in the geographical or social periphery of the settler society, as Yashar (1999, 84) observes for Latin America.

This intuition is depicted in Figure 3. The likelihood of language difference in a region that has been isolated in tribal governance prior to colonization will depend on whether the local population has been overwhelmed by settlers. Such regions are vulnerable to colonization, but in the absence of mass migration this will not wipe out the indigenous language. The absence of overarching governance is the political equivalent of geographical isolation, and our hunch is that language difference will persist to the extent that the indigenous population is not displaced or overwhelmed. In contrast, the likelihood of language difference in a region already encompassed in overarching governance (i.e. in a pre-colonial empire) is small, and is not affected by European settlement.

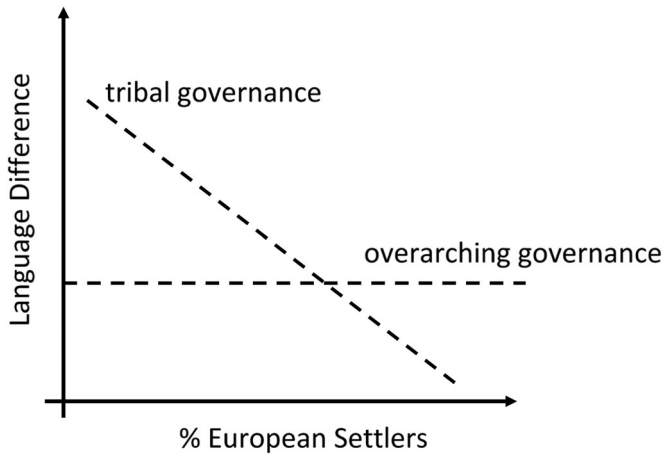


Figure 3. Governance, settlers, and language difference. Source: authors' production.

H4a: Tribal governance prior to colonization increases the likelihood that a region is a language region.

H4b: The greater the presence of European settlers in such a region, the less likely it will sustain language difference.

Language difference and regional authority

There are several reasons to believe that a linguistically distinct region will demand a greater measure of authority than a region in which the population speaks the core language of the state. Language difference is a source of identity (Brubaker 2013; Safran 2008). A population speaking a different vernacular from that of the state's core is more than a collection of individuals; it is a distinctive group sharing a capacity for expressive communication. A minority language binds a group in shared communication and meaning. At the same time, it can demarcate a boundary between who is one of *us* and who is one of *them*. Language difference can indicate the existence of a distinctive way of life. Those who speak a different language may have, in Kymlicka's words (1988, 83), 'a shared vocabulary of tradition and convention' which can reach deep into a community and shape an individual's views on social and political norms (Liu et al. 2018; Singh 2015).

Language

The status of a language is acutely sensitive to whether a language community has authority over education and public discourse. Public opinion surveys in minority language regions, such as Quebec, Catalonia, or Flanders,

show broad public support for subnational control over education, language, culture, and immigration (Keating 2001, 94). Consistent with this, Liu (2011, 126) finds that, of all post-1945 civil wars, more than half have involved language conflict.

Quebec separatism is a case in point. It gained mass support in the 1960s in response to urbanization and migration from the French-speaking rural hinterland to English-speaking cities. Language shaped life chances in palpable ways, fuelling a demand for authority on the part of French-speakers to control Quebec education, immigration, and the public sector. Canadian governments had managed to assuage Catholics in Quebec by allowing the Church a privileged role in communal life, but this was not an option for language.⁷ Minority nationalists demanded the political authority to overcome the incentive to speak English in a predominantly Anglo sphere. Quebec control over the right to make law was conceived as an indispensable tool alter individual cost/benefit calculations about language acquisition (Laitin 1998).

This motivates the following hypothesis:

H5: A language region will tend to have greater authority than a non-language region.

Democracy

The response to pressure on the part of a minority region for self-rule depends on the character of the regime. Democracies are less resistant to decentralizing authority than autocracies (Haggard 2000; Niedzwiecki et al. 2018; Shair-Rosenfield, Marks, and Hooghe 2014; Stepan, Linz, and Yadav 2011). A democracy creates more scope for a region to press its demands, and it provides the state with a more flexible repertoire of accommodation. By contrast, an autocrat may be induced to monopolize authority so that he has the resources to tame or suppress political opposition. A territorially concentrated language minority is a potential threat to an authoritarian regime, and an autocrat may be wary about granting it authority.

Hence, we model regional language difference in interaction with the character of the regime. Figure 4 hypothesizes that the effect of language difference for regional authority depends decisively on the extent to which a regime is democratic.

H6: The more democratic the regime, the greater the positive effect of language difference for regional authority.

Data and operationalization

We wish to understand how the authority of a region is shaped by state formation, colonialism, and physical geography. The effect of these forces is

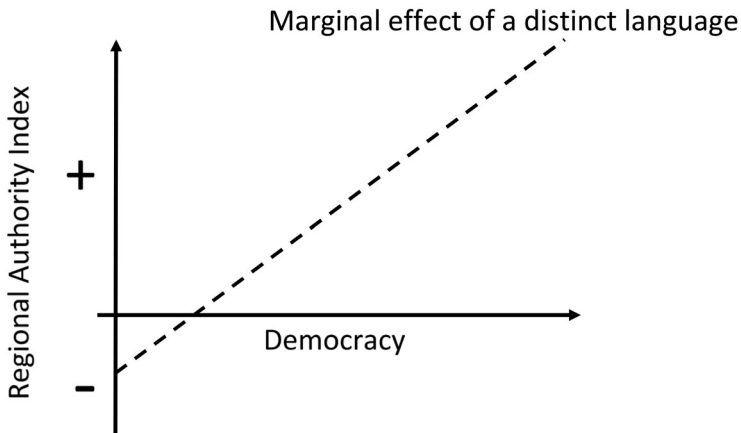


Figure 4. Expected marginal effect of a distinct language. Source: authors' production.

long-term and should be evident in cross-sectional comparison of contemporary regions. Our unit of analysis is the individual region in 2018, defined as the most authoritative general-purpose government between the local and the national level. Conceiving a region as a political unit reflects our concern with the architecture of multilevel governance and has the virtue of producing clearly articulated units for which data on language can be collected.

Taking the individual region as the unit of analysis also makes sense from a measurement perspective. The decision to collect data for regional jurisdictions allows us to match language data with estimates of authority provided by the Regional Authority Index (Hooghe et al. 2016; Hooghe, Marks, and Schakel 2010). The most accurate information on the proportion of language speakers at the sub-national level is available for individual jurisdictions through census data. Whereas *Ethnologue* (Eberhard, Simons, and Fennig 2020) provides maps depicting the geographical spread of language communities, we wish to estimate the percentage of the population speaking a language in a particular jurisdiction.

We impose a population threshold of 150,000, which is calculated as the average population of regions in a tier in 2011 or the prior census.⁸ So while regions *on average* meet the population criterion, the population of some regions in that tier may be smaller.⁹ In all, this delivers 1767 regions in 95 countries across Asia, Europe, and the Americas.

Our theory conceives regional authority as the result of a two-step process in which the distinctiveness of a region is first an outcome and then a predictor. We use original data to put this argument to work.

Regional authority scores in 2018 are provided by the authors' RAI which has annual estimates for all regions in the sample for 1950–2018. Authority is

disaggregated in ten dimensions, five of which tap the authority that a regional government exercises in its territory (self-rule) and five of which tap the authority that a regional government co-exercises in the country as a whole (shared rule) (Hooghe et al. 2016). Most regions have the same authoritative competences as units in their tier, while 151 regions in 53 countries have a separate jurisdictional arrangement that empowers (or disempowers) a region relative to standard regions in its tier. Values on the RAI for the sample population range from 1 for 105 regions that are outposts of the central state to 28 (Republika Srpska in Bosnia–Herzegovina). The sample mean is 11.0 with a standard deviation of 6.95. 296 regions across 19 countries have a RAI score greater than 20. This includes regions in all federal countries except Nepal and Venezuela, alongside the *comunidades* in Spain and regions in Britain, Denmark, Finland, and Papua New Guinea.¹⁰ Figure 5 shows the distribution in the sample.¹¹

A region is coded as a **Language region** if a majority of its population speaks one or more mother tongues that are different from that of the state core.¹² In the face of considerable disagreement about what constitutes a separate language, we code language difference for standard languages according to the ISO-639/2 classification. As the sociolinguist de Swaan (2013, 3) points out, languages ‘resemble clouds: it is hard to tell where one begins and the other ends, and yet most clouds and languages are obviously distinct.’ Dialects

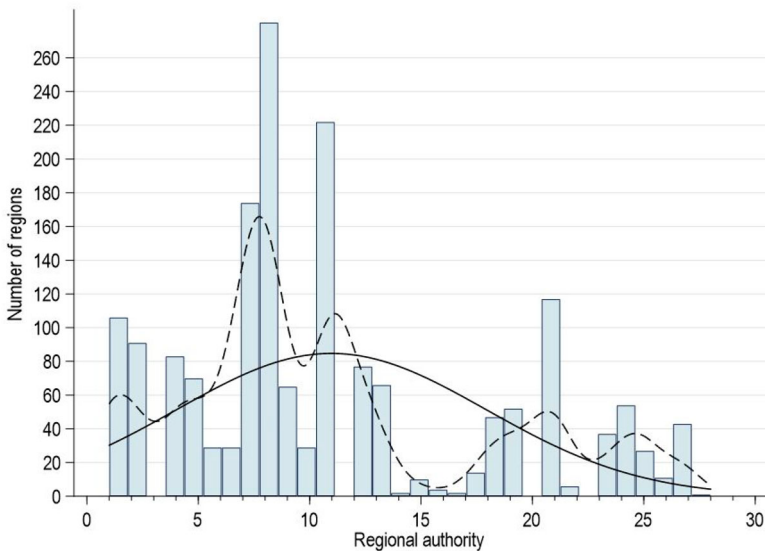


Figure 5. Regional authority in 2018.

Note: $n = 1749$ regions in 2018 (not including 18 states that do not have an intermediate tier of government). The bars display the number of regions with an RAI score, overlaid with the normal density curve (solid line) and the kernel density Gaussian curve (broken line). Source: RAI dataset v.3.

of a standard language (e.g. mutually intelligible variants of Malay) or diglossia involving a codified and vernacular version of a common language (e.g. Mandarin Chinese and its local variants) do not meet this criterion.¹³ We codify what the majority of a region's population indicate as their so-called 'mother tongue,' i.e. the chief language that a person grows up with in childhood (Liu et al. 2018). Where data availability allows, reported language is averaged over the course of the past three decades, and triangulated with information from census data, *Ethnologue*, *Wikipedia*, and secondary sources. Our coding identifies 345 language regions (19.7%).

Former state core takes the value of 1 if the region meets the following criteria: (a) it was part of a prior independent state or empire for a continuous period of at least thirty years since 1200AD; (b) it encompasses the core or capital of the prior state; (c) it does not encompass the capital of the contemporary state; and (d) at least half of its territory was part of the prior state. Of the sample 198 regions or 11.3% meet these criteria.

Tribal governance is a dichotomous variable that takes the value of 1 if the region was not encompassed in a durable state or empire prior to 1600AD or by the time the region was colonized, if colonization took place after 1600. A *durable* state or empire is one that encompassed the region uninterrupted for at least thirty years. Overall, 30 percent of the regions in this study meet the criterion of tribal governance. In former colonies, the proportion increases to 51.5 percent.

The coding for *Former state core* and for *Tribal governance* relies on historical atlases, *Encyclopedia Britannica*, and *Wikipedia*, and is cross-checked with secondary sources (Borcan, Olsson, and Putterman 2018; Wimmer and Min 2006).

Two variables assess geographical insulation. **Distance** is the distance in kilometres 'as-the-crow-flies' between a region's capital and the capital of the country in which it is located. We use a publicly available online search tool to calculate distance,¹⁴ and apply the logarithm (base 10) on the intuition that the marginal effect of a kilometre decreases with distance.

Island-region takes a value of 1 if the region is 30 km or more removed from any other region of its state. We measure the shortest distance between the coastal shore of the region and that of the nearest region of the state. Two neighbouring regions of the same state on the same island do not meet the 30 km distance criterion. Fifty regions (2.8%) are classified as island-regions.

An indigenous culture can be overwhelmed by migration. **Settlers** is the percentage of European settlers in a country's total population in 1900 (Acemoglu, Johnson, and Robinson 2001: Table A.5) complemented with our own estimates.¹⁵ This is 100% for all European countries, including Russia. The percentage allocated to each region is the percentage of settlers in the country that it was part of in 2018. Of the regions in the

dataset, 613 had no settlers and 576 had 100%; in the median country, 20% of the population in 1900 was estimated to consist of settlers.

Democracy is the average score of the Liberal (*v2x_libdem*) and Electoral (*v2x_polyarchy*) dimensions of the Varieties of Democracy Index in 2017 (Coppedge 2020). This country-level variable ranges from 0 to 1.

The models include controls for linguistic fractionalization (Alesina et al. 2003), a region or country's population and area (*statoids* and Wikipedia), and a country's GDP per capita (Feenstra, Inklaar, and Timmer 2015) in 2017.¹⁶ We report cluster-robust standard errors because regions within the same country may be affected in a similar way (Stock and Watson 2008). The online appendix provides more detail on operationalization, descriptive statistics, and robustness.

A model of language difference

We first model the incidence of language regions, i.e. regional jurisdictions where a majority of the population has a mother tongue different from that of the majority in the country. Our prior is that the language distinctiveness of a region depends on the ratio of internal to external interaction, and we investigate the effects of four variables that tap physical and political barriers to inter-regional interaction: island-region, distance from the national capital, whether the region is the core of a former state, and whether tribal governance survived into modernity. We conjecture this last factor to be conditional on pre-1900 colonial settlement, which we model by interacting tribal governance with the density of Euro-colonial inward migration. The statistical estimates for these variables are highly significant and in the expected direction. On the plausible premise that these variables are exogenous to language distinctiveness, the results provide strong confirmation for interaction theory.

Table 1 reports the results from a logit model where we control for linguistic fractionalization, that is the chance that two randomly selected individuals living anywhere in a country have different mother tongues. Controlling for *non-territorial* linguistic fractionalization gives us greater confidence that the independent variables of interest are picking up the effect of territorial barriers at the regional level.

A basic implication of interaction theory is that geographical insulation increases the ratio of internal relative to external interaction. A dichotomous variable, *Island-region*, yields an odds ratio of 5.53. Of the 50 regions that are more than 30 km distant from any other region in the same state, 29 have populations in which a majority has a distinctive mother-tongue. *Distance* yields an odds ratio of 3.26.¹⁷

Our theory posits a political logic alongside a geographical one. States facilitate human interaction across territory by reducing transaction costs

Table 1. The likelihood of a language region.

	Odds ratio	s.e.
Island-region	5.53***	2.29
Distance	3.26***	1.04
Former state core	5.23***	1.36
Tribal governance	7.33***	2.92
Settlers	1.00	0.00
Tribal governance* Settlers	0.95***	0.00
Control (country)		
Linguistic fractionalization	42.82***	22.72
Constant	0.00***	0.00
Number of observations		1767
McFadden R^2		0.37
Tjur's D coefficient		0.40
Wald χ^2 (df)		228.8 (7)
AIC		1121.4
BIC (df)		1165.2 (8)

Note: Logistic regression with cluster-robust standard errors (95 country clusters). Significance: *** $p < .001$ ** $p < .01$ * $p < .05$.

of trade and imposing common laws and institutions. Yet the causality may also run in reverse, as dense interaction among communities may facilitate the formation of overarching jurisdictions, including states. So it is not possible to gauge the causal effect of jurisdictional design on language homogeneity by merely looking at contemporary state structure. Our strategy is to estimate the effect of polity formation prior to 1900.

A region with a durable history of independent statehood is more likely to inherit some of its past institutions and norms, including a distinctive mother-tongue. The odds that a region comprising the core of a former state core is linguistically distinctive are 5.23 times higher than for a region with no such history. Regions in India and Spain illustrate the effect. In India, fifteen of the seventeen states that formerly were state-cores are also language regions. In Spain, the same is true of three of the five comunidades.

A region in which tribal governance survived until colonization is more likely to have a population that speaks one or more minority languages to this day.¹⁸ However, tribal areas are also vulnerable to colonial incursion, and so language difference is conditional on the extent of colonial settlement which can overwhelm the local population. Figure 6 provides strong confirmation of the effect of tribal governance conditional on colonial settlement. In the absence of extensive colonial settlement, the probability that a region with a tribal history until colonial conquest will be a language region in the early twenty-first century is 49.1% ($\pm 16\%$). If European settlers make up 30% of the population the probability declines drastically to 15.8% ($\pm 6.2\%$); and if the proportion of such settlers is 90%, the probability that the population is linguistically distinctive is just 0.7%.¹⁹ As expected, the distinctiveness of a region that has been part of an overarching

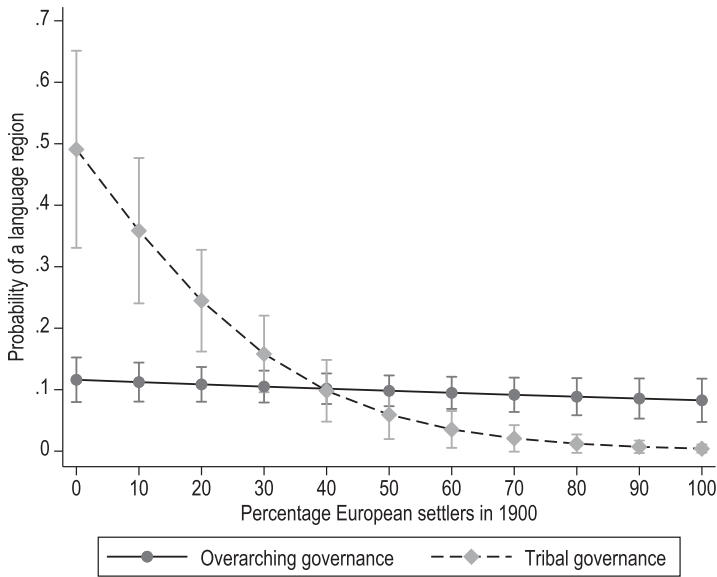


Figure 6. The effect of tribal governance for language difference conditional on European settlers.

Note: $n = 1767$ regions in 2018; 90% confidence intervals.

polity is much less sensitive to European settlement: the 90% confidence bars overlap across the entire zero to 100 percent range for *Settlers* in Figure 6.

The effect of language difference for regional authority

We begin by evaluating whether language regions have more, or less, authority than non-language regions. Table 2 assesses our expectations for 1698 regions that form part of regional tiers in 74 countries.²⁰ The first thing to note is that *Language region* on its own has no significant effect in Model 1, and Model 2 shows why: the effect of *Language region* is conditioned on *Democracy*.²¹ Whether language regions receive more authority depends on the openness of the regime.

In Model 2, a history of prior statehood in the region is a strong predictor of regional authority, with an estimated coefficient of +5.6 (± 2.3) holding other variables at their means. Regional authority also has a functional logic. All else equal, more populous regions and regions with a larger territorial footprint tend to have greater authority (Marks, Hooghe, and Schakel 2008). And national affluence is conducive to regional authority as wealthier countries tend to provide a wider range of public goods such as health, welfare, and education. Notably, these factors are significant while controlling for non-territorial linguistic fractionalization at the country level.

Table 2. Language regions and regional authority.

	Model 1		Model 2	
	Coeff.	s.e.	Coeff.	s.e.
Language region	-0.21	0.65	-4.03*	1.54
Democracy	6.59	4.99	5.08	5.18
Language region * Democracy			9.02**	3.37
Controls (region)				
Former state core	5.72***	1.43	5.56***	1.40
Population	2.38**	0.65	2.42***	0.62
Area	2.13*	0.84	2.16*	0.81
Controls (country)				
GDP per capita	4.75	2.80	5.09°	2.77
Linguistic fractionalization	1.83	2.18	1.56	2.23
Constant	-35.93**	11.00	-36.85**	10.77
Observations		1698		1698
R ²		0.46		0.47
F (df)		15.9 (7)		15.0 (8)

Note: OLS regression with cluster-robust standard errors (74 countries). Significance: *** $p < .001$ ** $p < .01$ * $p < .05$ ° $p < .10$.

There are two plausible mechanisms at play in the effect of language regions for regional authority conditioned on democracy. On the one hand, the findings reported in Model 2 might indicate that democratic governments empower minority language regions relative to core-language regions, and that authoritarian governments do the opposite. However, the same result could indicate that language regions under democracy had a positive knock-on effect for the authority of *all* regions in the country while language regions under an authoritarian regime had a negative knock-on effect.

We assess the first mechanism by modelling the raw *difference* in a region's authority index compared to that of standard regions in the same tier. [Figure 7](#) visualizes the marginal effect of a language region conditioned on the level of democracy with all other variables held at their means. The estimate reaches statistical significance ($p = 0.010$), but it is substantively small. A shift from 0.2 to 0.8 on the democracy scale (i.e. from authoritarian to strongly democratic) produces a slight increase in the authority of a language region relative to a standard region of less than one point on the 30-point RAI scale (0.92 points ± 0.39).

To assess the systemic effect of language regions we model regional authority *at the country level*. [Figure 8](#) plots the effect of the proportion of language regions in a country for the entire regional tier. The figure visualizes this at low and high democracy, 0.2 and 0.8 on the VDem scale. The conditional effect of regime type is statistically significant. Where one region in ten is a language region, the RAI for the regional tier in an authoritarian regime is 8.1 (± 2.4) and in a democratic regime it is 13.0 (± 1.5). Where one third of the regions in a country are language regions, the RAI falls to 5.2

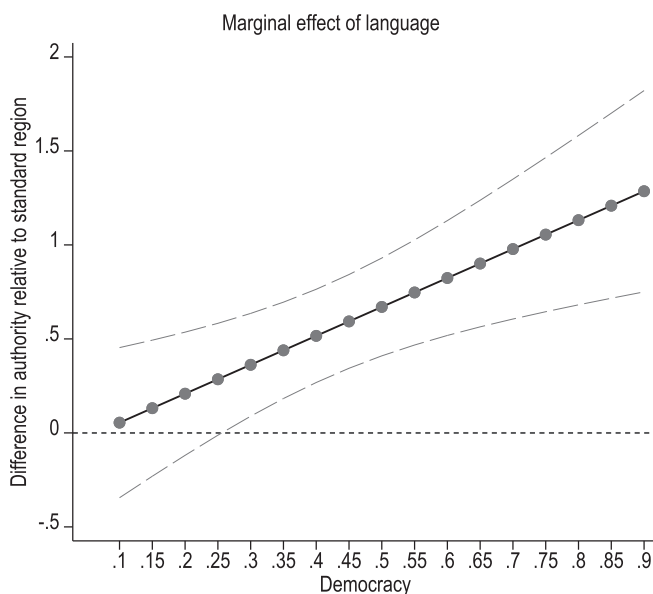


Figure 7. The effect of language difference for a region’s authority conditional on democracy.

Note: 90% confidence intervals. The dependent variable is calculated by subtracting, for each region, the RAI for a standard region from that region’s RAI.

(± 2.5) in an authoritarian regime, and it rises to 15.0 (± 1.9) under democracy.²²

Hence, to the extent that language regions affect regional authority, this is chiefly a systemic effect that is heavily dependent on the character of the regime. Spain, in which seven of nineteen comunidades are language regions, illustrates how this can happen. Democratization in the late 1970s produced a cascade of regional bargains, beginning with the historic regions of the Basque Country, Catalonia, and Galicia (Agranoff and Gallarín 1997). The central government sought to tame these demands by encasing them in a national quasi-federal system described as *café para todos*, ‘coffee for everyone instead of champagne for the historic regions’ (Agranoff 2005, 7; Chapman Osterkatz 2013; Keating 1998; Moreno 2001, 61).

In Belgium, the central government sought to contain Flemish nationalism by giving regions a role in decision making in the country as a whole involving a national chamber in which regions are represented and intergovernmental arrangements in which regional governments bargain directly with the central government. The strategy culminated in a leap to federalism in 1995 (Hooghe 2004; Swenden 2016).

Authoritarian regimes tend to produce the opposite result. President Putin sought to stifle regional autonomy in the Russian peripheries by centralizing

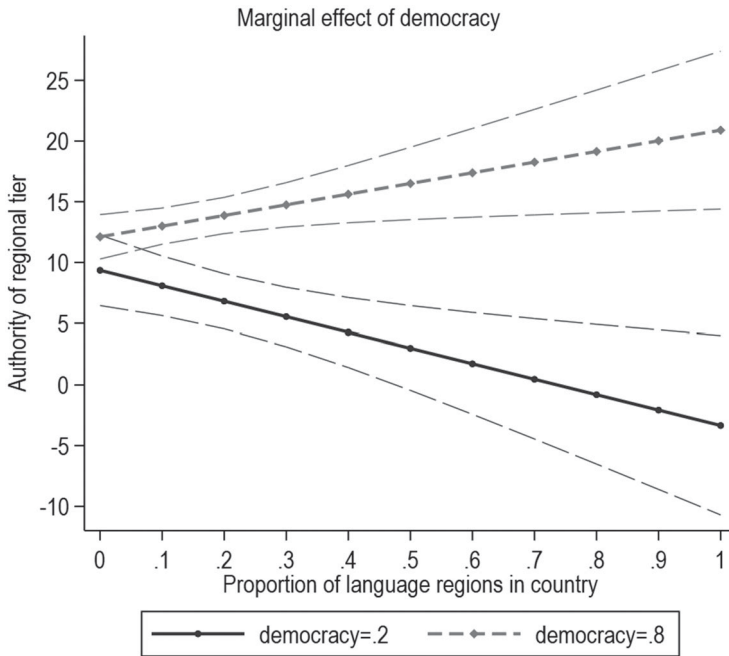


Figure 8. The effect of language difference for regional authority in a country conditional on democracy.

Note: $n = 74$ countries in 2018; 90% confidence intervals. The dependent variable is the mean RAI score of all regions in a country's most authoritative tier.

authority in the country as a whole and by coopting ethnic elites. In 2017 the distinctive regional pact of the last republic, Tatarstan, expired (Goloso and Konstantinova 2016; Moreno and Obydenkova 2013; Obydenkova and Swenden 2013; Zuber 2011). Standardization, centralization, and elite cooptation appears also to have been China's preferred strategy for managing Tibet, Xinjiang, and other regions with language minorities (Sheng 2009).

Conclusion

This article seeks to explain the incidence and authority of regional jurisdictions in which the majority of the population speaks a minority language in their respective state. We do so by drawing on literature in geography, linguistics, and political science. To evaluate our claims, we extend the Regional Authority Index to encompass 1767 regions in 95 countries up to 2018 and introduce a new dataset on language, prior statehood, and geography at the regional level. We propose a theory in two steps. First, on the premise that language difference reflects the ratio of internal interaction to external interaction, we find that the incidence of language regions depends both

on geography and political institutions. Sheer distance and travel time between the capital of a state and a region enhances the likelihood of observing a language region, as does the location of a region on an island. This confirms that language difference, like speciation, is sensitive to physical barriers to interaction.

We extend this line of reasoning to political institutions on the principle that a state increases cultural homogeneity within its borders while differentiating its population from that beyond. A region that was once the core of a state may leave a legacy of difference long after it has lost independence. By the same logic, a region with a history of tribal insulation is likely to retain its distinctive language, but only if it is spared mass inward migration.

Our second step is to model the effect of language difference on regional authority. We confirm the intuition that language regions tend to seek special rights and that this has a knock-on effect for all regions in a country. However, we also show that this depends decisively on the character of the regime. The more democratic the regime, the greater the authority-enhancing consequence of language difference. In contrast, authoritarian regimes rarely give special rights to language regions and the presence of language regions tends to suppress regional authority in the country.

Theories of language difference have generally taken Europe as their point of departure. This is understandable since the idea that a state should have a single language originated in Europe and was then disseminated through colonialism and geopolitical hegemony. In this article, we have taken a step in theorizing language and regional authority in a wide variety of contexts. However, there are several ways in which our analysis can be extended. One might go beyond colonial migration to investigate resettlement, war, and slavery, each of which has shaped the fate of indigenous languages. As one moves beyond Europe, it becomes all the more important to investigate the diverse ways in which states can seek to impose nations by promoting a state religion or state ideology, ethnic commonality, civic patriotism, or anti-imperial ideology, alongside language (Brubaker 2015; Reid 2010; Safran 2008). We find that language regions not only have an impact on their own RAI but their main effect is on the RAI of a regional tier within a country. The character of the regime is key, and here more research would be valuable in uncovering the disempowering strategies of political leaders in non-democracies.

While our analysis has focused on the deep structural sources of language difference, there are many insights to be gained from strategic theories of language use at the group or individual level (de Swaan 2013; Laitin 2007). This is an informational as well as theoretical challenge which takes us back to the opening sentences of this study. Comparative regional-level data on public opinion, elites, social movements, and political parties are now available for much of the developed world (e.g. Schakel and Romanova

2018; Tatham and Bauer 2016; Zuber and Szöcsik 2019). Commensurate data would be immensely valuable for a wider range of regions. The extent to which language is politically consequential for the structure of governance is a deep and fascinating puzzle. As new sources of comparative regional data become available, theory and empirics can speak more directly to each other.

Notes

1. Parallel to our approach here, Cederman, Rød, and Weidmann (2007), Wucherpfennig et al. (2011), and Wimmer (2015) have produced geo-data on the spatial diffusion of ethnic groups within and across states.
2. Oft-cited examples of ethnically distinct groups with extensive inter-ethnic interaction are Dalits in contemporary India and enslaved people in antebellum United States.
3. Perhaps the most influential variant is Deutsch's (1942, 1954) cybernetic theory, which explains language assimilation as a function of a group's internal relative to external communication. Deutsch (1954, 41) conceives communication in broad terms: 'Ties of transportation, economic intercourse, social stratification, cultural similarity, and similarity in already existing speech habits, as well as relative barriers and discontinuities in all these respects will all have their effects in determining what the actual speech community will be at any one time'.
4. 'The formation of different languages and of distinct species, and the proofs that both have been developed through a gradual process, are curiously parallel' (Darwin 1882, 90; see also Bromham 2017).
5. These include factors that facilitate or impede trade, migration, and geo-political interference; absolute and relative technological sophistication in communication, organization, and warfare; political or social factors that constrain or increase the supply of mainland migrants; the technological, organizational, and cultural capacity of the island population to resist assimilation; and international policy on the protection of minorities.
6. This argument parallels Acemoglu, Johnson, and Robinson (2001) who find that dense European settlement lay the institutional foundation for long-term economic development.
7. The Catholic Church's chief concern was usually to stem the dissemination of secular curricula rather than the expansion of education in the people's language (de Swaan 1988). Local priests did not always obey their hierarchy's orders, and several became prominent minority nationalists. For a deep study of the changing nature of Quebec nationalism in modernizing Canada, see Guindon (1988).
8. We lift the population criterion for a regional government that operates outside a regional tier. Examples include the Åland Islands (Finland), Tobago (Trinidad and Tobago), Pattaya (Thailand), as well as indigenous governance in six countries.
9. This includes 18 states that do not have an intermediate tier of government that meets these criteria. The results are robust when these states are excluded from the analysis.
10. Nepal's 2015 federal constitution is incompletely implemented. In Venezuela, authoritarian governments have recentralized state powers while retaining the pretense of federalism.
11. An average score of 11 is equivalent to the authority of New Zealand's regional governments, which are decentralized but remain subject to the central

government's veto, have a policy portfolio focused on economic development, transport, and environment, can set the base and rate of property taxes and borrow on domestic markets. New Zealand's regional assemblies are elected, and the assembly appoints the regional executive. Myanmar's fourteen regions and states also have a RAI score of 11, with less self-rule but greater shared rule by virtue of an upper chamber with twelve directly elected regional and state senators who form a majority and can veto national constitutional reform.

12. A more stringent criterion defines a language region as a region where a majority speaks the same distinctive non-core language. Results are robust using this specification.
13. This is consistent with de Swaan's notion that distinct languages tend to be mutually unintelligible while dialects tend to be mutually intelligible (de Swaan 2013, 12). What is mutually intelligible can be contested. While most Flemish speakers consider Flemish a variant of Dutch, some conceive it as a separate language. Similarly, while Chinese authorities emphasize the shared written form for Mandarin and Cantonese variants, many language specialists consider these separate languages because the spoken form is mutually unintelligible.
14. <https://www.freemaptools.com/how-far-is-it-between.htm>. Where a region has two capitals we estimate the average distance. The Indian Act Bands and Self-governing Aboriginal People in Canada; Indian Tribes in the United States; *Territorios autonomos indigenas* in Bolivia; *Resguardos indigenas* in Colombia; *Territorios indigenas* in Costa Rica are dispersed across the country, and we allocate the distance of the median region in their country.
15. The Acemoglu et al. dataset lacks estimates for Albania, Brunei, Cambodia, Cuba, Cyprus, Ireland, Kosovo, Malta, Montenegro, Slovenia, Taiwan, and Timor-Leste.
16. We use the logarithm (base 10) for regional or country population, area, and GDP per capita.
17. Estimating distance as a function of travel time by car in 2020 yields similar results.
18. An alternative line of inquiry focuses on the distinctive institutional legacies of the colonial empire in question. Robustness checks using dummies for British, French, Dutch, Portuguese, and Spanish colonies detect no significant impact on territorial language difference.
19. Results are robust when dropping particular countries, controlling for world region, and when restricting the sample to regions in former colonies. A model that employs a more restrictive operationalization of language region produces similar results.
20. We exclude regional jurisdictions that are not part of a country-wide system of regions (e.g. Indian Tribes in the United States) and regional jurisdictions that are themselves states (e.g. Barbados).
21. A model that substitutes regional self-rule for regional authority produces similar results.
22. The effect is similar with a population-weighted measure of regional authority and the incidence of language regions.

Acknowledgements

We would like to thank Weining Ai, Kyle Chan, Aleksandra Danailova, Noah Dasanaïke, Fanying Kong, Iana Gein, Sven Hegewald, Caroline Lancaster, Luigi Mendez, Juan

Diego Prieto, Alan James Simmons, and Emily Venturi for research assistance and Michaël Tatham and two anonymous reviewers for their comments. We are indebted to Brian C.H. Fong (The Education University of Hong Kong) for collaboration in estimating Hong Kong.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

Arjan Schakel received support from the European Commission [grant number 2016.CE.16.BAT.079]; the Trond Mohn Foundation; and the University of Bergen [grant number 812468]. Gary Marks and Liesbet Hooghe received support from the Reynolds Foundation; and the Robert Schuman Centre, EUI, Florence.

References

- Abrams, Daniel M., and Steven H. Strogatz. 2003. "Modelling the Dynamics of Language Death." *Nature* 424: 900.
- Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review* 91 (5): 1369–1401.
- Agranoff, Robert. 2005. "Federal Asymmetry and Intergovernmental Relations in Spain." Asymmetry Series. Kingston: Queens University.
- Agranoff, Robert, and Juan A.R. Gallarín. 1997. "Toward Federal Democracy in Spain: an Examination of Intergovernmental Relations." *Publius: The Journal of Federalism* 27 (4): 1–38.
- Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg. 2003. "Fractionalization." *Journal of Economic Growth* 8: 155–194.
- Barth, Fredrik. 1969. "Introduction." In *Ethnic Groups and Boundaries: The Social Organization of Cultural Difference*, edited by Fredrik Barth, 1–9. Long Grove, IL: Waveland Press.
- Borcan, Oana, Ola Olsson, and Louis Putterman. 2018. "State History and Economic Development: Evidence from Six Millennia." *Journal of Economic Growth* 23: 1–40.
- Bromham, Lindell. 2017. "Curiously the Same: Swapping Tools between Linguistics and Evolutionary Biology." *Biology & Philosophy* 32: 855–886.
- Brubaker, Rogers. 2013. "Language, Religion and the Politics of Difference." *Nations and Nationalism* 19 (1): 1–20.
- Brubaker, Rogers. 2015. *Grounds for Difference*. Cambridge, MA: Harvard University Press.
- Cederman, Lars-Erik, Jan Ketil Rød, and Nils Weidmann. 2007. "Geo-Referencing of Ethnic Groups: Creating a New Dataset." Proceedings from the Annual Convention of the International Studies Association, Chicago.
- Chapman Osterkatz, Sandra. 2013. "Commitment, Capacity, and Community: the Politics of Multilevel Health Reform in Spain and Brazil." Doctoral dissertation, University of North Carolina at Chapel Hill.

- Coppedge, Michael, et al. 2020. "V-Dem [Country–Year/Country–Date] Dataset v10." Varieties of Democracy (V-Dem) Project. doi:10.23696/vdemds20.
- Darwin, Charles R. 1882. *The Descent of Man, and Selection in Relation to Sex*. 2nd ed. London: John Murray.
- de Swaan, Abram. 1988. *In Care of the State: Health Care, Education and Welfare in Europe and the USA in the Modern Era*. Oxford: Oxford University Press.
- de Swaan, Abram. 2013. *Words of the World: The Global Language System*. Cambridge, UK: Polity.
- Deutsch, Karl W. 1942. "The Tread of European Nationalism—The Language Aspect." *American Political Science Review* 36 (3): 533–541.
- Deutsch, Karl W. 1954. *Nationalism and Social Communication*. Cambridge: MIT Press.
- Eberhard, David M., Gary F. Simons, and Charles D. Fennig, eds. 2020. *Ethnologue: Languages of the World*. 23rd ed. Dallas, TX: SIL International. <http://www.ethnologue.com>.
- Feenstra, Robert C., Robert Inklaar, and Marcel P. Timmer. 2015. "The Next Generation of the Penn World Table." *American Economic Review* 105 (10): 3150–3182.
- Gellner, Ernest. 1983. *Nations and Nationalism*. Ithaca, NY: Cornell University Press.
- Giraudy, Agustina, Eduardo Moncada, and Richard Snyder, eds. 2019. *Inside Countries: Subnational Research in Comparative Politics*. Cambridge, MA: Cambridge University Press.
- Golosov, Grigorii V., and Marian Konstantinova. 2016. "Gubernatorial Powers in Russia." *Problems of Post-Communism* 63 (4): 241–252.
- Guindon, Hubert. 1988. *Quebec Society: Tradition, Modernity and Nationhood*. Toronto/Montreal: University of Toronto Press.
- Haggard, Stephan. 2000. "The Politics of Decentralization in Latin America." In *Annual World Bank Conference on Development in Latin America and the Caribbean 1999: Decentralization and Accountability of the Public Sector*, edited by Shahid Javed Burki, and Guillermo E. Perry, 33–44. Washington, DC: World Bank Latin American and Caribbean Studies.
- Harbers, Imke, Jos Bartman, and Enrike van Wingerden. 2019. "Conceptualizing and Measuring Subnational Democracy across Indian States." *Democratization* 26 (7): 1154–1175.
- Hobsbawm, Eric J. 1990. *Nations and Nationalism since 1780*. Cambridge: Cambridge University Press.
- Hooghe, Liesbet. 2004. "Belgium: Hollowing the Center." In *Federalism and Territorial Cleavages*, edited by Ugo Amoretti and Nancy Bermeo, 55–92. Baltimore, MD: Johns Hopkins Press.
- Hooghe, Liesbet, and Gary Marks. 2016. *Community, Scale, and Regional Governance: A Postfunctionalist Theory of Governance, Vol. II*. Oxford: Oxford University Press.
- Hooghe, Liesbet, Gary Marks, and Arjan H. Schakel. 2010. *The Rise of Regional Authority: A Comparative Study of 42 Democracies (1950–2006)*. London: Routledge.
- Hooghe, Liesbet, Gary Marks, Arjan H. Schakel, Sara Niedzwiecki, Sandra Chapman Osterkat, and Sarah Shair-Rosenfield. 2016. *Measuring Regional Authority: A Postfunctionalist Theory of Governance, Vol. I*. Oxford: Oxford University Press.
- Hroch, Miroslav. 1985. *Social Preconditions of National Revival in Europe: A Comparative Analysis of the Social Composition of Patriotic Groups among the Smaller European Nations*. Translated by Ben Fowkes. Cambridge: Cambridge University Press.
- Jeffery, Charlie, and Arjan H. Schakel. 2013. "Insights: Methods and Data Beyond Methodological Nationalism." *Regional Studies* 47 (3): 402–404.

- Keating, Michael. 1998. *The New Regionalism in Western Europe: Territorial Restructuring and Political Change*. Aldershot: Edward Elgar.
- Keating, Michael. 2001. *Plurinational Democracy: Stateless Nations in a Post-Sovereignty Era*. Oxford: Oxford University Press.
- Kymlicka, Will. 1988. *Multicultural Citizenship*. Oxford: Oxford University Press.
- Laitin, David D. 1992. *Language Repertoires and State Construction in Africa*. Cambridge, MA: Cambridge University Press.
- Laitin, David D. 1998. *Identity in Formation: The Russian-Speaking Populations in the Near Abroad*. Ithaca, NY: Cornell University Press.
- Laitin, David D. 2000. "What Is a Language Community?" *American Journal of Political Science* 44 (1): 142–155.
- Laitin, David D. 2007. *Nations, States, and Violence*. Oxford: Oxford University Press.
- Liu, Amy. 2011. "Linguistic Effects of Political Institutions." *Journal of Politics* 73 (1): 125–139.
- Liu, Amy. 2015. *Standardizing Diversity: The Political Economy of Language Regimes*. Philadelphia, PA: University of Pennsylvania Press.
- Liu, Amy H., Sarah Shair-Rosenfield, Lindsay Vance, and Zombor Csata. 2018. "Linguistic Origins of Gender Equality and Women's Rights." *Gender & Society* 32 (1): 82–108.
- Marks, Gary. 2012. "Europe and Its Empires: From Rome to the European Union." *JCMS: Journal of Common Market Studies* 50 (1): 1–20.
- Marks, Gary, Liesbet Hooghe, and Arjan H. Schakel. 2008. "Patterns of Regional Authority." *Regional and Federal Studies* 18 (2-3): 167–181.
- Moreno, Luis. 2001. *The Federalization of Spain*. London: Frank Cass.
- Moreno, Luis, and Anastassia Obydenkova. 2013. "Federalization in Russia and Spain: The Puzzle of Reversible and Irreversible Outcomes." *Regional and Federal Studies* 23 (2): 151–168.
- Niedzwiecki, Sara, Sandra Chapman-Osterkat, Gary Marks, and Liesbet Hooghe. 2018. "The RAI Travels to Latin America: Measuring Regional Authority Under Regime Change." *Regional and Federal Studies*. doi:10.1080/13597566.2018.1489248.
- Nosil, Patrik, Luke J. Harmon, and Ole Seehausen. 2009. "Ecological Explanations for (Incomplete) Speciation." *Trends in Ecology & Evolution* 24 (3): 145–156.
- Obydenkova, Anastassia, and Wilfried Swenden. 2013. "Autocracy-Sustaining Versus Democratic Federalism: Explaining the Divergent Trajectories of Territorial Politics in Russia and Western Europe." *Territory, Politics, Governance* 1 (1): 86–112.
- Patriarca, Marco, and Els Heinsalu. 2009. "Influence of Geography on Language Competition." *Physica A: Statistical Mechanics and Its Applications* 388 (s2–3): 174–186.
- Prochazka, Katharina, and Gero Vogla. 2017. "Quantifying the Driving Factors for Language Shift in a Bilingual Region." *Proceedings of the National Academy of Sciences* 114 (17): 4365–4369.
- Pungetti, Gloria. 2012. "Islands, Culture, Landscape and Seascape." *Journal of Marine and Island Cultures* 1: 51–54.
- Reid, Anthony. 2010. *Imperial Alchemy: Nationalism and Political Identity in Southeast Asia*. Cambridge, MA: Cambridge University Press.
- Safran, William. 2008. "Language, Ethnicity and Religion: A Complex and Persistent Linkage." *Nations and Nationalism* 14 (1): 171–190.
- Safran, Rebecca J., and Patrick Nosil. 2012. "Speciation: The Origin of New Species." *Nature Education Knowledge* 3 (10): 17.
- Schakel, Arjan H., and Valentina Romanova. 2018. "Towards a Scholarship on Regional Elections." *Regional and Federal Studies* 28 (3): 233–252.

- Shair-Rosenfield, Sarah, Gary Marks, and Liesbet Hooghe. 2014. "A Comparative Measure of Decentralization for Southeast Asia." *Journal of East Asian Studies* 14 (1): 85–108.
- Sheng, Yumin. 2009. "Authoritarian Co-optation, the Territorial Dimension: Provincial Political Representation in Post-Mao China." *Studies in Comparative International Development* 44: 71–93.
- Singh, Perna. 2015. "Subnationalism and Social Development: A Comparative Analysis of Indian States." *World Politics* 67 (3): 506–562.
- Smith, Anthony D. 2009. *Ethnosymbolism and Nationalism: A Cultural Approach*. London: Routledge.
- Sobel, James M., Grace F. Chen, Lorna R. Watt, and Douglas W. Schemske. 2010. "The Biology of Speciation." *Evolution* 64 (2): 295–315.
- Solé, Richard V., Bernat Corominas-Murtra, and Jordi Fortuny. 2010. "Diversity, Competition, Extinction: The Ecophysics of Language Change." *Journal of The Royal Society Interface* 7: 1647–1664.
- Stepan, Alfred C., Juan J. Linz, and Yogendra Yadav. 2011. *Crafting State-Nations: India and Other Multinational Democracies*. Baltimore, MD: Johns Hopkins University Press.
- Stock, James H., and Mark W. Watson. 2008. "Heteroskedasticity-Robust Standard Errors for Fixed Effects Panel Data Regression." *Econometrica* 76 (1): 155–174.
- Swenden, Wilfried. 2016. "Why Is Belgian Federalism Not More Asymmetrical?" In *Federalism beyond Federations: Asymmetry and Processes of Resymmetrisation*, edited by Ferran Requejo, and Klaus-Jürgen Nagel, 13–36. London: Routledge.
- Tatham, Michaël, and Michael W. Bauer. 2016. "Competence Ring-Fencing from Below? The Drivers of Regional Demands for Control over Upwards Dispersion." *Journal of European Public Policy* 21 (9): 1367–1385.
- Wimmer, Andreas. 2015. "Race-Centrism. A Critique and a Research Agenda." *Ethnic and Racial Studies* 38 (13): 2186–2205.
- Wimmer, Andreas, and Brian Min. 2006. "From Empire to Nation-State. Explaining Wars in the Modern World, 1816–2001." *American Sociological Review* 71 (6): 867–897.
- Wucherpfennig, Julian, Nils B. Weidmann, Luc Girardin, Lars-Erik Cederman, and Andreas Wimmer. 2011. "Politically Relevant Ethnic Groups across Space and Time: Introducing the GeoEPR Dataset." *Conflict Management and Peace Science* 28 (5): 423–437.
- Yashar, Deborah. 1999. "Democracy, Indigenous Movements, and the Postliberal Challenge in Latin America." *World Politics* 52 (1): 76–104.
- Zuber, Christina Isabel. 2011. "Understanding the Multinational Game: Toward a Theory of Asymmetrical Federalism." *Comparative Political Studies* 44 (5): 546–571.
- Zuber, Christina Isabel, and Edina Szöcsik. 2019. "The Second Edition of the EPAC Expert Survey on Ethnationalism in Party Competition – Testing for Validity and Reliability." *Regional and Federal Studies* 29 (1): 91–113.