French*

Guillaume Blanc[1] and Masahiro Kubo[2]

[1]The University of Manchester
[2]Brown University

January 14, 2024
Link to the latest version

Abstract

This paper studies nation-building in a fragmented society. We document the adoption of a common language and the construction of a national identity in France. Using a natural experiment and drawing on a novel dataset on the languages spoken across municipalities on the eve of the twentieth century, we establish that state intervention in the provision of education brought homogenization. To understand why nation-building was successful, we study heterogeneity and find that elites and the demand for education were instrumental in driving assimilation. Finally, we document further impacts on identity and ideology in the twentieth century.

***JEL codes**: P00, H52, I20, N40, Z13*
***Keywords**: nation-building, language, education*

Article 2. *La langue de la République est le français.*
— French Constitution, October 4, 1958

1 Introduction

The transition from fragmented states to cohesive nations, with the construction of imagined communities and common identities, represents a critical juncture in human history. During the nineteenth century, policies of homogenization and assimilation dramatically advanced social cohesion and political stability, with the emergence of modern nation-states, at the expense of diversity (Anderson, 1983; Gellner, 1983; Hobsbawm, 1990; Tilly, 1975).

We study the making of France, the canonical example of nation-building in the literature since the publication of *Peasants into Frenchmen* (Weber, 1976). At the time of the French Revolution, the French language, which was one of more than forty-six dialects and languages

* Blanc: The University of Manchester, Arthur Lewis Building, Oxford Road, Manchester M13 9PL, UK, **guillaume.blanc@manchester.ac.uk**; Kubo: Department of Economics, Robinson Hall, 64 Waterman Street, Providence, RI 02912, USA, **masahiro_kubo@brown.edu**.We would like to warmly thank O. Galor, S. Michalopoulos, L. Putterman, R. Wacziarg, and D. Weil for their kind support and dedicated supervision. This research has benefited from insightful comments and suggestions from A. Bergeron, N. Blanc, J. Bruhn, K. Desmet, Y. Dupraz, E. Durante, J. Goldsmith, V. Fouka, P. Giuliano, L. Heldring, P. Hull, F. Kersting, M. Kumagai, A. Montalbo, E. Murard, S. Ottinger, T. Piketty, G. Ponzetto, M. Senouci, C. Tricaud, X. Wang, N. Wolf, and G. Xu. We also thank seminar and conference participants at ASSA, Aix-Marseille, ASREC, Brown, Collegio Carlo Alberto, Copenhagen, CREI, Cultural Evolution Society, Econometric Society, EIEF, ENS, King's College London, Humboldt, Manchester, NES, NEUDC, PSE, Stellenbosch, UC3M, UCLA Anderson, and York for helpful comments and discussions. Noémie Feuillette and Clara Lefevre provided excellent research assistance. Funding from the Orlando Bravo Center for Economic Research at Brown University and the Institute for Humane Studies is gratefully acknowledged. All remaining errors are our own.

historically spoken in France, was foreign to the vast majority of the population. Today, French is the common language. In this research, we empirically document nation-building and the process of homogenization. Using a natural experiment and drawing on a novel, detailed historical dataset, we establish, for the first time, the causal effect of mass state-sponsored education on the adoption of French and on the construction of a national identity and political ideology, with increased preferences for centralization.

We document homogenization at a particularly granular level, relying on a detailed survey of the languages spoken in municipalities in France around 1900, the *Atlas Linguistique de la France*. The data records the common language and pronunciation for 1,920 unique words or expressions in 577 municipalities at the time. We digitize a representative sample of words from the *Atlas* selected from a standard collection of words with universal meanings, the Swadesh list. Our main outcome variable, linguistic distance from French, the dialect spoken in the region near Paris, is measured using a standard measure of distance across these words.

Using a regression discontinuity design exploiting quasi-experimental variation in the building of schools, we establish that the provision of state-sponsored education brought about the adoption of the French language.

The *Loi du 28 Juin 1833 sur l'Instruction Primaire*, or *Loi Guizot*, named after the Minister of Public Instruction François Guizot, laid the foundations of mass public primary schooling in France. In municipalities above five hundred inhabitants, the law mandated the building of schools for boys as well as major changes in the curriculum and in the training of schoolteachers. These changes encompassed the teaching of the national language and a national history, the creation and distribution of state-approved textbooks, as well as the establishment of both a national system of teacher-training colleges and a corps of nationally recruited school inspectors. A close advisor of Guizot argued at the time that "each school shall be a colony of French language in a conquered land" (Lorain, 1837, p. 29), while Guizot (1833) himself wrote that "through the teaching of the French language, primary schools will instill and spread the spirit and unity of nationality everywhere" (p. 102).

On the eve of the twentieth century, the language spoken in municipalities above the threshold, that had received state-sponsored education, was significantly closer to French than in those that had not. We also demonstrate that, under a particularly weak assumption, our empirical design allows us to estimate a lower bound on how fast homogenization took place at the threshold after the *Loi Guizot*, although we only observe language in a single cross-section. We argue that state-sponsored education contributed to a total decrease in linguistic distance from French of at least 20 percent during that period, a particularly rapid homogenization.

Our results are robust across a wide range of standard tests and alternative or placebo specifications. We first provide evidence to support our empirical design, by testing for manipulation and showing that observables are balanced at the threshold. Then, we show the robustness of the estimated coefficients to using alternative methods of estimation, kernels, polynomial order, and bias correction; to using alternative bandwidths; to using different distance metrics and measuring distance across the entire corpus of words in the *Atlas*; to accounting for individual-level characteristics of survey respondents; and to accounting for

bilingualism near the border. We also vary the threshold for schools and show that only the five-hundred-inhabitant cutoff, based on population at the time of the law, returns statistically significant estimates.

Additionally, we discuss the interpretation of our results and explore why nation-building was successful. We find that state-sponsored education spread the language spoken near Paris and in the Loire Valley, the language of the elites. Our research also reveals that homogenization was significantly influenced by elites and returns to education, suggesting that successful nation-building could only occur during the transition from stagnation to growth, as societies demanded education.

Finally, we further document persistent impacts on national identity and ideology, suggesting that nation-building dramatically reshaped society. Using data on the birth places of Resistance heroes and Nazi collaborators during World War II, as well as on votes against the regionalization of political authority in the 1969 French constitutional referendum, we find that the nation-building policy increased the salience of national identity and preferences for centralization.

Our research adds to a large literature on nations and nationalism, in fields spanning economics, history, philosophy, political science, and sociology, building upon the foundational studies of Anderson (1983); Gellner (1983); Hobsbawm (1990); Smith (1991); Tilly (1975); Weber (1976). Our contribution is to empirically study one of the most extensively debated episodes of nation-building in the literature, using a novel dataset on spoken languages and a range of different outcomes on identity and ideology.

Second, we add to an emerging and vibrant literature on nation-building and assimilation, which has shed light on the role of various factors such as conflict (Aghion et al., 2019; Alesina, Reich and Riboni, 2020; Dell and Querubin, 2018), inter-group contact (Cáceres-Delpiano et al., 2021), leadership (Assouad, 2021), media (Hara, 2022), military conscription (Ronconi and Ramos-Toro, 2023), population resettlement (Bazzi et al., 2019), propaganda (Blouin and Mukand, 2019; Kersting and Wolf, 2021; Li, 2022), shared experiences (Depetris-Chauvin, Durante and Campante, 2020), and welfare policies (Caprettini and Voth, 2022). Rohner and Zhuravskaya (2023) provide an extensive overview of this literature.[1] Our contribution is to focus on the role of education, one of the most important drivers of indoctrination.[2]

Third, we add to a literature on the causes and consequences of assimilationist educational policies.[3] Alesina, Giuliano and Reich (2021) study the incentives of elites to homogenize their population using education and compulsory schooling, and provide a qualitative documentation of numerous examples from France, Italy, England, Prussia, the Russian Empire, Chile, Argentina, Japan, Thailand, Turkey, the British Caribbean, Kenya, India, and various colonies. Similarly, a number of papers focus on the role of immigration (Bandiera et al., 2019) and war (Paglayan, 2022) in the provision of education. Other studies have explored the role of specific components of education policies, such as school building (Bazzi, Hilmy

[1] See also Fouka (2022), on assimilation.

[2] See also the related literature on schools, teaching practices, and ideology (Algan, Cahuc and Shleifer, 2013; Friedman et al., 2016; Lott, 1999).

[3] We also add to a broader literature on the effects of education on socio-economic outcomes (Duflo, 2001, 2004; Wantchekon, Klašnja and Novta, 2015).

and Marx, 2023; Izumi and Park, 2023), language of instruction (Clots-Figueras and Masella, 2013; Fouka, 2020; You, 2018), curriculum (Cantoni et al., 2017; Chen, Lin and Yang, 2023), or central spending (Cinnirella and Schueler, 2018), on specific outcomes of nation-building. In the context of the *Loi Guizot*, Montalbo (2021*b*) finds a positive impact on fiscal capacity. Our contribution is threefold. While most of these papers exploit cross-cohort variation, we exploit quasi-experimental, cross-sectional variation after the policy to estimate the causal effect of state-sponsored education on nation-building, locally accounting for unobservables in a regression-discontinuity framework. Additionally, we capture nation-building more comprehensively by documenting both the adoption of a common language and the building of a national identity and ideology. Finally, we capture community-level and long-term effects. A weakness of our approach is that we rely on a bundled treatment that includes all the dimensions of education policies studied in the literature. However, we provide suggestive evidence, both qualitative and quantitative, that changes in the curriculum played a major role.

Fourth, we add to a literature that has illuminated the phenomenon of minority backlash against policies of forced assimilation (Bazzi, Hilmy and Marx, 2023; Carvalho, Koyama and Williams, forthcoming; Dehdari and Gehring, 2022; Fouka, 2020; Marciante, 2023). Our contribution is to document assimilation and explore the factors that made state-sponsored education successful, namely local elites and returns to education.[4]

Last but not least, we add to a literature on political economy documenting state-building after the French Revolution (Acemoglu et al., 2011; Chambru, Henry and Marx, 2022; Jha and Wilkinson, 2023; Rosenberger and Ottinger, 2023). Our contribution is to study nation-building, the construction of a national identity, rather than state-building, the creation or consolidation of governmental institutions.

2 Historical Background

The founder of the *Annales* school, Lucien Febvre, famously asserted that "France's name is diversity" (Febvre, 1946, p. 271). Fernand Braudel, one of the most influential historians of the twentieth century, amended this statement, contending instead that "France *is* diversity, since this is no mere appearance or label, but corresponds to concrete reality: it is the dazzling triumph of the plural, of the heterogeneous, of the never-quite-the-same, of the never-quite-what-you-find-elsewhere" (Braudel, 1986, p. 38), arguing that "France is diverse to the point of absurdity" (p. 37) and that "France is not one, but many" (p. 43).

More than forty different languages or dialects were spoken historically in France, and French was a foreign language to most until the twentieth century.[5] The *lenga d'òc* (*langues*

[4]Additionally, there are also a number of literatures, on assimilation, identity, or language, that we contribute to. For assimilation and migration, e.g., Abramitzky, Boustan and Eriksson (2020); Bleakley and Chin (2010). For identity and culture, e.g., Akerlof and Kranton (2000); Alesina and Giuliano (2015). For ethno-linguistic diversity, e.g., Ahlerup and Olsson (2012); Cervellati, Chiovelli and Esposito (2019); Michalopoulos (2012). For the roots of linguistic characteristics, e.g., Blasi et al. (2019); Blouin and Dyer (2021); Dickens (2021); Galor, Özak and Sarid (2018). See also Ginsburgh and Weber (2020) for a review on the economics of language.

[5]France is also characterized by the coexistence of diverse family structures (Todd, 1990; Todd and Le Bras, 1981) along with the presence of both common law and civil law systems (le Bris, 2019). Ch. 1. of Braudel (1986) provides a

d'oc, or Occitan) were spoken in the south, while the *langues d'oïl*, including the French language, were spoken in the north. Several other languages, such as Alsatian, Basque, Breton, Catalan, Flemish, or Francoprovençal, were spoken in the rest of the country. Figure 2, Panel A, displays the spatial distribution of these languages and dialects historically.[6]

According to the Glottolog language classification (Hammarström et al., 2023), the *langues d'oc* share their closest linguistic relationship with Catalan, which, as noted by Friend (2012), was sometimes regarded as a dialect of Occitan. Although nearly extinct today, this language family holds enormous historical importance dating back to at least the tenth century (Kay, 2022). In fact, Italian poet Dante Alighieri considered *oc* (Occitan) one of the only three major families of vernacular Romance languages, along with *oïl* (French) and *sì* (Italian): "some now say *oc*, some *oïl*, and some *sì*, when they answer in the affirmative", "the language of *oc* argues in its own favour that eloquent writers in the vernacular first composed poems in this sweeter and more perfect language" (Alighieri, 1305, p. 19). Because it was the language of troubadours, Occitan also wielded a profound impact on the emergence of European poetry and culture during the Renaissance (Kay, 2013; Lafont, 2004).[7]

The earliest effort to promote the adoption of the French language can be traced back to the Ordinance of Villers-Cotterêts of August 10, 1539. It called for the use of French in all legal documents including parish records, notarized contracts, and legislation. The ordinance made French the language of the government and of local administrations, and it relegated all other dialects and languages to the status of spoken only. Nevertheless, the main target of the ordinance was the use of Latin, widespread among elites and in official documents. This process is documented in Binzel, Link and Ramachandran (2023), on the impact of the Protestant Reformation on vernacularization. According to Lodge (1993), local languages were virtually unaffected: "They naturally used the King's French in their administrative documents. However, it took a further three centuries for even the 'best families' of the town and the surrounding countryside to adopt the King's French in everyday speech" (p. 125).

At the time of the French Revolution, French was still a foreign language to the rural population, the vast majority of the overall population, and was only spoken by the elites, in the cities and in the region around Paris. In *Report on the necessity and means to annihilate dialects and to universalize the use of the French language*, Grégoire (1794) argues that "we are still, for language, the Tower of Babel"; and estimates that only three million French, out of twenty-nine million, could speak fluently the French language.[8] In 1838, Stendhal, the

thoughtful discussion of the diversity of norms and cultures in France.

[6]Note that there is no formal distinction separating a dialect from a language (Ginsburgh and Weber, 2020); rather, "a language is a dialect with an army and navy" (Weinreich, 1945).

[7]Occitan is the only foreign language used in Dante's *Divine Comedy*, which established the Florentine dialect as the standard Italian language. On the emergence of European poetry and culture, French linguist Lafont (2004) writes "This is but an observation: the Europe of poets was born Occitan. After this, the French kept it stifled for centuries under the cloak of occupation by way of their weapons and their language." According to Kay (2013), "no one questions the centrality of the troubadours to the development of European culture and, indeed, of European sensibilities" (p. 2). We also note that the Nobel Prize in Literature was awarded in 1904 to Occitan poet Frédéric Mistral, in recognition of his contributions to the preservation and promotion of Provençal literature and culture. He remains the only laureate to have been honored for work in a language not officially recognized by any state. Appendix Figure A3 displays the poem "Mirèio", by Frédéric Mistral, in Provençal, a dialect of Occitan, and its translation into French, while Appendix Figure A4 displays street names in Provençal and Alsacien.

[8]The French revolutionary governments considered the use of dialects to be a threat to the French Revolution, and tried to implement forced homogenization through compulsory schooling (Perrot, 1997). In their reports on public

writer of The Red and the Black, one of the most important and influential novels in French literature, wrote that "All nuances are constantly disappearing in France. In fifty years, there may not be any Provençals or Provençal language" (Stendhal, 1838, p. 323, own translation). Indeed, French is the only recognized official language in France today, while most other languages and dialects spoken historically are endangered (Moseley, 2010). According to mostly anecdotal and qualitative evidence, the French language was not widely adopted until the early twentieth century, with mass state-sponsored education and the development of railways (Lodge, 1993; Weber, 1976).

3 The Data

3.1 Linguistic distance

Atlas Linguistique de la France. We gather data on the language spoken around 1900 in 577 municipalities from the *Atlas Linguistique de la France*.[9] The *Atlas* was published in nine volumes from 1902 to 1910 (Gilliéron and Edmont, 1902-1910) and relies on a survey carried out by linguist Jules Gilliéron from 1897 to 1901 (henceforth "1900") to study Romance languages in rural parts of the country, where the vast majority of the population lived and did not speak French historically.

For four years, Gilliéron's assistant Edmond Edmont traveled across the country and asked locals for the standard, common pronunciation of 1,920 words or expressions in their municipality. The survey captures spoken language, including vocabulary and pronunciation, and paints a particularly detailed picture of the words used by ordinary people in their daily life at the time. Importantly, the *Atlas* aimed to document the common language spoken in the surveyed municipality—that is, the local population's "average" language as described by Gilliéron and Edmont (1902, p. 8)—rather than recording the language spoken by individual respondents.[10]

To assess selection into the sample, Appendix Table A1, Panel A, displays summary statistics for observable variables of interest both in the sample of full municipalities in France and in the *Atlas* only. We do not find any particularly relevant differences—except from the fact that municipalities in the *Atlas* are on average less populated, but the difference is not statistically significant. In fact, Gilliéron and Edmont (1902) argue that they "never searched

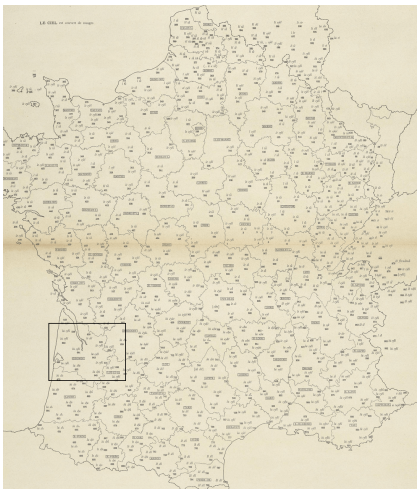
instruction, Talleyrand (1791) and Condorcet (1792) proposed universal education and a system of public primary schools in order to spur the adoption of French. For example, de Talleyrand-Périgord (1791) writes that "the language of the Constitution and of the laws will be taught to all; and this crowd of corrupted dialects, the last remnant of feudalism, will be forced to disappear." Yet, none of these projects was carried out to its term because of the high political instability. On July 20, 1794, the decree of 2 Thermidor, Year II, extended the provisions of the Ordinance of Villers-Cotterêts to local dialects, but Robespierre was arrested and executed by guillotine a few days later, leading to the suspension of the decree.

[9] The *Atlas* also contains information on sixty-two additional towns in neighboring countries, including Italy, Switzerland, and Germany. We did not digitize the data for these towns.

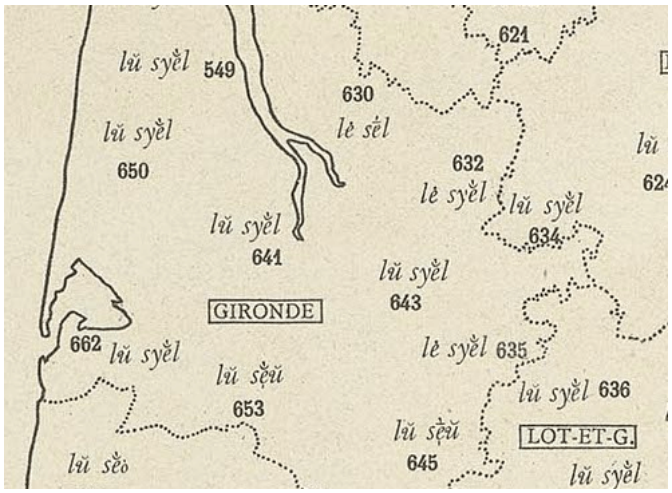
[10] For example, Gilliéron and Edmont (1902) argue that "Whether one takes as the basis of a community's language the entire population (and establishes an average), or whether, as we have generally done, only a sample is taken, consulting only one of its representatives, is inconsequential" (p. 8, own translation). To capture the common language spoken in a municipality by asking only one or at most a couple persons, Edmond Edmont mostly asked bilingual informants about the average language spoken by the population in their municipality (Gilliéron and Edmont, 1902; Le Dû, Le Berre and Brun-Trigaud, 2005). Consistent with this, Appendix Tables A8 and A9 find that individual-level characteristics were irrelevant.

Figure 1: Map of *ciel* ("sky")

Note: This figure displays a map from the *Atlas Linguistique de la France* (Gilliéron and Edmont, 1902-1910) showing the pronunciation of *ciel* ("sky") across municipalities in France (Panel A) and in the southern part of France only (Panel B). The map displays *le ciel* ("the sky"), but we only digitized the word "sky."



(Panel A) France



(Panel B) Zoomed in

for the places that could have stood out in any way" (p. 4). Most of the municipalities in the *Atlas* were selected before the start of the survey with the sole purpose of covering places "at approximately equal distances from each other" (Gilliéron and Edmont, 1902, p. 4). For a more comprehensive view, Appendix Figure A5 displays the spatial distribution of the municipalities surveyed in the *Atlas*, and Appendix Figure A6 shows that the distance to the closest town in the *Atlas* was twenty-five kilometers on average, with very little variance.

The inclusion of a large number of words in the survey allows for a direct, word-by-word comparison between languages. The words were mostly chosen before the survey was carried out. Additionally, Gilliéron carefully thought about how to not bias the survey and mostly included standard, common words (Le Dû, Le Berre and Brun-Trigaud, 2005).[11] Finally, only spontaneous replies were recorded, to further limit potential biases (Baiwir, 2019; Gilliéron and Edmont, 1902). To capture the complexity of the languages surveyed, the *Atlas* relied on the *alphabet Rousselot-Gilliéron*, a system of phonetic transcription developed at the time. The pronunciations of the words are displayed in maps—one map for each word or expression.[12] Appendix A1.1 details how we digitized the phonetical representations in the *Atlas*.

We display the pronunciation of *ciel* ("sky," the 285th map in the *Atlas*) in Figure 1. Panel A maps the pronunciation of the word across space in France. In Panel B, which covers an area of roughly 100 × 80 miles, we zoom in to the *département* of Gironde, in the historical region of Aquitaine in the southwest of France and at the frontier between the historical *langues d'oc* and *langues d'oïl* regions. The map paints a particularly detailed picture of the languages spoken at the time. We observe substantial heterogeneity, with both persistent

[11] Also included but more marginal were ancient words, words of recent origin, and words used only in some dialects.
[12] The scanned maps are publicly available on the website **cartodialect.imag.fr**.

historical differences as well as recent variations resulting from the diffusion of the French language.[13] Some municipalities had adopted the French word *ciel*, while others were still using dialects of *langues d'oc*, with *cel* or *ceu*.

Finally, we note that, while some of the variation is only phonological, it also arises from different structures, vocabularies, or languages. Here, the change from *ceu* to *ciel* involves both vowel breaking, or diphthongization, from "e" to "ie" and a change of the final letter of the word. Differences can be more profound and structural, and for that reason we underestimate the variation coming purely from language and the effect of state-sponsored education on homogenization in the historical *langues d'oc* region.

Timing of the survey. The timing of the linguistic survey is ideal for studying the effect of state-sponsored education during Guizot and to abstract from potential confounding factors resulting from the Freycinet Plan and Ferry Laws—the two other important drivers of homogenization at the time (Weber, 1976). Most adults in 1900, at the time of the *Atlas*, had not been affected by the Jules Ferry Laws of 1881 and 1882, which made education free, universal, and secular, or by the Freycinet Plan of 1878, which mandated the construction of railways, canals, and ports and was only completed in 1914.

Measure of linguistic distance. We rely on the *Atlas* to measure linguistic distance across pairs of municipalities. We use a Levenshtein-distance algorithm, defined as the minimal number of edits by insertions, deletions, or substitutions between phonetic representation of words, to capture linguistic dissimilarity in the *Atlas*. Our measure of distance is then normalized to the unit interval by dividing by the greater number of letters between the two words. We provide more details on our measure in Appendix A1.2.[14]

Table 1: Digitized words from the Swadesh list in the *Atlas Linguistique de la France*

Note: This table displays the words (or maps) that we have digitized from the *Atlas Linguistique de la France*. We also report the corresponding map number in the *Atlas* for each word.

43	animal	142	drink	796	hand	923	new	1241	sun
210	ashes	946	ear	314	how	929	night	894	swim
1	bee	1299	earth	638	knee	908	nose	1120	tail
916	black	558	fire	1200	know	76	others	233	to sing
1187	blood	1052	fish	559	leaf	1039	rain	344	to spit
250	cat	582	flowers	120	lots	1159	river	750	tongue
461	child	178	fog	788	moon	285	sky	52	trees
612	cold	615	fruit	874	mountain	418	sleep	397	two
882	die	1031	full	151	mouth	494	star	432	water
279	dog	270	hair	328	neck	1014	stone	594	wood

To measure distance, we digitized the maps showing the pronunciation of fifty representative words in the *Atlas*, such as "cat," "dog," "drink," "fruit," "hand," "mouth," "night," "rain," "sun,"

[13] Appendix Figure A7 displays a similar map of *étoile* ("star"). The words used instead of *étoile* were mostly *estello* and *estella*, from Provençal—a dialect of Occitan.

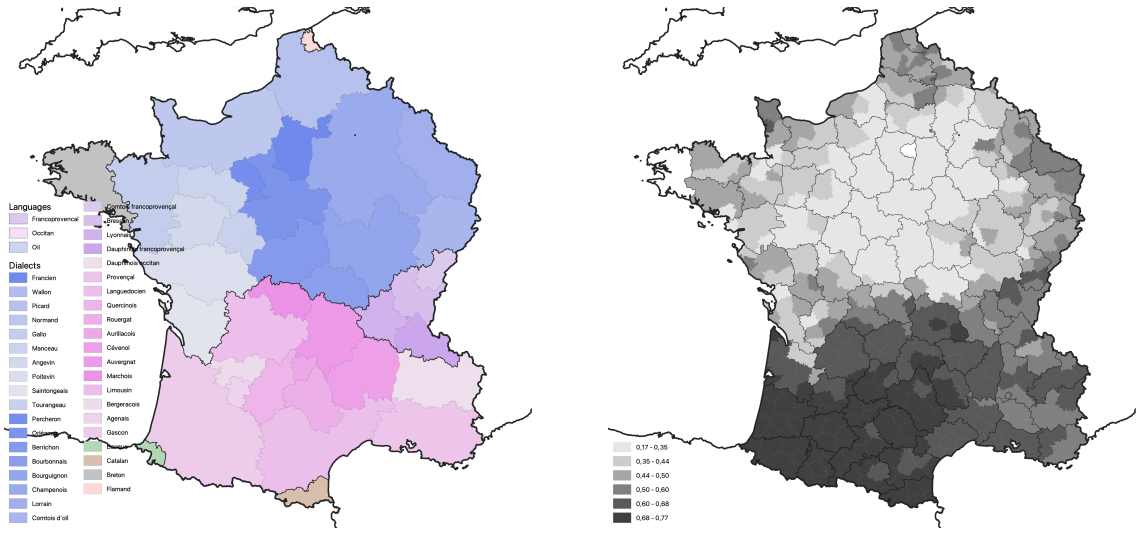
[14] Levenshtein distance was first introduced in dialectology by Kessler (1995) to study Irish dialects, and is used by Adsera and Pytlikova (2015); Dickens (2018, 2021) in more recent research by economists.

"trees," and "water," as shown in Table 1. These words were chosen from the comprehensive list established by Swadesh (1952), representing words prevalent across languages and cultures and aligning with a recognized "standardized universal list of meanings" (Ginsburgh and Weber, 2020, p. 367). We take the average Levenshtein distance over these fifty words to measure linguistic distance from French.[15] Since the *Atlas* does not provide any information on the standard French language, we use the language spoken in the town closest to Paris in the *Atlas* (Le Plessis-Robinson) as the reference point for French.[16] We refer to our main measure as the linguistic distance from French in 1900.

Figure 2, Panel B, displays the spatial distribution of our measure of linguistic distance from French in 1900. The darker areas indicate a larger linguistic distance from the standard language. Our measure tracks closely the historical language regions, displayed in Panel A, further suggesting that it captures recent changes but also deep-rooted historical differences.[17]

Figure 2: Historical language regions and linguistic distance from French in 1900

Note: This figure displays the spatial distribution of historical languages and dialects (Panel A) and linguistic distance from French in 1900 (Panel B). Linguistic distance from French is defined as the Levenshtein distance from Le Plessis-Robinson. Observations are at the municipality level, but, to ease readability, we generate Thiessen polygons around each town surveyed in the *Atlas* to plot our measure. Sources: Billy, Nadiras and Moufflet (n.d.) for historical language regions, Gilliéron and Edmont (1902-1910) and authors' calculations for linguistic distance from French.



(Panel A) Historical language regions

(Panel B) Linguistic distance from French (1900)

[15] We also measure linguistic distance across the total of 166,176 pairs of municipalities, but we focus in the paper on the linguistic distance from French.

[16] The French language is believed to originate, in part, from the francien dialect spoken in the region near Paris (Lodge, 1993, 2002). In Section 5.2, we also turn to alternative towns of reference for robustness and attempt to empirically locate the regions of origin of the French language by letting the data speak.

[17] The only other measure of linguistic distance in France not using the *Atlas* is cladistic distance, a measure of the theoretical, structural distance between languages families in large delineated areas (Desmet, Ortuño-Ortín and Wacziarg, 2012; Desmet, Ortuño-Ortín and Weber, 2009; Fearon, 2003; Fearon and Laitin, 1999; Laitin, 2000). In France, Spolaore and Wacziarg (2021) compute cladistic distance from the Oïl language family using the classification and trees in the *Ethnologue* (Eberhard, Simons and Fennig, 2021). Appendix Figure A8 displays their measure for the regions where Occitan, Francoprovencal, Basque, Breton, and Flemish were spoken historically. However, this approach suffers from limitations: it gives no sense of what the timing is since cladistic distance is only a geographical attribute of the regions in which these languages were spoken historically; we don't observe variations between dialects or within dialects; finally, the spatial distribution of speakers and their usage of dialects or languages is unknown.

In the bulk of the analysis we use our measure of linguistic distance from French, which was calculated based on the representative words from the Swadesh list whose maps we digitized.

We also show the robustness of our results to using alternative measures of linguistic distance and to relying on the entire corpus of words from the *Atlas*, using data from the Salzburg dialectometry team (Goebl et al., 2019). Goebl (2002, 2003, 2006, 2011) has been studying and digitizing the *Atlas* for more than two decades. The Salzburg dialectometry team released their data in late 2019, after we started digitizing the *Atlas*, and we discovered their website in December 2021. Their distance metrics, Relative Identity Value (RIW) and Weighted Identity Value (GIW), are closely related to ours. Appendix Figure A15 plots our measure of average Levenshtein distance across 50 words from the Swadesh list against their measures across 1,681 words. The analysis indicates a robust correlation exceeding .96. Although their data shows the average distance across the entire set of words in the *Atlas*, raw data for any specific map, or average distance for a set of representative words, is unavailable as of October 2023. Additionally, while the Levenshtein algorithm is purely automated, their distance metrics initially rely on a qualitative analysis by linguists before the data is fed into an algorithm computing distance (Goebl, 1984; Séguy, 1973).[18]

3.2 National identity and other outcomes

We rely on a set of different municipality-level measures to capture the salience of national identity, including data on participation in the French Resistance and on collaboration with the Nazis during World War II, as well as on the vote in the 1969 constitutional referendum on regionalization. Additionally, we use data on migration and trade flows across pairs of districts and *départements* to study the effect of linguistic distance on these outcomes. We detail the process of data collection of these measures in Section 6, in particular regarding the 1969 referendum, which required digitization.

3.3 Additional details on the sample

Throughout the paper, we focus on the parts of mainland France that were not annexed at any point in the nineteenth century, hence excluding Corsica, Savoy, Nice, and Alsace—as well as parts of Brittany, in Section 5, since the *Atlas* only surveyed Romance languages. We further restrict the sample to rural municipalities with recorded population that have not been split or been merged since the French Revolution.[19] Including such municipalities would bias our results to zero since historical population is the running variable. Last but not least, we focus on rural municipalities, where 80.7 percent of the French population lived

[18]Goebl (2011) provides the following description: "the original data from a given linguistic atlas are analyzed by trained geolinguists according to traditional linguistic criteria (phonetics, morphology, vocabulary, etc.). This results in a great number of disjunctive (qualitative) units called taxates" (p. 437). Then, the distance measure is computed using "the percentage of pairwise (qualitative) matchings of the elements (types, taxates) of two measuring point profiles" (Goebl et al., 2019)—with the GIW measure giving a greater weight than RIW to rare features.

[19]Out of the 577 municipalities in France, 552 are within the parts of mainland France that were not annexed at any point in the nineteenth century; 510 had a non-missing population at the time of the law; 479 were rural, with less than five thousand inhabitants; and 375 have not been merged or split since the French Revolution.

in 1836. This is crucial for understanding nation-building, the assimilation of non-elites, and the transformation from "Peasants into Frenchmen" (Weber, 1976).[20]

4 Empirical Strategy

4.1 The *Loi Guizot* of June 28, 1833

Before the July Monarchy. When François Guizot assumed the position of Minister of Public Instruction in 1832, the majority of the French population was illiterate, and the educational system was still in its early stages (Day, 1983; Furet and Ozouf, 1977; Meyers, 1976; Montalbo, 2021*a*; Weber, 1976). A comprehensive survey, the *Enquête Guizot*, was undertaken in twenty-two *départements* in 1833 to capture the state of primary instruction in France. The data shows that classrooms were only open part of the year, with irregular schedules and without trained teachers. According to Meyers (1976), the curriculum placed heavy emphasis on religious and moral education. In fact, more municipalities had religious education than had schools. Weber (1976) argues that most teachers were clergy members, or "could well have been a retired soldier, a rural constable, the local barber, innkeeper, or grocer, or simply a half-educated peasant's son" (p. 305).

The provision of mass state-sponsored education. The July Monarchy laid the foundations of education for the masses and the provision of state-sponsored education with the *Loi du 28 Juin 1833 sur l'Instruction Primaire* (*Loi Guizot*) and the *Loi du 21 Mars 1831 sur l'Organisation Municipale* (Municipal Law). In less than a decade, the Ministry of Public Instruction increased its annual budget for primary education by a staggering sixtyfold, while the budget for secondary education remained nearly constant (Reboul, 1991, p. 170).

The *Loi Guizot* mandated that municipalities with a population over five hundred inhabitants establish primary schools for boys, adhering to a state-designed national curriculum that emphasized the teaching of the national language and history and geography through state-produced and distributed textbooks; teachers received standardized training through newly state-established teachers schools, and nationally recruited school inspectors oversaw teachers and the enforcement of the law (Furet and Ozouf, 1977, p. 136).[21] The Municipal Law established the legal framework for the application of the *Loi Guizot*, promoting local democratization with the introduction of voting rights in all municipalities and addressing nepotism by forbidding family members from simultaneously serving on the municipal council in municipalities above five hundred inhabitants. Montalbo (2021*b*) shows that municipalities over five hundred inhabitants at the time of the *Loi Guizot* raised significantly more taxes to

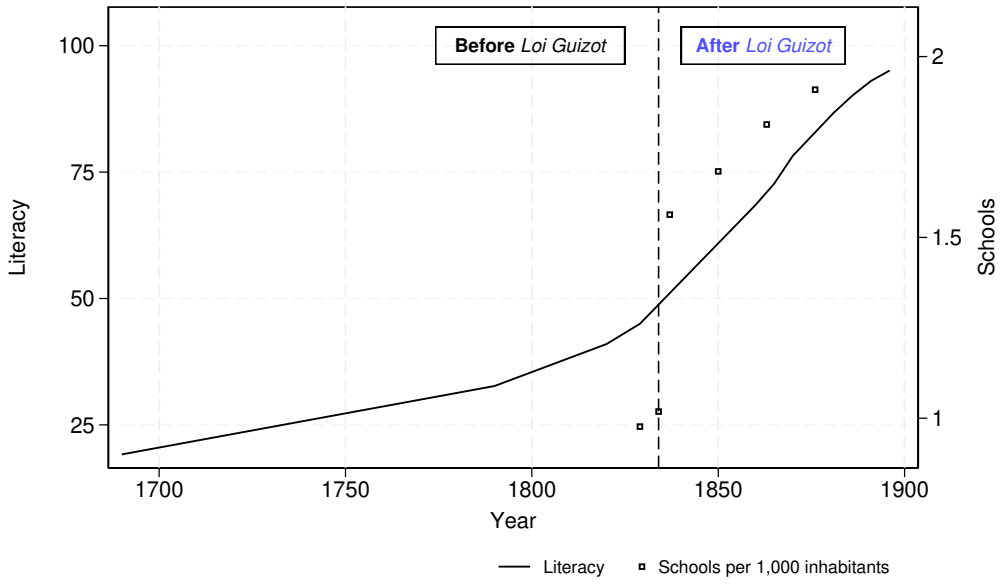
[20] Additionally, this enhances comparability and is further motivated by representativeness of the *Atlas* in rural areas. Also, the language spoken and the schools in urban settings were likely very different from those in rural places. In practice, including urban municipalities in the sample does not change our results significantly since we compare municipalities near the threshold.

[21] The Ferry Laws of 1881–82 marked a significant milestone by introducing free, inclusive, and non-religious education. The *Loi Guizot* laid the foundation for these changes by requiring the construction of schools and establishing comparable principles, although schooling was not made mandatory and schools were not necessarily required to be publicly funded. We exploit variation in the returns to education and in religiosity at the time of the *Loi Guizot* in Section 5.5.

fund its provisions, especially to support the construction of school buildings and hiring of schoolteachers.[22]

Figure 3: Literacy and schools

Note: This figure displays the literacy rate (left) and number of primary schools per thousand inhabitants (right) over time in France. Source: Statistique Générale de la France (n.d.*b*).



The number of primary schools increased nearly twofold in less than a decade, from one school for every thousand inhabitants to one for every six hundred, and universal literacy was achieved in the aftermath of the law (Figure 3).[23] Public schools came to account for more than 90 percent of schools on the eve of the Ferry Law of 1881–82 (Appendix Figure A10), and students who could not afford tuition had to be fully funded (Appendix Figure A11). In 1829, only 1.3 million children attended primary schools in France; by 1872, this number had risen to more than 4.6 million (Appendix Figure A9). Yet Guizot radically modernized primary education well beyond the building of schools; his reforms brought about the advent of state intervention in the provision of education in France (Furet and Ozouf, 1977; Nique, 1990; Prost, 1968).[24] We provide a detailed summary of the provisions of the *Loi Guizot* and its application in the remainder of the subsection, in particular regarding standardization and state intervention in the design of the curriculum and training of schoolteachers, along with additional details on the July Monarchy Laws in Appendix Section 2.

[22]The building of primary schools had to be financed by the municipalities, but the *département* or the state had to provide financial assistance to municipalities that did not have the resources or refused to raise additional taxes (*centimes*).

[23]The bulk of the increase in the number of schools built took place from 1835 to 1837, when the policy took effect, since municipalities had to buy or build a school within six years of the law being passed (Ordinance of July 16, 1833). After that, the increase can be accounted for by the Falloux and Duruy Laws, which extended the provisions of the *Loi Guizot* to schools for girls at the eight-hundred-inhabitant threshold and five-hundred-inhabitant threshold, respectively.

[24]Rosanvallon (1985) provides an interesting discussion of the motivations and political philosophy of Guizot.

Standardization of curriculum. The public primary schools built after the law were required to teach "elements of French language" and French history and geography (Article 1); and teachers were to teach in French. In 1863, local dialects were the language of instruction in only 92 out of 65,338 public primary schools in France (Weber, 1976, p. 498). Meyers (1976) documents this in his study of rural teachers in nineteenth-century France: "The language they spoke and taught was French, not patois; the history they taught, even the holidays they celebrated, were national, not regional" (p. 552). Interestingly, the expression "our ancestors the Gauls", coined by Ernest Lavisse in 1878 and depicting a fabricated national ideal, can be traced out to the early nineteenth century; and a similar form is used by Guizot in his 1875 book *L'histoire de France* (Bourdon, 2017).

The creation of five main textbooks to be used in all public primary schools was mandated by Guizot (Choppin, 1986). Then the Royal Council for Public Instruction had the authority to determine which textbooks were authorized in the classrooms of public schools, and only authorized books could be used (Article 9, *Statut portant règlement des écoles primaires élémentaires* of April 25, 1834).[25] The textbooks were free for poor students. In three years, the Ministry of Public Instructions sent millions of these textbooks to schools throughout France, including a million copies of a book on the French language (*Alphabet and Reading Primer*). At the same time, Guizot created the Society of the History of France (1833) and the Committee for Historic and Scientific Works (1834) to encourage the creation of a national narrative in history textbooks (Rosanvallon, 1985). Last, he created the *Manuel général*, a pedagogical review first published in 1832–33, to provide direct instructions and up-to-date information on recommended teaching methods.[26]

Finally, the *Loi Guizot* imposed the teaching of French and standardization through state coercion. Although it was not officially mandated, schoolteachers often relied on corporal punishments and on the use of the *symbole*, a token of shame, when students were caught speaking a language other than French (Weber, 1976, p. 313). The quasi-systematic use of coercion was especially strong in Brittany and Occitanie (Calvet, 1974; Polard, 2004). Weber (1976) argues that "Breton was hunted out of the schools" (p. 313). Today, Occitan speakers use the word *vergonha* ("shame") to describe the policies of the French government intended to shame them in the nineteenth century.[27]

Training of schoolteachers. The state took steps to dramatically improve the training of schoolteachers to help standardize the curriculum in primary schools. Guizot (1860) wrote in his memoirs that he "was trying to penetrate the soul of every schoolteacher" (p. 75). The creation of a national system of normal schools (or teachers colleges) led to radical improvements in the training of primary school teachers. All *départements* were required

[25]The Royal Council would decide which books were authorized for use in public schools only, not private or religious schools (in towns below the cutoff). See *Circulaire aux recteurs ayant pour objet de leur transmettre la liste des livres dont l'usage est autorisé dans les établissements d'instruction primaire* of June 21, 1837.

[26]For example, we find the following quote in the first edition of the *Manuel général*: "The study of the French language is paramount where provincial dialects and foreign languages have been preserved" (*Manuel général, ou Journal de l'instruction primaire, destiné à guider les instituteurs dans le choix des méthodes et à répandre dans toutes les communes de France les meilleurs principes d'éducation, publié sous la direction d'un inspecteur général des études et de plusieurs autres membres de l'Université* Novembre-Décembre 1832, Paris: Hachette, p. 12).

[27]See **Wikipedia** for reference.

to open a normal school in the departmental capital, and schoolteachers had to graduate with a *brevet de capacité* (certification) to teach in public schools. There were only three normal schools in France in 1828; by 1834, that number had risen to seventy-two (Code de l'instruction primaire, 1834, p. 42). As a result, graduates of normal schools quickly replaced older teachers. In 1846, seventeen thousand of the forty thousand primary school teachers in France had attended a normal school (Day, 1983, p. 29).

Schoolteachers became de facto civil servants with the creation of the *engagement décennal*, which required graduates of normal schools to remain on the job as schoolteachers for at least ten years (Day, 1983).[28] Their salary improved significantly with the creation of a minimum annual salary that increased sixfold by 1880.[29] According to Meyers (1976), on the eve of the Ferry Law, teachers "were earning considerably more even than many artisanal occupations" (p. 553). They were under the direct supervision of both the local municipal council and departmental school inspectors, who were themselves civil servants. The strong influence and control of the state therefore left schoolteachers with a very thin margin for acting outside of what was recommended by the state.

Finally, a body of school inspectors was created by the Royal Ordinances of February 26, 1835, and November 13, 1837 (Ravier, 2012). School inspectors were recruited nationally and were under the aegis of the Ministry of Public Instruction. Schools had to be reviewed yearly, and inspectors attended classes to review the quality of the teaching and the teachers' conduct. In particular, they had to check that only authorized books were being used (per the *Arrêté du 27 février 1835*). Ravier (1998) argues that inspectors played an important and direct role in the imposition of the French language in the north of France (p. 23).

4.2 Regression discontinuity design

Identification strategy We exploit the five-hundred-inhabitant threshold in a regression discontinuity (RD) design to study the effects of state-sponsored education on the homogenization of language and the construction of a national identity. Our empirical specification takes the following form:

$$y_i = \alpha + \tau \times state_educ_i + f(population_i) + \varepsilon_i \tag{1}$$

where y_i is the linguistic distance from French in 1900 in municipality i ($y_{i,1900}$) or some other variable capturing national identity; $state_educ_i$ is a deterministic and discontinuous function of population that equals 1 if the population of town i ($population_i$) was above five hundred at the time of the *Loi Guizot* and 0 otherwise, capturing state-sponsored education. Following Montalbo (2021*b*), we use population in 1836, when the law took effect and most schools were built (see Figure 3).[30] $f(population_i)$ is a polynomial controlling for smooth

[28] Only graduates on partial or full scholarships were required to do so, but they constituted the majority of graduates of the normal schools.

[29] While it was on par with the salary of agricultural workers at the time of the law; it then rose significantly relative to other occupations (Morrisson and Snyder, 2000).

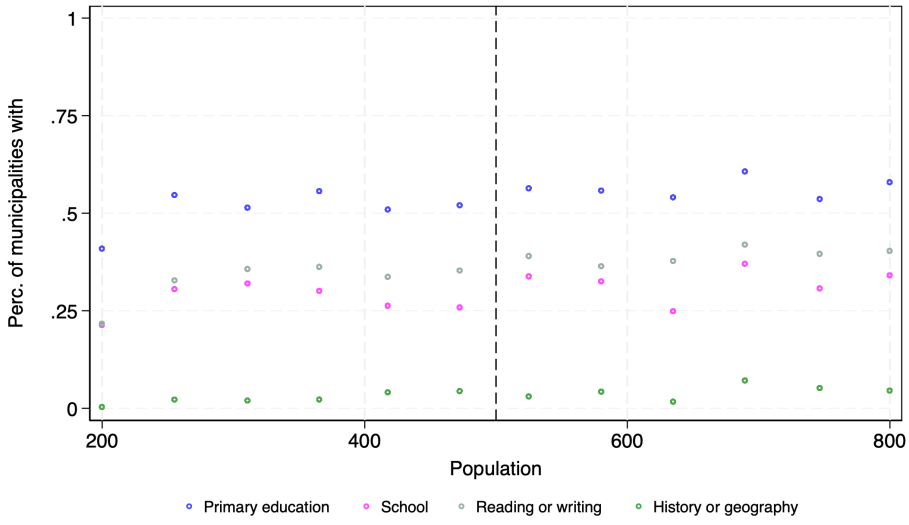
[30] Anecdotal and historical evidence suggests that the application of the law was indeed based on population at the time. In fact, municipalities had to buy or build a school within six years of the law being passed (Ordinance of July

functions of population in town i and allowing for different slopes on different sides of the cutoff to account for the conditional expectation of the outcome.

Under the assumption of continuity of conditional expectation functions of potential outcomes, τ identifies the causal effect of the treatment on linguistic distance from French in 1900 (Cattaneo, Idrobo and Titiunik, 2020; Imbens and Lemieux, 2008; Lee and Lemieux, 2010). We report conventional RD estimates of τ, following Hahn, Todd and Van der Klaauw (2001), throughout the paper, and we use the bias-corrected estimator of Calonico, Cattaneo and Titiunik (2014) in robustness analysis. Additionally, we rely on three different ways of specifying the polynomial fit—mean comparison (order 0), local linear regression (order 1), and quadratic polynomial (order 2)—and we avoid higher-order polynomials in order to limit bias and overfitting (Gelman and Imbens, 2019). Last, we follow Calonico, Cattaneo and Titiunik (2014) and Imbens and Kalyanaraman (2012) in using mean squared error (MSE) optimal bandwidths, as well as a range of other bandwidths for robustness.

Figure 4: Education before the *Loi Guizot*

Note: This figure displays the share of municipalities with primary education, with a school building (classroom), as well as those with reading, writing (grammar), geography, or history in the curriculum, at the time of the *Enquête Guizot* in Fall 1833, available at **www.inrp.fr/she/guizot**, against population at the time of the *Loi Guizot*, in 1836, around the discontinuity introduced by the policy. Observations are at the municipality level, and twenty-two *départements* are included in the data. Each point plots the average value within a bin. We rely on the mean-squared-error optimal bandwidth and on the evenly spaced mimicking-variance optimal number of bins using spacing estimators from Figure 7. Source: Guizot Survey (1833); Montalbo (2021*a*).



Nature of the treatment. We estimate a local effect of a bundled treatment capturing the prohibition of nepotism and state intervention in the provision of education, with the building of schools, the provision of education, the creation of a standardized curriculum, and the training of schoolteachers.

16, 1833). See also Figure 3 and the May 8, 1834, speech of François Guizot at the Chambre des députés (Guizot, 1863, p. 244). In Appendix Table A12, we show that there was no jump in linguistic distance at the five-hundred-inhabitant threshold using the population in other years.

We argue that the standardization of the curriculum and of the training of schoolteachers were the most important factors driving nation-building, consistent with Cantoni et al. (2017). Figure 4, using data from the *Enquête Guizot* of 1833, shows that, among surveyed municipalities near the threshold, while about one-third had schools, primary education was available in over half. The *Enquête Guizot* was conducted in Fall 1833, immediately after the passing of the *Loi Guizot*. Because the law did not immediately come into effect, the vast majority of schools were built between 1835 and 1837 and the survey predominantly captures the education system before the law. Among those that offered education, two-thirds included instruction on reading or writing, predominantly with religious education.[31] However, only 6 percent included geography in their curriculum, and a mere 2 percent included history. Since the *Loi Guizot* mandated the teaching of French history and geography—subjects not included in the curriculum before the law, even though most municipalities provided education—this data suggests that changes in the curriculum, and the accompanying changes in the training of schoolteachers, are the driving forces behind our results. Additionally, we provide quantitative evidence in Section 5.3 that the prohibition of nepotism by the Municipal Law is unlikely to directly affect linguistic homogenization.

Last but not least, because there is no municipality-level data available on schools built in the aftermath of the law, we estimate an intention-to-treat effect.[32]

Balance at the threshold. To assess the validity of our design, we test the main assumption of continuity of the conditional expectation function of potential outcomes. Using data on altitude, distance from Paris, distance from coastlines, temperature, precipitation, subscriptions to the *Encyclopedie* in 1776–79, and distance from railways in 1870, we show that observables vary smoothly at the threshold. Figure 5, Panel A, plots linguistic distance, predicted linearly using these covariates of interest, against population at the time of the *Loi Guizot*—allowing us to aggregate observables into a single conditional expectation function. We do not find any discontinuity at the threshold. Appendix Table A1, Panel B, presents comparable results for each of these variables considered individually.[33]

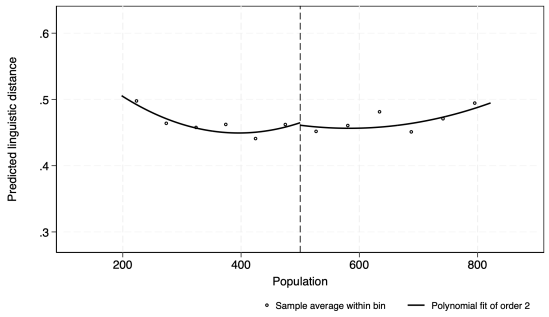
[31] In line with this, Figure 3 does not show a significant or discontinuous increase in literacy in the immediate aftermath of the *Loi Guizot*.

[32] Figure 4 shows a small increase in the number of schools above the threshold, suggesting that municipalities responded to the law, consistent with the increase in the *département*-level data. Using the available *département*-level data, Appendix Table A2 estimates the association between exposure to the *Loi Guizot*, defined as the share of municipalities with more than five hundred inhabitants, and the growth in the number of schools per capita in the aftermath of the law—after controlling for the number of schools before, in 1829. The results suggest a strong impact of the law on school building, in line with the time series in Figure 3. In order to account for endogeneity, we instrument the overall share of municipalities with more than five hundred inhabitants with the share—with more than five hundred inhabitants—in the narrower samples of municipalities within one hundred, two hundred, three hundred, and four hundred inhabitants of the five-hundred-inhabitant threshold for school building. We report first-stage, reduced-form, and two-stage-least-squares results, and note that the instrument is weaker when using smaller bandwidths. We find a large effect of exposure to the law on the building of schools across specifications.

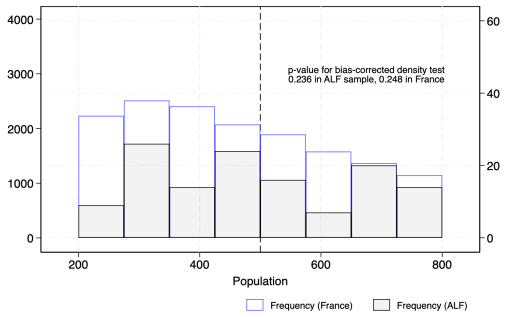
[33] We report coefficients on the difference in means above and below the threshold for each variable of interest, among municipalities that fall within different bands around the threshold. In Panel B1, we rely on the sample of all municipalities in France in order to compare a large number of places near the cutoff—for example, the 601 municipalities with between 490 and 510 inhabitants. In Panel B2, we solely rely on the sample of municipalities surveyed in the *Atlas*. We do not report the results for the ten-inhabitant bandwidth in Panel B because the sample size would be too small. We obtain similar results. In particular, we find no jump at the threshold in distance from railways, which is one the main competing explanations for homogenization (see Weber, 1976). Finally, we note that municipalities near Paris and in the historical *langues d'oïl* region are less populated near the threshold, both in the full sample and in the *Atlas* only. We also find this in Appendix Figure A12, which displays the spatial distribution of towns in the *Atlas* that are near (above or below), or away from the threshold.

Figure 5: Balance and density

Note: This figure plots predicted linguistic distance from French against population at the time of the *Loi Guizot*, in 1836, around the discontinuity introduced by the policy (Panel A), as well as the distribution of population in rural municipalities in the *Atlas Linguistique de la France* and in all of France (Panel B). In Panel A, linguistic distance from French is first regressed on the set of geographic, climatic, cultural, and economic covariates used in Appendix Table A1. We use the predicted values from this regression. Each point plots the average value within a bin, partialled out of the full set of controls used in Appendix Table A4. Linguistic distance from French is defined as the Levenshtein distance from Le Plessis-Robinson. We apply a local-polynomial fit of order 2 and rely on the mean-squared-error optimal bandwidth estimated in Figure 7 for local-polynomial estimation. In Panel B, the p-value for the bias-corrected density test relies on Cattaneo, Jansson and Ma (2018). Appendix Figure A13 plots the data for all rural municipalities in France. Observations are at the municipality level.



(Panel A) Balance in observables at the threshold: CEF of predicted linguistic distance using a set of geographic, climatic, cultural, and economic covariates



(Panel B) Density test

Finally, additional threats to identification may come from sorting and migration. Figure 5, Panel B, plots the density of the population of municipalities in France and in the *Atlas* around the threshold. We do not find any jump in the density at the threshold and, using the Cattaneo, Jansson and Ma (2020) density test, conclude that manipulation is unlikely in both samples.[34] Indeed, after 1831, municipalities were required to maintain and annually update a population register, which then had to be submitted to the central administration for verification (Biraben, 1963).[35] Additionally, we argue that migration is unlikely to be an important factor. Although there is no comprehensive data on migration at the municipality level, we find that treated municipalities did not have higher population growth after the *Loi Guizot*, as shown in Appendix Figure A14. Different concerns would arise from out-migration, in particular if there was migration from municipalities below to those above the threshold. However, gravity forces would lead to out-migration to large cities, since we estimate a local effect in small rural municipalities. Consistent with this, Blanc (2023) shows that, on average, rural-born individuals were significantly more likely to die in urban centers after 1830.

5 The Homogenization of Language

This section delves into the adoption of the French language, deemed "essential to the concept of France" by Hobsbawm (1990, p. 60). This view is echoed by French President

[34] If anything, positive sorting is particularly unlikely to affect our results since, according to François Guizot, most municipalities did not want to build the schools mandated by the law. In his May 8, 1834, speech at the Chambre des députés, Guizot (1863) noted that "there are 21,000 municipalities in France that do not feel the need for primary instruction or that do not dare to do what it takes to satisfy this need" (p. 244). Instead, negative sorting would bias our results toward zero.

[35] According to Biraben (1963, p. 309), this was incorporated into law in 1836.

Emmanuel Macron, who asserts that it "builds the unity of the nation" (Discours à l'occasion de l'inauguration de la Cité internationale de la langue française, October 30, 2023, own translation). We perform a variety of analyses and explore the drivers of homogenization.

5.1 Correlates of linguistic distance

Appendix Table A3 shows the correlates of linguistic distance from French as found through ordinary least squares (OLS) regressions. We report standardized beta coefficients and cluster standard errors at the *département* level. In column (1), we show that municipalities in the historical *langues d'oc* region are much more distant from French than those in the historical *langues d'oïl* region (where French originates from). Hence, our measure of linguistic distance captures both recent variations and, to a large extent, deep-rooted differences. We also show, in column (2), that a one-standard-deviation increase in geographical distance from Paris is associated with three-fourths of a standard deviation increase in linguistic distance. In column (3), we show a positive association between log population at the time of the *Loi Guizot* and linguistic distance. This surprising association likely stems from a combination of factors. The most important being that municipalities in the historical *langues d'oïl* region, around Paris, were on average smaller than those in the *langues d'oc* region in southern France.[36] In column (4), we report a large negative association between linguistic distance and disposable income per capita. In column (5), we show that, surprisingly, the increase in the number of schools built following the law is positively associated with linguistic distance and not statistically significant. The sign of the correlation likely underscores that the building of schools was not random but was correlated with historical linguistic distance. Indeed, we expect that municipalities in which the language spoken was initially very different from French built more schools. However, when we include all variables in column (6), in particular controlling for dummies for historical language regions and for geodesic distance from Paris, we find that the building of schools had a negative and statistically significant effect on linguistic distance from French.

5.2 Regression discontinuity results

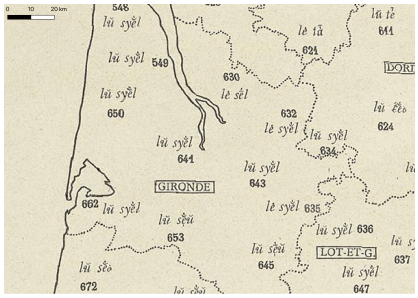
Case study. We first present a detailed preview of our data, the historical context, our empirical strategy, and our results in the historical region of Aquitaine, in the southwest of France, in Figure 6. Panel A plots the map showing the pronunciation of *ciel* ("sky") in the *Atlas*, as in Figure 1. We observe that French was spoken in municipalities located in the middle of the map. Others, in the lower part of the map, were still speaking Gascon, one of the dialects spoken in the south of France, with the word *cèl* or *cèu* used in place of *ciel*. Panels B and C superimpose the boundaries of municipalities and historical language

[36] Moreover, the *Atlas* was likely a less representative sample in larger municipalities. Additionally, the fact that population is a relatively poor proxy for income after the demographic transition (Galor and Weil, 2000) could play a role, as France had already experienced the demographic transition 100 to 150 years earlier (Blanc, 2024). Finally, it is very likely that both representativeness and selection into the sample change at higher levels of population (see Section 3.1).

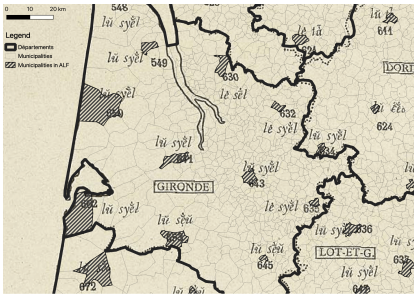
regions. Panel D displays the treatment status of municipalities near the cutoff following the *Loi Guizot* of 1833, and Panel E displays the Levenshtein distance from the pronunciation of *ciel*. Finally, Panel F displays our measure of linguistic distance across all fifty maps. Untreated municipalities were less likely to speak French. For example, the linguistic distance from French was 0.7 in Saint-Côme (town 645), which was not treated, but only 0.63 in Vélines (town 634), which was treated, even though the two municipalities are only fifty kilometers, or thirty miles, from each other and had very similar populations.[37]

Figure 6: Case study in Aquitaine

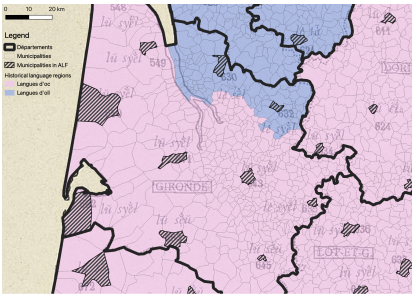
Note: This figure constitutes a case study of the historical region of Aquitaine and is designed to understand our data collection (Panels A and B), the historical framework (Panel C), the provision of state-sponsored education (Panel D), our measure of linguistic distance (Panel E), and our regression discontinuity results (Panel F). We use the map of *ciel* ("sky") from the *Atlas Linguistique de la France* (Gilliéron and Edmont, 1902-1910). In Panels D and F, "no school" refers to a town with fewer than 500 inhabitants, "school" a town with more than 500 inhabitants, and "school (away from the cutoff)" a town more than 327 inhabitants (the optimal bandwidth, as reported in Table 2) away from the cutoff at the time of the *Loi Guizot*.



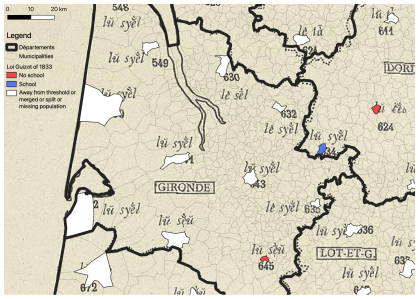
(Panel A) Map of *ciel* ("sky")



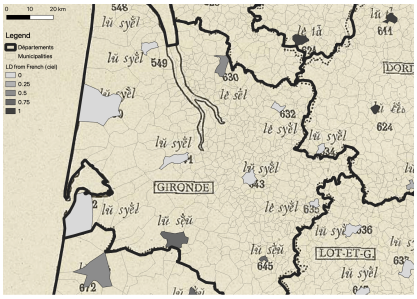
(Panel B) With municipalities



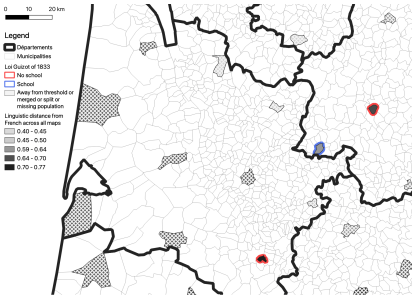
(Panel C) With historical language regions



(Panel D) With treatment status after *Loi Guizot* (state-sponsored education)



(Panel E) Linguistic distance from French (*ciel*)



(Panel F) Average linguistic distance from French across 50 maps

Regression discontinuity results. Table 2 explores the causal effect of state intervention in the provision of education on the adoption of the French language using the discontinuity introduced by the *Loi Guizot*. In the first column, we report results from controlling for a zero-th degree polynomial in population and compare the average linguistic distance with and without state intervention in the provision of education in a particularly narrow bandwidth around the threshold. The regression in column (2) accounts for linear trends in population and relies on a larger optimal bandwidth. Column (3) accounts for nonlinear,

[37] When the *Loi Guizot* was passed, there were 446 inhabitants in Saint-Côme and 801 in Vélines.

quadratic functions of population. Linguistic distance from French decreases by 0.1 at the threshold. The effect is both economically and statistically significant, at the 1 percent level, across specifications. A one-standard-deviation increase in state intervention in the provision of education is associated with 32 to 41 percent of a standard deviation decrease in linguistic distance. We also estimate that the policy decreased linguistic distance by 20 percent relative to the counterfactual mean.

Table 2: State-sponsored education and linguistic homogenization

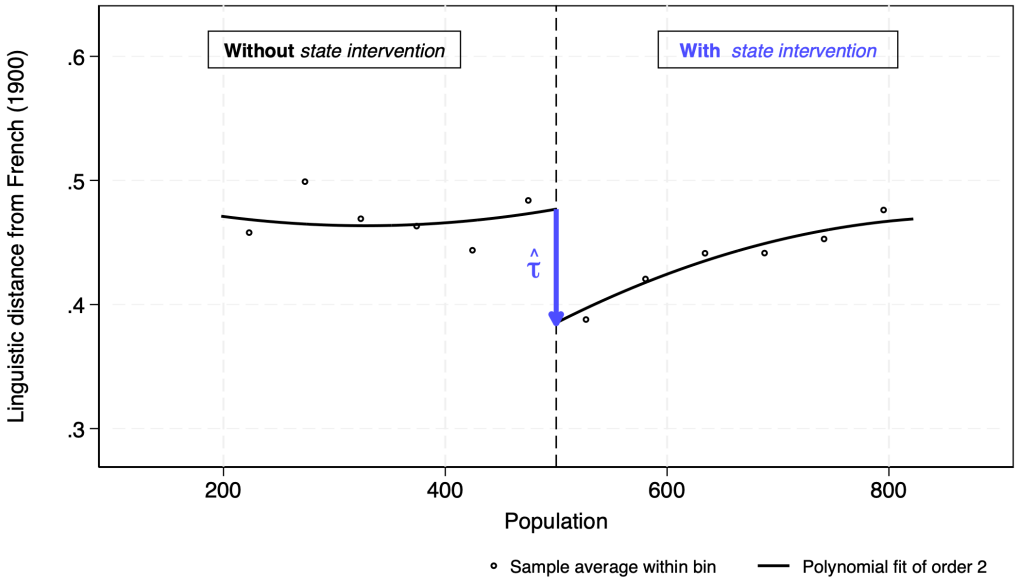
Note: This table displays regression discontinuity estimates of the effect of state-sponsored education on linguistic distance from French using the discontinuity introduced by the policy. We also report estimates of homogenization (the rate of growth over time; see Section 4.2) and standardized beta coefficients. For the estimate of the lower bound (τ/α), we compute 90 percent confidence intervals (CIs) using both the delta method (in brackets [ ]) and the Fieller method (in brackets ⟨ ⟩). We note that Fieller CIs are asymmetric. Significance stars for the estimate of τ/α are reported when both estimates (using delta and Fieller) are significantly different from 0 at the same level (here at the 1 percent level). For the estimate of the upper bound ($(\tau+\alpha)/\xi - 1$), we compute 90 percent confidence intervals using the delta method (in brackets [ ]). ξ is defined as the maximum linguistic distance in municipalities without state-sponsored education in 1900. See Appendix A3.1 for more details. We use the Stata command `nlcom` and the program `fieller` provided by Coveney (2004). Linguistic distance from French is defined as the Levenshtein distance from Le Plessis-Robinson. State-sponsored education is defined as a dummy variable that equals 1 if population at the time of the law was above five hundred inhabitants. We use local-polynomial fits of orders 0, 1, and 2 (mean, linear, quadratic) and mean-squared-error optimal bandwidths for local-polynomial estimation. Observations are at the municipality level. $^{+}$ $p < 0.2$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

	dep var: Linguistic distance from French (1900)		
	(1) Poly. degree 0	(2) Poly. degree 1	(3) Poly. degree 2
State-sponsored education			
τ	-0.09***	-0.11***	-0.11***
	(0.03)	(0.03)	(0.04)
Percentage change relative to counterfactual			
τ/α	-0.19***	-0.22***	-0.22***
	[-0.28,-0.10]	[-0.34,-0.11]	[-0.35,-0.09]
	⟨-0.28,-0.10⟩	⟨-0.34,-0.11⟩	⟨-0.36,-0.09⟩
Cutoff	500	500	500
Bandwidth selection	mse	mse	mse
Standardized beta coefficient			
$\tau * sd(school)/sd(y)$	-0.32***	-0.40***	-0.39***
Optimal bandwidth (inhab.)	139	287	327
Effective observations	60	123	141

We plot linguistic distance from French against population in Appendix Figure A16. We find a locally positive relationship between linguistic distance from French and population, which we argue captures pre-existing linguistic differences since municipalities near Paris are smaller. Appendix Table A4 presents our results after accounting for distance from Paris as well as *département* and historical-dialect fixed effects. Our results remain unchanged. Figure 7 plots linguistic distance from French against population, after accounting for the full set of controls. The relationship between linguistic distance and population is much weaker but remains positive above the threshold, likely because the sample size does not allow us to fully account for pre-existing unobservables. This is not a concern, since we expect unobservables to vary smoothly at the threshold, as confirmed by Figure 5.

Figure 7: State-sponsored education and linguistic homogenization

Note: This figure plots linguistic distance from French against population at the time of the *Loi Guizot*, in 1836, around the discontinuity introduced by the policy. Each point plots the average value within a bin, partialled out of the full set of controls used in Appendix Table A4. Linguistic distance from French is defined as the Levenshtein distance from Le Plessis-Robinson. We apply a local-polynomial fit of order 2 and a mean-squared-error optimal bandwidth for local-polynomial estimation. We rely on an evenly spaced mimicking-variance optimal number of bins. Observations are at the municipality level.



Rate of homogenization. The primary goal of nation-building is to homogenize populations by assimilating non-elites. However, we only observe linguistic distance in a single cross-section and lack temporal variation. Below, we show that our regression discontinuity design allows us to recover key parameters of interests, the change over time and the rate of change over time, under a weak assumption.[38] First, we note that we estimate the effect of state intervention in the provision of education on the homogenization of language over time. Indeed, because linguistic distance before Guizot should be continuous at the threshold, we can account for unobserved deep-rooted historical differences in the regression discontinuity framework.[39]

Second, because how fast homogenization took place is also a parameter of interest, we demonstrate in Proposition 1 that the percentage change relative to the counterfactual is a lower bound on the speed of homogenization over time, the rate of decrease in linguistic distance from French from 1833 to 1900, at the threshold. The intuition is straightforward. If homogenization took place around the threshold and state-sponsored education contributed

[38]Other papers extend the classical RD framework to incorporate variation over time. In particular, Lemieux and Milligan (2008) observe both the outcome and the treatment in intermediate years and propose a first-difference RD estimator to account for the fact that the treatment changes over time; Cellini, Ferreira and Rothstein (2010) proposes a dynamic RD design that tracks the treatment effect over time, with a dynamic treatment assignment; Grembi, Nannicini and Troiano (2016) implement a difference-in-discontinuities design in the context of treatments for the same threshold at different times.

[39]Appendix A3.1 shows that the effect of state-sponsored education on linguistic distance from French in 1900 ($y_{i,1900}$) is equal to its effect on the change in linguistic distance over time ($y_{i,1900} - y_{i,1833}$). Alternatively, we can also account for linguistic distance before the policy with historical-language or historical-dialect fixed effects, which will account for regional-level variation in the historical use of dialects.

to this process above the threshold, that is if the total change over time and the effect of treatment are of the same sign, then the rate of decrease in linguistic distance over time will be larger than the decrease relative to the counterfactual in the cross-section, which we observe.[40]

Proposition 1. *If the effect of treatment, in the cross-section, is of the same sign as the total change over time, then the percentage change relative to the counterfactual, $|\tau|/\alpha$, is a lower bound on the speed of homogenization over time, $|\boldsymbol{g}|$:* $0 \leq |\tau|/\alpha \leq |\boldsymbol{g}|$.

Proof. See Appendix A3.2.

Our results suggest that linguistic distance from French decreased by at least 20 percent from 1833 to 1900. Because we estimate τ/α using $\widehat{\tau}/\widehat{\alpha}$, which is a nonlinear combination of coefficient estimates, we cannot easily provide traditional 90 percent confidence intervals (CIs). In Table 2, we report approximate CIs using the delta method, which finds the standard error of the first-order Taylor expansion of the lower bound. We also rely on the Fieller method, which yields asymmetric CIs for the estimate of the lower bound—not its linear approximation.[41]

Finally, we note that we document an assimilation into the French culture instead of a backlash, which contrasts with the findings of Bazzi, Hilmy and Marx (2023); Carvalho, Koyama and Williams (forthcoming); Dehdari and Gehring (2022); Fouka (2020); Marciante (2023). Because the *Atlas* captured the common language spoken in each municipality, our results are driven by the adoption of French in everyday life, rather than being solely influenced by its forced use at school, which was not compulsory. If they were, the students would speak French at school and the other language outside of school, but our results do not support this. Instead, the use of French reflects the adoption of a national identity. We explore the drivers of successful assimilation in heterogeneity in Section 5.5.

5.3 Robustness and placebo

Alternative methods of estimation and bias-correction. Appendix Table A5 demonstrates the robustness of the estimated coefficients across different methods of estimation, kernels, and bias correction. In particular, we find that the bias-corrected estimator developed by Calonico, Cattaneo and Titiunik (2014), which addresses the issue of mean-squared-error optimal bandwidths being excessively large, yields somewhat larger coefficients.

[40] Additionally, if there was no change from 1833 to 1900 other than that caused by the *Loi Guizot*, $|\widehat{\tau}|/\widehat{\alpha}$ would be an unbiased estimator of the speed of homogenization. Although the assumption of absence of secular change other than the policy is strong, historical and anecdotal evidence suggest it may hold in our context since there was no other nation-building policy until much later in the nineteenth century (Weber, 1976). In that case, the estimator can also be interpreted as the effect of state-sponsored education on the rate of homogenization: when there is no homogenization below the threshold, the rate of homogenization at the threshold is also the effect of the treatment on the rate of homogenization. Note that we view this as a suggestive benchmark, but the estimation of $|\tau|/\alpha$ is highly informative about the speed of homogenization over time regardless since it is a lower bound.

[41] This is derived from the fact that $\widehat{\alpha}g - \widehat{\tau}$ is normally distributed with mean zero and known variance (Fieller, 1954). The Fieller method is less intuitive and more computationally complex than the delta method, yet it has been shown to be superior in a variety of contexts, in particular when the sample size is small (Hirschberg and Lye, 2010).

Alternative bandwidths. We rely on mean-squared-error optimal bandwidths in our main results. Appendix Table A6 demonstrates the robustness of the estimated coefficients across a range of alternative bandwidths. The effect size increases with the bandwidth when using linear and cubic specifications, while the zero-th degree polynomial specifications yield smaller and less significant results with larger bandwidths—likely because it does not account for trends and smooth functions of population to capture pre-existing differences.

Alternative distance metrics and lists of words. Appendix Table A7 demonstrates the robustness of our results across different distance metrics and lists of words, measuring distance across the entire corpus of words in the *Atlas* (Goebl et al., 2019) instead of only representative words from the Swadesh list.

Accounting for individual-level characteristics. The design of the survey aimed at capturing the average language spoken by the population in each municipality (Gilliéron and Edmont, 1902). Yet, the *Atlas* reports information on the main respondents of the survey. Using this data, Appendix Tables A8 and A9 report individual-level regressions accounting for age and gender, with standard errors clustered at the municipality level. Our results remain unchanged, suggesting that individual characteristics of the respondents did not matter and confirming that the *Atlas* is representative of the average language spoken in municipalities near the threshold.

Accounting for bilingualism. We expect places closer to the border to be more likely to be bilingual. This would complicate our interpretation of language as a proxy of identity, since municipalities near the national border could speak French as a *lingua franca*. Appendix Table A10 reports our results in the sample of municipalities more than 25 kilometers away from the border. Our results remain unchanged, except from a loss of statistical significance in the cubic specification, due to the lower number of observations.

Placebo thresholds. Appendix Table A11 reports estimates of τ using a range of placebo population thresholds. Only the five-hundred-inhabitant threshold returns statistically significant estimates, which further suggests the relevance of the *Loi Guizot*, together with the Municipal Law, for the interpretation of our results.

Placebo year of measurement of population. Appendix Table A12 reports estimates of τ relying on the five-hundred-inhabitant threshold, based on population measured at different, placebo points in time: at the time of the French Revolution, in 1793; at the time of the Municipal Law, in 1831; at the time of the Duruy Law, in 1866; at the time of the Ferry Law, in 1881; and at the time of the *Atlas*, in 1901. We only consistently find statistically significant coefficients at the time of the *Loi Guizot*, in 1836, when most schools were built, suggesting that our results predominantly capture the effect of the *Loi Guizot*, and not that of any of the other laws.

While we find small negative jumps at the threshold based on population at the time of the Municipal Law and at the time of the Duruy Law, neither is statistically significant

at conventional levels. For the Municipal Law, this could be explained by the fact that, while most schools were built between 1835 and 1837 (see Figure 3), some schools were built immediately after the *Loi Guizot* was passed, in 1833 and 1834. Therefore, using the threshold based on population at the time of the Municipal Law, in 1831 captures the direct effect of the Municipal Law but also, to a limited extent, the effect of the *Loi Guizot*.

For the Duruy Law of 1867, which extended the provisions of the Falloux Laws of 1850-51 and *Loi Guizot* to girls, requiring municipalities above 500 inhabitants to open a school for girls, we argue that the results are not statistically significant because most of the population at the time of the survey, in 1900, had not been treated by the Duruy Law, and a large number of municipalities around the threshold had already been treated by the *Loi Guizot*. Additionally, the Church's influence increased after the Falloux Laws, potentially accounting for the absence of impact: "clerical influence in French education increased after 1850 because of the large number of Catholic schools which were established" (Huckaby, 1965, p. 205).

5.4 Discussion and interpretation

The adoption of the French language. To understand if our findings reflect the adoption of a new language, or other changes such as phonetic convergence, we compute a measure of linguistic distance accounted for by language only. While our baseline measure of linguistic distance across representative words considers phonetic accents and captures both language and pronunciation, including sound replacements, the measure of distance excluding accents captures only differences in language. For example, in Figure 1, the "e" of *seu* is pronounced as a long vowel in town 645 and as a short vowel in town 653 (see Gilliéron and Edmont, 1902, p. 19). When including phonetic accents, the linguistic distance between 645 and 653 is one third. However, when excluding phonetic accents, it becomes zero, capturing the fact that both of these municipalities used the word *seu*, from the same language family, the *langues d'oc*. Appendix Figure A17 plots the share of linguistic distance from French that is accounted for by language only (that is, the ratio of linguistic distance accounted for by language to linguistic distance from French, which is accounted for by both language and pronunciation), against linguistic distance from French. For municipalities speaking a language similar to French, almost half of the measure of distance is explained by differences in pronunciation. For municipalities speaking a very different language, often in the historical *langues d'oc* region, 90 percent of the difference is explained by language and less than 10 percent by differences in pronunciation.

Table 3 reports estimates of the effect of state-sponsored education on linguistic distance including and excluding phonetic accents, in odd and even columns respectively. Our estimates show that about three-fourths of the effect of state-sponsored education on linguistic distance from French comes from the effect on language. The effect on language only is statistically significant, suggesting that state intervention in the provision of education spread not only the French pronunciation but also, and predominantly, the French language. Lodge (1993) records a process of language shift, or abandonment, in the historical *langues d'oc* region and a gradual convergence of vernacular dialects toward the common language in the *langues d'oïl*

Table 3: State-sponsored education and linguistic homogenization: language and pronunciation

Note: This table displays regression discontinuity estimates of the effect of state-sponsored education on linguistic distance from French (including and excluding phonetic accents) using the discontinuity introduced by the policy. Linguistic distance from French is defined as the Levenshtein distance from Le Plessis-Robinson. By excluding phonetic accents (in even columns), we compute linguistic distance without taking into account different pronunciations. State-sponsored education is defined as a dummy variable that equals 1 if population at the time of the law was above five hundred. Columns (1), (3), and (5) show our baseline estimates. We use local-polynomial fits of orders 0, 1, and 2 (mean, linear, quadratic) and, across all specifications, rely on the baseline mean-squared-error optimal bandwidths estimated in the baseline (including accents) for local-polynomial estimation. Observations are at the municipality level. $^{+}$ $p < 0.2$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

	dep var: Linguistic distance from French (1900) with and without phonetic accents					
	Poly. degree 0		Poly. degree 1		Poly. degree 2	
	(1)	(2)	(3)	(4)	(5)	(6)
State-sponsored education	-0.09***	-0.07**	-0.11***	-0.09**	-0.11***	-0.09*
	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)
Cutoff	500	500	500	500	500	500
Bandwidth selection	mse		mse		mse	
Without phonetic accents: language only		Yes		Yes		Yes
Bandwidth (inhab.)	139	139	287	287	327	327
Effective observations	60	60	123	123	141	141

region in the north (p. 190). We find supportive empirical evidence in Appendix Table A13, which reports estimates of the effect of state-sponsored education on phonetic distance from French—defined as the difference between linguistic distance including and excluding phonetic accents—across historical language regions. The coefficient on the interaction between state-sponsored education and historical *langues d'oïl* region is negative and statistically significant, suggesting a process of language convergence took place in that region following the adoption of state-sponsored education. We report similar results in Appendix Table A14, where we find that state-sponsored education led to the homogenization of language, not pronunciation, in the historical *langues d'oc* region, suggesting a language-shift process in this region.

Origins of French: the language of the elites. To understand the process of top-down nation-building, we empirically trace out the geographical origins of the French language; in addition, we validate our reliance on the town of Le Plessis-Robinson, ten kilometers from Paris, as the reference point for French. While historical and anecdotal evidence suggests that French was the language of the political elites, near Paris (see Lodge, 1993; Weber, 1976), no empirical evidence exists. Using the *Atlas*, we compute linguistic distance across all pairs of municipalities to understand from which regions the language adopted following the introduction of state-sponsored education came from.[42] We estimate, for each reference municipality $\boldsymbol{j}$, the following modified version of Equation 1, where $LD_{i,\boldsymbol{j}}$ is the linguistic distance between i and $\boldsymbol{j}$ (in Equation 1, we used the closest town to Paris in the *Atlas*,

[42]Barjamovic et al. (2019) undertake a similar attempt, using a structural-gravity model, to estimate the geographical locations of historical cities based on Assyrian trade links.

Le-Plessis-Robinson, as $\boldsymbol{j}$):

$$LD_{i,\boldsymbol{j}} = \alpha^{\boldsymbol{j}} + \tau^{\boldsymbol{j}} \times state_educ_i + f^{\boldsymbol{j}}(population_i) + \varepsilon_{i,\boldsymbol{j}} \tag{2}$$

In Appendix Table A15, we display RD estimates of the effect of state-sponsored education on linguistic distance from each of the six towns closest to Paris: Le-Plessis-Robinson (10 km), Sartrouville (14 km), Liancourt-Saint-Pierre (51 km), Ormoy-la-Rivière (52 km), Droue-sur-Drouette (54 km), Gommecourt (58 km). We find that state-sponsored education spread the language spoken in all of these towns, with the largest coefficient on Le-Plessis-Robinson. We extend the analysis to the rest of the municipalities in the *Atlas* in Appendix Figure A18. Panel A plots the estimated coefficients against the distance between reference municipalities and Paris, while Panel B plots the t-statistics. We find that state-sponsored education predominantly spread the language spoken in the municipalities within four hundred kilometers from Paris—that is, in the historical *langues d'oïl* region.

Appendix Figure A19 maps our estimates of the effect of state-sponsored education on linguistic distance from different reference towns across space. We plot each coefficient $\tau^{\boldsymbol{j}}$ in the location of the reference municipality $\boldsymbol{j}$ in the regression. We find that state-sponsored education predominantly led to the adoption of the language spoken near Paris and Touraine, in the Loire Valley, since the largest statistically significant estimated coefficients are for the reference towns in these regions, suggesting that French originated in Paris and the Loire Valley. The castles of the Loire Valley housed the royal court during the Renaissance, and the region (along with Paris) is believed to be one of the historical geographical centers of French Lodge (1993). Hence, our results lend credence to this hypothesis and confirm the anecdotal and historical evidence suggesting that French was the language of the elites.

5.5 Heterogeneous effects

To understand where and why nation-building was successful, we explore the heterogeneous effects of state intervention in the provision of education, noting that our findings should be interpreted with caution since we split a small sample.

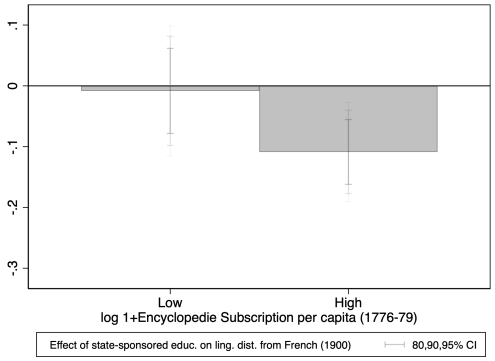
Local elites. Because nation-building was a top-down policy, and French was the language of the elites, we expect that elites were instrumental to the execution and the enforcement of the *Loi Guizot*. At the intensive margin, local elites were directly involved in the administration of public schools and in the curriculum: the schoolteachers were supervised by local municipal councils, and district councils (*comités d'arrondissement*) had the authority to influence the curriculum, within the limits of the law, by choosing, from the list of textbooks authorized by the state, the textbooks schoolteachers would use (Choppin, 1986, p. 37).[43] Second, local municipal councils voted for the building of schools and arranged for the

[43] Appendix Figure A20 displays evidence supporting this; namely, the effect of state-sponsored education was smaller in municipalities farther away from departmental capitals (*préfectures*).

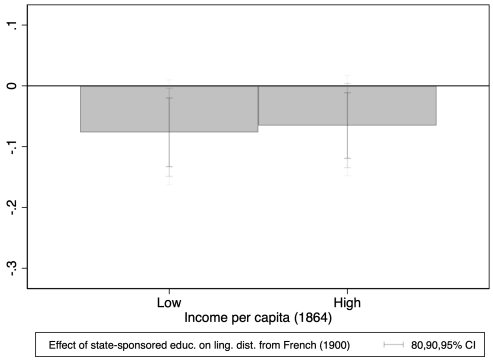
collection of additional resources in accordance with the law.[44] Below, we exploit variation in the distribution of local elites to understand their role in homogenization.

Figure 8: Heterogeneous effect of state-sponsored education

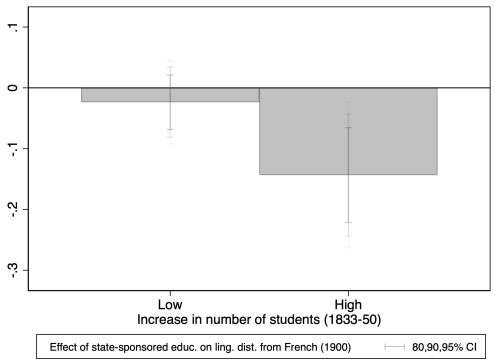
Note: This figure displays the effect of state-sponsored education on linguistic distance from French—at low and high levels of presence of presence of local elites (Panel A), economic development (Panel B), educational demand (Panel C), and returns to education (Panel D)—using the discontinuity introduced by the policy. Linguistic distance from French is defined as the Levenshtein distance from Le Plessis-Robinson. State-sponsored education is defined as a dummy variable that equals 1 if population at the time of the law was above five hundred. The presence of local elites is proxied by the log of subscriptions to the *Encyclopedie* per capita, within a 50 km buffer around each municipality, and is available at the municipality level. Economic development is proxied by disposable income per capita in 1864, available at the *département* level. Educational demand is proxied by the rate of increase in the number of students in public schools from 1833 to 1850, available at the *département* level. Returns to education are proxied by the total horsepower of steam engines in 1861, available at the *département* level All variables are coded as high if above their median value. We apply a local-polynomial fit of order 0, for ease of interpretation of the interaction term, and a mean-squared-error optimal bandwidth for local-polynomial estimation. We present the results for other polynomial degrees in Appendix Table A16. The regression discontinuity coefficient is estimated with ordinary least squares (as in Appendix Table A5). Observations are at the municipality level, and standard errors are clustered at the municipality level in Panel A and *département* level in Panels B, C, and D. Source: Darnton (1973) for presence of local elites, Delefortrie and Morice (1959) for income per capita, Statistique Générale de la France (n.d.*b*) for increase in the number of students, Statistique Générale de la France (n.d.*a*) for horsepower of steam engines.



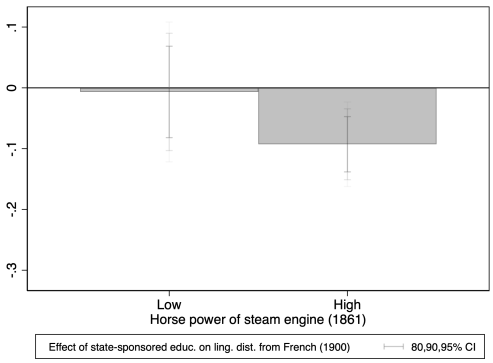
(Panel A) Local elites



(Panel B) Economic development



(Panel C) Educational demand



(Panel D) Returns to education

Figure 8 shows that the effect of state-sponsored education is larger in municipalities with local elites nearby, as proxied by *Encyclopedie* subscriptions per capita in the period 1776—79 (Darnton, 1973; Squicciarini and Voigtländer, 2015).[45] Darnton (1973) (p. 1350) discusses the occupations of 137 subscribers in the city of Besançon. The vast majority were political

[44] See Blanc and Wacziarg (2020) for a detailed account of this in the municipality of Saint-Germain-d'Anxure, as well as anecdotal evidence of a positive effect of the *Loi Guizot* on literacy. See also footnote 34.

[45] We use a municipality-level measure of all *Encyclopedie* subscriptions per capita within a fifty-kilometer radius of each municipality, while Squicciarini and Voigtländer (2015) use a *département*-level measure and discard rural towns with subscriptions from their sample.

elites: parliamentarians, members of the military, lawyers, local administrators. Squicciarini and Voigtländer (2016) show that the presence of elites is indeed associated with school building under the *Loi Guizot*. Furet and Ozouf (1977) argue that, for the elites, "school was first and foremost an instrument of control, designed to moralize and discipline the masses" (p. 120). Our results lend credence to this hypothesis, and are consistent with what Assouad (2021) finds in Turkey, with Mustafa Kemal Atatürk's nation-building policies. Yet, one could be concerned that we may not be capturing the effect of elites but rather that of income.[46] Figure 8, Panel B, shows that the effect of state-sponsored education was the same across all levels of development, which alleviates this concern.

Demand for schooling. Finally, we show that returns to education and educational demand played an important role in shaping the impact of the policy. While the *Loi Guizot* increased the supply of state-sponsored education, it did not make schooling compulsory, allowing us to understand heterogeneity. Additionally, although there was coercion within schools, the population was not coerced into attending them, which could have contributed to the success of the policy.

Figure 8, Panel C, finds that the effect of state-sponsored education was larger in *départements* where the number of students in public schools increased the most after the passing of the law. Where the benefits of schooling outweighed its costs, that is where the returns to education and the social returns to the adoption of a common language were important enough to drive educational demand up, the policy was more successful.[47]

To further capture the benefits of education, Figure 8, Panel D, shows that state-sponsored education had a larger effect in places where returns to education were high, as proxied by the horsepower of steam engines (Franck and Galor, 2021*a*). The steam engine was one of the major innovations of the Industrial Revolution. Using this data, (Franck and Galor, 2021*b*) document that the adoption of the steam engine fostered human capital accumulation in France. Indeed, we expect returns to education to be higher where technological progress was fast-paced (Ashraf and Galor, 2011; Galor, 2011, 2022; Galor and Weil, 2000). Additionally, because we are concerned that these results could be driven by geographical features, we report in Appendix Figure A24 that the effect of state-sponsored education was the same in places both near and far from coal fields.

Our results suggest that nation-building could only have been successful along the transition from stagnation to growth, as societies entered modernity, but not before, in the Malthusian epoch, when the returns to education were negligible, consistent with the hypothesis of Anderson (1983); Gellner (1983). According to Emmanuel Macron "The spread of the French language in our regions and throughout the world, in our colonies, also occurred through coercion" (October 30, 2023, Discours à l'occasion de l'inauguration de la Cité internationale de la langue française, own translation). We show that incentives mattered as well.

[46] For example, Squicciarini and Voigtländer (2015) document that subscriber density is positively associated with disposable income and various measures of development.

[47] Note that, while the supply of schools could have been impacted by educational demand in a limited number of cases, it probably was not a widespread phenomenon, since the law made *départements* or the state pay for the building of the school when municipalities would not or could not raise additional taxes to finance it. For that reason, educational demand is very unlikely to introduce endogeneity at the threshold and affect our results.

State capacity and legitimacy. We also expect that the government wanted to homogenize regions where state capacity was weak and where the legitimacy and authority of the state were contested.[48] Appendix Figure A21 shows that state-sponsored education had a larger effect in the historical *langues d'oc* region, consistent with our findings in Table 3. Appendix Figure A22 documents that state-sponsored education had a stronger impact outside the Cinq Grosses Fermes (CGF) before the French Revolution. Johnson (2019) argues that the CGF bolstered national institutions and fiscal capacity while weakening the feudal system. This implies that nation-building was more successful in less integrated regions, where state capacity and national identity were initially weak. Last but not least, Appendix Figure A23 shows that state intervention in the provision of education had a stronger effect in places where most clergy members took an oath of allegiance to the secular state during the French Revolution (Blanc, 2024; Squicciarini, 2020; Tackett, 1986). Our results shed light on the importance of state legitimacy. Koepke (1995) documents, for example, that the clergy in Morbihan, which was mostly refractory, "did not want the children to learn French, as opposed to their native Breton, for such knowledge would open the young to dangerous books and even more dangerous conversations and liaisons with the irreligious French" (p. 602).

6 Consequences of Nation-Building

This section further explores the effects of nation-building, providing a range of additional and suggestive evidence of a lasting impact on society. We first provide evidence suggesting that the adoption of a standard language promoted migration and trade. Then, we document increased participation in the French Resistance and reduced collaboration with the Nazis, alongside an increase in votes supporting political centralization.

Migration and trade. We use data on migration flows across *départements* and trade links across districts to estimate the relationship between linguistic proximity and the exchange of ideas and goods. There is unfortunately no data on migration and trade flows across municipalities, neither historically nor in contemporary times, hence we can only provide suggestive evidence of the role of the policy on these variables using data aggregated at the district or at the *département* level. We estimate Equation 3 with ordinary least squares.[49]

$$y_{d(i),d(j)} = \alpha + \beta \times LD_{i,j} + \boldsymbol{X}'_{i,j}\delta + \varepsilon_{i,j} \tag{3}$$

with $LD_{i,j}$ is the linguistic distance between municipalities i and j; $\boldsymbol{X}_{i,j}$ a vector of control variables to account for gravity factors (namely, the geographical distance between i and j, the log of the population of $d(i)$, and the log of the population of $d(j)$); and $y_{d(i),d(j)}$ a measure

[48] For example, Alesina, Giuliano and Reich (2021) argue that the threat of democratization is the strongest determinant of homogenization. The threat was particularly high at the time, following the French Revolution of 1789 and the July Revolution of 1830, and before the February Revolution of 1848. When the legitimacy of the state is contested, the population is more likely to revolt and overthrow a ruler seen as illegitimate.

[49] We cannot use RD because the level of observation (pairs of districts or pairs of *départements*) is more aggregated than the municipality level and because using pairs creates estimation issues that, to the best of our knowledge, have not been theoretically discussed.

of migration or trade flows from administrative unit $d(j)$ to unit $d(i)$.[50] Migration is observed at the pair-of-*départements* level while trade is observed at the pair-of-districts level.

Table 4: Linguistic distance, migration, and trade across pairs of towns

Note: This table displays ordinary least squares estimates of the effect of linguistic distance on migration and trade across pairs of towns. Linguistic distance between towns i and j is defined as the Levenshtein distance between i and j. Migration and trade are observed between the *département* or district of i, $d(i)$, and that of j, $d(j)$. Migration is defined as the inverse hyperbolic sine of the percentage of the population living in $d(i)$ who were born in $d(j)$ in the years 1891, 1896, and 1911, for every 10,000 inhabitants, and is available at the pair-of-*départements* level. Trade is defined by a dummy variable that equals 1 if district $d(i)$ is being supplied in goods by district $d(j)$ in 1794. Coefficients are standardized and standard errors clustered at the pair-of-administrative-units level; that is, $d(i)$ and $d(j)$ on the one hand and $d(j)$ and $d(i)$ on the other hand are considered to be the same cluster since the distance measure is symmetric. We control for the log of the population of i and the log of the population of j in even columns, and we include census-year fixed effects in the first two columns. Sources: Statistique Générale de la France (n.d.c) for population Daudin (2010) for trade. $^{+}$ $p < 0.2$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Standardized beta coefficient (clustered se)	*dep var* is:			
	Migration		Trade	
	(1)	(2)	(3)	(4)
Linguistic distance between i and j (1900)	-0.56***	-0.13***	-0.19***	-0.05***
	(0.02)	(0.01)	(0.01)	(0.01)
Geographical distance between i and j		-0.77***		-0.25***
		(0.02)		(0.01)
log population of i and j		Yes		Yes
Observations	914,112	914,112	55,752	53,400
Clusters (pairs of *départements*)	3,403	3,403		
Clusters (pairs of districts)			24,321	22,803
Adjusted R^2	0.31	0.74	0.04	0.08

Table 4 presents our results. We report standardized beta coefficients for ease of interpretation, and we cluster standard errors at the pair-of-*départements* or pair-of-districts level. We first consider migration, defined as the inverse hyperbolic sine of the percentage of the population living in $d(i)$ who were born in $d(j)$ in the years 1891, 1896, and 1911, for every 10,000 inhabitants, using data from the census on eighty-two *départements* (Statistique Générale de la France, n.d.c). We find that a one–standard deviation increase in linguistic distance between municipalities is associated with a decrease in migratory flows of more than half of a standard deviation. We account for gravity factors in column (2). The coefficient on geographical distance is particularly large, yet the effect of linguistic distance remains meaningful and statistically significant, though smaller, at the 1 percent level.

Then we study the association between linguistic distance and trade links across districts, defined by a dummy variable that equals 1 if district $d(i)$ was supplied with goods from district $d(j)$ in 1794.[51] Daudin (2010) provides data on 88 receiving districts and 551 supplier districts; less than 20 percent of pairs were linked by trade.[52] We match 67 supplier and 396

[50] Because we observe migration in 1891, 1896, and 1911, we additionally control for census-year fixed effects in the migration regressions.

[51] Recall that linguistic distance predominantly captures deep-rooted, historical differences (see Figure 2).

[52] When we aggregate the existence of links across pairs of *départements*, we find that 50 percent were linked by trade. The coefficient on linguistic distance is larger at this level, likely because there are fewer zeros in the data.

consumer districts to municipalities in the *Atlas*. In column (3), we report that a one–standard deviation increase in linguistic distance is associated with decrease in trade of 19 percent of a standard deviation. After accounting for geographical distance in the last column, the coefficient on linguistic distance diminishes but remains statistically significant. Geographical factors appear to play a smaller role in trade links than in migratory flows.

Our results suggest that language shapes migration and trade and that the homogenization brought about by state-sponsored education likely had far-reaching and persistent impacts. Although we cannot rely on the discontinuity introduced by the *Loi Guizot* to account for confounding and pre-existing factors—because of the high level of aggregation of the data and because of the nature of the regression (namely, the use of pairs)—the fact that coefficient on linguistic distance is large and survives the inclusion of controls for gravity is remarkable. Additionally, our study contributes to a growing body of research documenting an association between linguistic distance and migration and trade at the subnational level, following Falck et al. (2012); Fenske and Kala (2021).

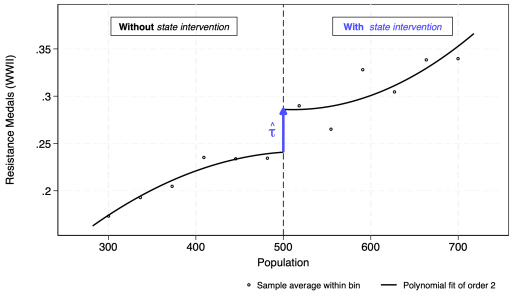
Identity and ideology. Then, we explore the persistent effects of state intervention in the provision of education, together with the adoption of the French language, on various outcomes capturing the salience of national identity and ideology. Although the (bundled) treatment now also includes the adoption of the French language, we continue to use the term 'state-sponsored education', or alternatively 'the nation-building policy'.

Figure 9, Panel A, shows that state-sponsored education and the adoption of the French language significantly increased the likelihood that individuals born in treated municipalities were heroes of the French Resistance and were awarded a Médaille de la Résistance for documented "remarkable acts of faith and courage that contributed to the resistance of the French people against the enemy." We carefully match 42,441 medals to 11,783 of the 33,814 municipalities in our sample and report the results of our estimation for different polynomial degrees, as well as for different transformations of the outcome, including dummies for the extensive margin and inverse hyperbolic sine transformations for the intensive margin, in Appendix Table A17. We also find a positive effect at the intensive margin, but only find a statistically significant effect at the extensive margin across specifications, with or without controlling for geographical distance from Paris and for *département*, historical-dialect, and Vichy-regime (the French government that chose the path of collaboration with the Nazis) fixed effects to account for local institutional and cultural factors.[53]

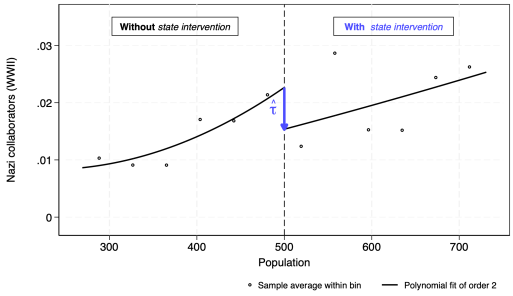
[53]Additionally, we use data on 375,346 unverified Resistance fighters from a corpus of reported acts of resistance, also taken from Site Mémoire des hommes (n.d.). Most of these individuals were not awarded Resistance Medals. This data therefore includes less heroic acts that are often unverified and unconfirmed. According to Site Mémoire des hommes (n.d.), "This corpus is not a list of all resistance fighters. Indeed, many people who engaged against the occupier during World War II never made themselves known. In addition, certain forms of resistance have not given rise to administrative procedures leading to the constitution of an individual file. Conversely, the mention of an individual in the database does not presume his capacity as a resistance fighter because the corpus contains all files opened by the Service historique de la défense, including those of individuals who, after examination, have not been recognized or approved for acts of Resistance." As a result, almost 90 percent of municipalities were the birthplace of someone with self-reported acts of resistance. We display the results at the extensive and at the intensive margins in Appendix Table A18. Our results are less precise, in particular at the extensive margin, since we lack variation, but suggest an effect of one to ten additional Resistance fighters per ten thousand inhabitants in treated municipalities. After controlling for *département*, historical-dialect, and Vichy-regime fixed effects to account for pre-existing institutional and cultural differences and differences in the reporting of acts of resistance, we improve the precision of the estimator.

Figure 9: State-sponsored education and national identity and ideology

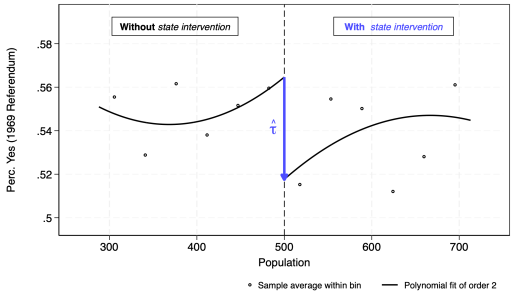
Note: This figure plots measures of the salience of national identity against population at the time of the *Loi Guizot*, in 1836, around the discontinuity introduced by the policy. Each point plots the average value within a bin, partialled out of the full set of controls used in Appendix Tables A17 (Panel A) and A22 (Panel B). In Panel A, French Resistance during World War II is defined by a dummy variable that equals 1 if at least one individual born in the municipality was awarded a Médaille de la Résistance for documented "remarkable acts of faith and courage that contributed to the resistance of the French people against the enemy." In Panel B, votes on the 1969 constitutional referendum on regionalization are defined as the vote share in favor of the referendum. We apply a local-polynomial fit of order 2 and a mean-squared-error optimal bandwidth for local-polynomial estimation. We rely on the evenly spaced mimicking-variance optimal number of bins using spacing estimators from Figure 7. Observations are at the municipality level. Sources: Site Mémoire des hommes (n.d.) for French Resistance and Dehdari and Gehring (2022) and Archives départementales du Var (1969) for the 1969 referendum.



(Panel A) Resistance heroes (WWII)



(Panel B) Nazi collaborators (WWII)



(Panel C) 1969 Referendum on regionalization

Using declassified intelligence data on collaborators from Cagé et al. (2023), Figure 9, Panel B, shows that treated municipalities were also less likely to have been the birthplace of Nazi collaborators who joined the Gestapo, the German Intelligence Service, or the Waffen-SS. Interestingly, only about 2 percent of municipalities near the threshold had Nazi collaborators, compared to almost 30 percent for heroes who were awarded a Medal of the Resistance. Therefore, while the previous section documented an effect on the language spoken by ordinary people in their everyday life, these results capture extreme cases of integration and disintegration. Appendix Table A19 displays the results at the extensive and at the intensive margin. The effect is also large and statistically significant at the intensive margin across specifications, with or without controlling for geographical distance from Paris and for *département*, historical-dialect, and Vichy-regime fixed effects, to account for local institutional and cultural factors, as well as controlling for whether the infantry regiments based in the region fought in Verdun, and in Verdun under Pétain, to account for the process documented by Cagé et al. (2023). Appendix Table A20 further shows that, while other types of collaborations, such as political and paramilitary collaboration, were also less likely to occur in treated

municipalities, the effect is largest and more consistently significant on Nazi collaborators. Indeed, Nazi collaboration involved joining organizations such as the Gestapo, the German Intelligence Service, or the Waffen-SS, and typically swearing personnel allegiance to Hitler. This constitutes a direct betrayal of the nation. Other forms of collaboration, while also indicators of betrayal, may also reflect varying degrees of alignment with anti-communist, anti-democratic, or authoritarian ideologies rather than a direct betrayal of the nation (Cagé et al., 2023, Table 7).

Then, we use data on votes against the regionalization of political authority in the 1969 referendum on regionalization, where Charles De Gaulle proposed to return significant power and autonomy to regional authorities. However, the outcome of the vote is not readily available at the municipality level. Dehdari and Gehring (2022) study this referendum in the context of Alsace-Lorraine and digitize the municipality-level results in that region from a local newspaper that published the results the day after the election. Similarly, we obtain municipality-level results in Provence from another local newspaper, *Le Petit Varois*. We searched for the edition published the day after the election in the Archives départementales du Var (1969) and digitized it. Appendix Figure A25 displays scanned pages of the newspapers. We use the data from both Lorraine and Provence and display the spatial distribution of our data in Appendix Figure A26.

Figure 9, Panel C, establishes that state-sponsored education significantly reduced the vote share in favor of regionalization. We detail the results in Appendix Table A22, where we show that the vote share in favor of regionalization decreased by 2 to 6 percentage points. This is a particularly important effect since the referendum was rejected by 52.4 percent of voters. The referendum would have been adopted and the south of France would look more like Catalonia today had nation-building not taken place. We break down the effect of the treatment by region in Appendix Table A23. The magnitude is similar, but, because of the small number of observations in each region, we barely reach conventional levels of statistical significance.

Finally, we perform a number of additional analyses, suggesting a persistent effect until today.[54] First, in Appendix Table A24, we find negative effects on votes for Jean-Marie and Marine Le Pen in the 2002 and 2017 presidential elections. Because far-right political parties, including the 'National Front' in France, often promote nationalist and xenophobic policies, we expect a positive effect of the nation-building policy. Yet, we find a negative effect. While the results are not statistically significant for Jean-Marie Le Pen, they are in most specifications for Marine Le Pen, who won around 28 percent of the votes around the threshold for state-sponsored education. Our results suggest that nation-building had a negative impact on nationalism, which is surprising since historians and political scientists such as Gellner (1983); Hobsbawm (1990); Ther (2014) view nationalism as an inevitable byproduct of the emergence of nation-states. Second, we report in Appendix Table A25 a small but positive effect of the policy on votes in favor of the 1992 referendum on the Maastricht Treaty, which laid the

[54] Appendix Table A21 also reports a placebo test and shows that the number of individuals *Morts pour la France*—individuals who died in action or from an injury or an illness contracted during service during World War I or II—is not different in treated and untreated municipalities. The results suggest that, on average, soldier death is likely to be more the product of idiosyncratic forces than of endogenous forces somewhat related to national identity.

foundation for the European Union, and the 2005 French European Constitution Referendum, which aimed to establish a constitution for the European Union but was ultimately rejected by the French electorate. The results align with far-right voting trends, as the National Front not only discussed topics such as immigration, security, and sovereignty, but also consistently opposed European integration. This suggests that nation-building not only affected national identity may have led to greater moral universalism,"the extent to which people exhibit the same level of altruism and trust towards strangers as towards in-group members" (Enke, Rodríguez-Padilla and Zimmermann, 2023, p. 1934).

7 Concluding Remarks

Using a novel, detailed dataset on spoken languages in France and a natural experiment, our paper documents the adoption of a common language and the construction of a national identity and ideology in the process of nation-building in a fragmented society.

References

Abramitzky, Ran, Leah Boustan, and Katherine Eriksson. 2020. "Do Immigrants Assimilate More Slowly Today Than in the Past?" *American Economic Review: Insights*, 2(1): 125–41.

Acemoglu, Daron, Davide Cantoni, Simon Johnson, and James A. Robinson. 2011. "The Consequences of Radical Reform: The French Revolution." *American Economic Review*, 101(7): 3286–3307.

Adsera, Alicia, and Mariola Pytlikova. 2015. "The role of language in shaping international migration." *The Economic Journal*, 125(586): F49–F81.

Aghion, Philippe, Xavier Jaravel, Torsten Persson, and Dorothée Rouzet. 2019. "Education and military rivalry." *Journal of the European Economic Association*, 17(2): 376–412.

Ahlerup, Pelle, and Ola Olsson. 2012. "The roots of ethnic diversity." *Journal of Economic Growth*, 17(2): 71–102.

Akerlof, George A, and Rachel E Kranton. 2000. "Economics and identity." *The quarterly journal of economics*, 115(3): 715–753.

Alesina, Alberto, and Paola Giuliano. 2015. "Culture and Institutions." *Journal of Economic Literature*, 53(4): 898–944.

Alesina, Alberto, Bryony Reich, and Alessandro Riboni. 2020. "Nation-building, nationalism, and wars." *Journal of Economic Growth*, 25(4): 381–430.

Alesina, Alberto, Paola Giuliano, and Bryony Reich. 2021. "Nation-Building and Education." *The Economic Journal*. ueab001.

Algan, Yann, Pierre Cahuc, and Andrei Shleifer. 2013. "Teaching practices and social capital." *American Economic Journal: Applied Economics*, 5(3): 189–210.

Alighieri, Dante. 1305. *Dante: De Vulgari Eloquentia (2005 edition, edited and translated by Steven Botterill). Cambridge Medieval Classics*, Cambridge University Press.

Anderson, Benedict. 1983. *Imagined communities: Reflections on the origin and spread of nationalism.* London:Verso.

Archives départementales du Var. 1969. *Le Petit Varois, la Marseillaise.* Cote 326 PRS 79.

Ashraf, Quamrul, and Oded Galor. 2011. "Dynamics and Stagnation in the Malthusian Epoch." *American Economic Review*, 101(5): 2003–2041.

Assouad, Lydia. 2021. "Charismatic Leaders and Nation Building." Working paper.

Baiwir, Esther. 2019. "Francisation des dialectes d'oïl : de l'usage des atlas linguistiques comme termes de comparaison." *Langages*, 215(3): 27–42.

Bandiera, Oriana, Myra Mohnen, Imran Rasul, and Martina Viarengo. 2019. "Nation-building through compulsory schooling during the age of mass migration." *The Economic Journal*, 129(617): 62–109.

Barjamovic, Gojko, Thomas Chaney, Kerem Coşar, and Ali Hortaçsu. 2019. "Trade, Merchants, and the Lost Cities of the Bronze Age." *The Quarterly Journal of Economics*, 134(3): 1455–1503.

Bazzi, Samuel, Arya Gaduh, Alexander D Rothenberg, and Maisy Wong. 2019. "Unity in diversity? How intergroup contact can foster nation building." *American Economic Review*, 109(11): 3978–4025.

Bazzi, Samuel, Masyhur Hilmy, and Benjamin Marx. 2023. "Religion, Education, and the State." Working paper.

Billy, Pierre-Henri, Sébastien Nadiras, and Jean-François Moufflet. n.d.. "Carte linguistique de la France, Archives Nationales (**www.archives-nationales.culture.gouv.fr/fr_FR/web/guest/cartographie-historique**)."

Binzel, Christine, Andreas Link, and Rajesh Ramachandran. 2023. "Language, Knowledge, and Growth: Evidence from Early Modern Europe." Working Paper.

Biraben, Jean-Noël. 1963. "Inventaire des listes nominatives de recensement en France." *Population*, 2.

Blanc, Guillaume. 2023. "Crowdsourced genealogies." Working paper.

Blanc, Guillaume. 2024. "The Cultural Origins of the Demographic Transition in France." Lewis Lab Working Paper 2024-02.

Blanc, Guillaume, and Romain Wacziarg. 2020. "Change and Persistence in the Age of Modernization: Saint-Germain-d'Anxure, 1730-1895." *Explorations in Economic History*, 78.

Blasi, Damián E, Steven Moran, Scott R Moisik, Paul Widmer, Dan Dediu, and Balthasar Bickel. 2019. "Human sound systems are shaped by post-Neolithic changes in bite configuration." *Science*, 363(6432).

Bleakley, Hoyt, and Aimee Chin. 2010. "Age at Arrival, English Proficiency, and Social Assimilation among US Immigrants." *American Economic Journal: Applied Economics*, 2(1): 165–92.

Blouin, Arthur, and Julian Dyer. 2021. "How Cultures Converge: An Empirical Investigation of Trade and Linguistic Exchange." University of Toronto, Department of Economics Working Papers tecipa-691.

Blouin, Arthur, and Sharun W Mukand. 2019. "Erasing ethnicity? Propaganda, nation building, and identity in Rwanda." *Journal of Political Economy*, 127(3): 1008–1062.

Bourdon, Étienne. 2017. *Forging the nation ahead with the Gauls: Ernest Lavisse and the making of ancestors.* Lyon:ENS Éditions.

Braudel, Fernand. 1986. *The identity of France: History and environment ; translated from the French by Sian Reynolds.* Vol. 1. 1990 ed., London:Collins.

Cáceres-Delpiano, Julio, Antoni-Italo De Moragas, Gabriel Facchini, and Ignacio González. 2021. "Intergroup contact and nation building: Evidence from military service in Spain." *Journal of Public Economics*, 201: 104477.

Cagé, Julia, Anna Dagorret, Pauline Grosjean, and Saumitra Jha. 2023. "Heroes and Villains: The Effects of Heroism on Autocratic Values and Nazi Collaboration in France." *American Economic Review*, 113(7): 1888–1932.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica*, 82(6): 2295–2326.

Calvet, Louis-Jean. 1974. *Linguistique et colonialisme : Petit traité de glottophagie.* Paris:Éditions Payot.

Cantoni, Davide, Yuyu Chen, David Y Yang, Noam Yuchtman, and Y Jane Zhang. 2017. "Curriculum and ideology." *Journal of Political Economy*, 125(2): 338–392.

Caprettini, Bruno, and Hans-Joachim Voth. 2022. "New Deal, New Patriots: How 1930s Government Spending Boosted Patriotism During World War II." *The Quarterly Journal of Economics*, 138(1): 465–513.

Carvalho, Jean-Paul, Mark Koyama, and Cole Williams. forthcoming. "Resisting Education." *Journal of the European Economic Association.*

Cattaneo, Matias D., Michael Jansson, and Xinwei Ma. 2018. "Manipulation Testing Based on Density Discontinuity." *The Stata Journal*, 18(1): 234–261.

Cattaneo, Matias D., Michael Jansson, and Xinwei Ma. 2020. "Simple Local Polynomial Density Estimators." *Journal of the American Statistical Association*, 115(531): 1449–1455.

Cattaneo, Matias D., Nicolás Idrobo, and Rocío Titiunik. 2020. *A Practical Introduction to Regression Discontinuity Designs: Foundations. Elements in Quantitative and Computational Methods for the Social Sciences*, Cambridge University Press.

Cellini, Stephanie Riegg, Fernando Ferreira, and Jesse Rothstein. 2010. "The Value of School Facility Investments: Evidence from a Dynamic Regression Discontinuity Design." *The Quarterly Journal of Economics*, 125(1): 215–261.

Cervellati, Matteo, Giorgio Chiovelli, and Elena Esposito. 2019. "Bite and Divide: Malaria and Ethnolinguistic Diversity."

Chambru, Cédric, Emeric Henry, and Benjamin Marx. 2022. "The Dynamic Consequences of State-Building: Evidence from the French Revolution." Sciences Po Economics Discussion Papers No. 2021-13.

Chen, Wei-Lin, Ming-Jen Lin, and Tzu-Ting Yang. 2023. "Curriculum and national identity: Evidence from the 1997 curriculum reform in Taiwan." *Journal of Development Economics*, 163: 103078.

Choppin, Alain. 1986. "Le cadre législatif et réglementaire des manuels scolaires. I, De la Révolution à 1939." *Histoire de l'éducation*, 29.

Cinnirella, Francesco, and Ruth Schueler. 2018. "Nation building: The role of central spending in education." *Explorations in Economic History*, 67: 18–39.

Clots-Figueras, Irma, and Paolo Masella. 2013. "Education, language and identity." *The Economic Journal*, 123(570): F332–F357.

Code de l'instruction primaire. 1834. *Code de l'instruction primaire contenant l'historique de la législation primaire depuis 1783.* Paris:Librairie Normale de P. Dupont.

Coveney, Joseph. 2004. "FIELLER: Stata module to calculate confidence intervals of quotients of normal variates." Statistical Software Components, Boston College Department of Economics.

Darnton, Robert. 1973. "The Encyclopédie Wars of Prerevolutionary France." *The American Historical Review*, 78(5): 1331–1352.

Daudin, Guillaume. 2010. "Domestic Trade and Market Size in Late-Eighteenth-Century France." *The Journal of Economic History*, 70(3): 716–743.

Day, C. R. 1983. "The Rustic Man: The Rural Schoolmaster in Nineteenth-Century France." *Comparative Studies in Society and History*, 25(1): 26–49.

Dehdari, Sirus H., and Kai Gehring. 2022. "The Origins of Common Identity: Evidence from Alsace-Lorraine." *American Economic Journal: Applied Economics*, 14(1): 261–92.

Delefortrie, Nicole, and Janine Morice. 1959. *Les Revenus departmentaux en 1864 et en 1954.* Paris:Armand Colin.

Dell, Melissa, and Pablo Querubin. 2018. "Nation Building Through Foreign Intervention: Evidence from Discontinuities in Military Strategies." *The Quarterly Journal of Economics*, 133(2): 701–764.

Depetris-Chauvin, Emilio, Ruben Durante, and Filipe Campante. 2020. "Building nations through shared experiences: Evidence from African football." *American Economic Review*, 110(5): 1572–1602.

Desmet, Klaus, Ignacio Ortuño-Ortín, and Romain Wacziarg. 2012. "The political economy of linguistic cleavages." *Journal of Development Economics*, 97(2): 322–338.

Desmet, Klaus, Ignacio Ortuño-Ortín, and Shlomo Weber. 2009. "Linguistic Diversity and Redistribution." *Journal of the European Economic Association*, 7(6): 1291–1318.

de Talleyrand-Périgord, Charles-Maurice. 1791. "Rapport sur l'instruction publique." *Comité de Constitution, Assemblée nationale.*

Dickens, Andrew. 2018. "Ethnolinguistic favoritism in african politics." *American Economic Journal: Applied Economics*, 10(3): 370–402.

Dickens, Andrew. 2021. "Understanding Ethnolinguistic Differences: The Roles of Geography and Trade." *The Economic Journal.*

Duflo, Esther. 2001. "Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment." *American Economic Review*, 91(4): 795–813.

Duflo, Esther. 2004. "The medium run effects of educational expansion: Evidence from a large school construction program in Indonesia." *Journal of Development Economics*, 74(1): 163–197.

Eberhard, David M., Gary F. Simons, and Charles D. Fennig. 2021. *Ethnologue: Languages of the World (24th edition).* Dallas, TX:SIL International.

Enke, Benjamin, Ricardo Rodríguez-Padilla, and Florian Zimmermann. 2023. "Moral Universalism and the Structure of Ideology." *The Review of Economic Studies*, 90(4): 1934–1962.

Falck, Oliver, Stephan Heblich, Alfred Lameli, and Jens Südekum. 2012. "Dialects, cultural identity, and economic exchange." *Journal of urban economics*, 72(2-3): 225–239.

Fearon, James D. 2003. "Ethnic and Cultural Diversity by Country." *Journal of Economic Growth*, 8(2): 195–222.

Fearon, James D, and David D Laitin. 1999. "Weak states, rough terrain, and large-scale ethnic violence since 1945." 2–5, Citeseer.

Febvre, Lucien. 1946. "Que la France se nomme diversité. À propos de quelques études jurassiennes." *Annales*, 1(3): 271–274.

Fenske, James, and Namrata Kala. 2021. "Linguistic Distance and Market Integration in India." *The Journal of Economic History*, 81(1): 1–39.

Fieller, E. C. 1954. "Some Problems in Interval Estimation." *Journal of the Royal Statistical Society: Series B (Methodological)*, 16(2): 175–185.

Fouka, Vasiliki. 2020. "Backlash: The unintended effects of language prohibition in US schools after World War I." *The Review of Economic Studies*, 87(1): 204–239.

Fouka, Vasiliki. 2022. "Assimilation in Historical Political Economy." In *The Oxford Handbook of Historical Political Economy*. Oxford University Press.

Franck, Raphaël, and Oded Galor. 2021*a*. "Flowers of evil? Industrialization and long run development." *Journal of Monetary Economics*, 117: 108–128.

Franck, Raphaël, and Oded Galor. 2021*b*. "Technology-Skill Complementarity in Early Phases of Industrialisation." *The Economic Journal*, 132(642): 618–643.

Friedman, Willa, Michael Kremer, Edward Miguel, and Rebecca Thornton. 2016. "Education as Liberation?" *Economica*, 83(329): 1–30.

Friend, Julius W. 2012. *Stateless Nations: Western European Regional Nationalisms and the Old Nations*. New York City:Palgrave Macmillan.

Furet, François, and Jacques Ozouf. 1977. *Lire et écrire : l'alphabétisation des Français de Calvin à Jules Ferry*. Paris:Les Editions de Minuit.

Galor, Oded. 2011. *Unified Growth Theory*. Princeton:Princeton University Press.

Galor, Oded. 2022. *The Journey of Humanity: The Origins of Wealth and Inequality*. Dutton.

Galor, Oded, and David Weil. 2000. "Population, Technology, and Growth: From the Malthusian Regime to the Demographic Transition and Beyond." *American Economic Review*, 90(4): 806–828.

Galor, Oded, Ömer Özak, and Assaf Sarid. 2018. "Geographical roots of the coevolution of cultural and linguistic traits." *Available at SSRN 3284239*.

Gellner, Ernest. 1983. *Nations and Nationalism*. Ithaca:Cornell University Press.

Gelman, Andrew, and Guido Imbens. 2019. "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs." *Journal of Business & Economic Statistics*, 37(3): 447–456.

Gilliéron, Jules, and Edmond Edmont. 1902. *Atlas linguistique de la France: notice servant a l'intelligence des cartes*. Paris:Champion.

Gilliéron, Jules, and Edmond Edmont. 1902-1910. *Atlas linguistique de la France*. Vol. 1-9, Paris:Champion.

Ginsburgh, Victor, and Shlomo Weber. 2020. "The Economics of Language." *Journal of Economic Literature*, 58(2): 348–404.

Goebl, Hans. 1984. *Dialektometrische Studien. Anhand italoromanischer, ratoromanischer und galloromanischer Sprachmaterialien aus AIS und ALF. 3 Bande. Beihefte zur Zeitschrift fur romanische Philologie*, Tubingen:Niemeyer.

Goebl, Hans. 2002. "Analyse dialectométrique des structures de profondeur de l'ALF." *Revue de linguistique romane*, 66: 261–262.

Goebl, Hans. 2003. "Regards dialectométriques sur les données de l'Atlas linguistique de la France (ALF): relations quantitatives et structures de profondeur." *Estudis Romanics*, 25: 59–120.

Goebl, Hans. 2006. "Recent Advances in Salzburg Dialectometry." *Literary and Linguistic Computing*, 21(4): 411–435.

Goebl, Hans. 2011. "22. Dialectometry and quantitative mapping." *Volume 2 Language Mapping: An International Handbook of Linguistic Variation*, , ed. Alfred Lameli, Roland Kehrein and Stefan Rabanus, 433–464. De Gruyter Mouton.

Goebl, Hans, Salzburg dialectometry team, Edgar Haimerl, Pavel Smečka, Bernhard Castellazzi, and Yves Scherrer. 2019. "Dialectometry ALF (**www.dialektkarten.ch/dmviewer/alf/index.en.html**), retrieved December 13, 2021."

Grégoire, Henri. 1794. "Rapport sur la nécessité et les moyens d'anéantir les patois et d'universaliser l'usage de la langue française." *Convention nationale.*

Grembi, Veronica, Tommaso Nannicini, and Ugo Troiano. 2016. "Do Fiscal Rules Matter?" *American Economic Journal: Applied Economics*, 8(3): 1–30.

Guizot, François. 1833. "64. 2 janvier 1833. Exposé des motifs de la loi sur l'instruction primaire."

Guizot, François. 1860. *Mémoires pour servir à l'histoire de mon temps: La Révolution de 1830 (Tome 3).* Paris:Michel Lévy Frères.

Guizot, François. 1863. *Histoire parlementaire de France: recueil complet des discours prononcés dans les Chambres de 1819 à 1848 (Vol. 2).* Paris:Michel Lévy Frères.

Guizot Survey. 1833. *L'enquête Guizot.* Institut national de recherche pédagogique. Available at **inrp.fr/she/guizot/**.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw. 2001. "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design." *Econometrica*, 69(1): 201–209.

Hammarström, Harald, Robert Forkel, Martin Haspelmath, and Sebastian Bank. 2023. "Glottolog database 4.8."

Hara, Tomohiro. 2022. "Building the "rainbow nation" through mass media: television, cultural diversity, and national unity in post-Apartheid South Africa." Working paper.

Hirschberg, Joe, and Jenny Lye. 2010. "A Geometric Comparison of the Delta and Fieller Confidence Intervals." *The American Statistician*, 64(3): 234–241.

Hobsbawm, E. J. 1990. *Nations and Nationalism since 1780: Programme, Myth, Reality.* Cambridge:Cambridge University Press.

Huckaby, John K. 1965. "Roman Catholic Reaction to the Falloux Law." *French Historical Studies*, 4(2): 203–213.

Imbens, Guido, and Karthik Kalyanaraman. 2012. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." *The Review of Economic Studies*, 79(3): 933–959.

Imbens, Guido W., and Thomas Lemieux. 2008. "Regression discontinuity designs: A guide to practice." *Journal of Econometrics*, 142(2): 615–635.

Izumi, Yutaro, and Sangyoon Park. 2023. "Education and Wartime Mobilization: Evidence from Colonial Korea during WWII." Working paper.

Jha, Saumitra, and Steven I. Wilkinson. 2023. "Revolutionary Contagion." Stanford University Graduate School of Business Research Paper No. 4084.

Johnson, Noel D. 2019. "Taxes, National Identity, and Nation Building: Evidence from France." Working paper.

Kay, Sarah. 2013. *Parrots and Nightingales: Troubadour Quotations and the Development of European Poetry.* Philadelphia, PA:University of Pennsylvania Press.

Kay, Sarah. 2022. *Occitan (Ch. 8.), in Literary Beginnings in the European Middle Ages (edited by Mark Chinca and Christopher Young). Cambridge Studies in Medieval Literature*, Cambridge:Cambridge University Press.

Kersting, Felix, and Nikolaus Wolf. 2021. "On the Origins of National Identity: German Nation-Building after Napoleon."

Kessler, Brett. 1995. "Computational Dialectology in Irish Gaelic." *EACL '95*, 60–66. San Francisco, CA, USA:Morgan Kaufmann Publishers Inc.

Koepke, Robert L. 1995. "Cooperation, Not Conflict: Curés and Primary School Inspectors in July Monarchy France, 1833-1848." *Church History*, 64(4): 594–609.

Lafont, Robert. 2004. *Préface. In G. Agresti (Ed.), Anthologie de la nouvelle écriture occitane.* Montpeirós:Jorn.

Laitin, David D. 2000. "What Is a Language Community?" *American Journal of Political Science*, 44(1): 142–155.

le Bris, David. 2019. "Testing legal origins theory within France: Customary laws versus Roman code." *Journal of Comparative Economics*, 47(1): 1–30.

Le Dû, Jean, Yves Le Berre, and Guylaine Brun-Trigaud. 2005. *Lectures de l'atlas linguistique de la France de Gilliéron et Edmont: du temps dans l'espace ; essai d'interprétation des cartes de L'Atlas linguistique de la France de Jules Gilliéron et Edmond Edmont.* Paris:Éditions du Comité des Travaux Historiques et Scientifiques.

Lee, David S., and Thomas Lemieux. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature*, 48(2): 281–355.

Lemieux, Thomas, and Kevin Milligan. 2008. "Incentive effects of social assistance: A regression discontinuity approach." *Journal of Econometrics*, 142(2): 807–828.

Li, Peiyuan. 2022. "Political Repression, Media Propaganda, and Nation-Building." Working Paper.

Lodge, Raymond Anthony. 1993. *French, from Dialect to Standard.* London:Routledge.

Lodge, Raymond Anthony. 2002. "« Francien » et « français de Paris »." *Linx*, 12: 149–172.

Lorain, Paul. 1837. *Tableau de l'instruction primaire en France: d'après des documents authentiques, et notamment d'après les rapports adressés au ministre de l'instruction publique par le 490 inspecteurs chargés de visiter toutes les écoles de France, à la fin de 1833.* Paris:Hachette.

Lott, Jr., John R. 1999. "Public Schooling, Indoctrination, and Totalitarianism." *Journal of Political Economy*, 107(6): 127–157.

Marciante, Gianni. 2023. "When nation building goes south: draft evasion, government repression, and the origins of the Sicilian mafia." Working paper.

Meyers, Peter V. 1976. "Professionalization and Societal Change: Rural Teachers in Nineteenth Century France." *Journal of Social History*, 9(4): 542–558.

Michalopoulos, Stelios. 2012. "The origins of ethnolinguistic diversity." *American Economic Review*, 102(4): 1508–39.

Montalbo, Adrien. 2021*a*. "Schools without a law: Primary education in France from the Revolution to the Guizot Law." *Explorations in Economic History*, 79: 101364.

Montalbo, Adrien. 2021*b*. "State intervention, education supply and economic growth in nineteenth-century France." Working paper.

Morrisson, Christian, and Wayne Snyder. 2000. "The income inequality of France in historical perspective." *European Review of Economic History*, 4(1): 59–83.

Moseley, Christopher. 2010. *Atlas of the World's Languages in Danger (3rd ed.).* Paris:UNESCO Publishing.

Nique, Christian. 1990. *Comment l'école devint une affaire d'Etat (1815-1840).* Paris:Nathan.

Paglayan, Agustina S. 2022. "Education or Indoctrination? The Violent Origins of Public School Systems in an Era of State-Building." *American Political Science Review*, 116(4): 1242–1257.

Perrot, Marie-Clémence. 1997. "La politique linguistique pendant la Révolution française." *Mots/Les langages du politique*, 52.

Polard, André. 2004. *Cet obscur sentiment qu'est la honte, in: L'épilepsie du sujet.* Paris:Harmattan.

Prost, Antoine. 1968. *Histoire de l'enseignement en France, 1800-1967.* Paris:Armand Colin.

Ravier, Joël. 1998. "L'imposition du français dans le département du Nord : de la Monarchie de Juillet au Second Empire." *Histoire de l'éducation*, 77.

Ravier, Joël. 2012. "L'évaluation des enseignants du primaire en France de 1835 à 1850. L'inspection primaire de Guizot et Falloux." *Spirale - Revue de recherches en éducation*, 49.

Reboul, Fabienne. 1991. *Guizot et l'instruction publique, in François Guizot et la culture politique de son temps, Colloque de la Fondation Guizot-Val Richer.* Paris:Seuil.

Rohner, Dominic, and Ekaterina Zhuravskaya. 2023. *Nation Building: Big Lessons from Successes and Failures.* London:CEPR Press.

Ronconi, Juan Pedro, and Diego Ramos-Toro. 2023. "Nation-Building Through Military Service." Working paper.

Rosanvallon, Pierre. 1985. *Le moment Guizot.* Paris:Gallimard.

Rosenberger, Lukas, and Sebastian Ottinger. 2023. "The American Origin of the French Revolution." Working paper.

Séguy, Jean. 1973. "La dialectométrie dans l'Atlas linguistique de la Gascogne." *Revue De Linguistique Romane*, , (1-24).

Site Mémoire des hommes. n.d.. "Portail culturel du Ministère des Armées, données téléchargées sur **www.memoiredeshommes.sga.defense.gouv.fr**, mise à jour du 21 juin 2021."

Smith, Anthony D. 1991. *National identity.* Vol. 11, University of Nevada press Reno.

Spolaore, Enrico, and Romain Wacziarg. 2021. "Fertility and Modernity." *The Economic Journal*, 132(642): 796–833.

Squicciarini, Mara P. 2020. "Devotion and Development: Religiosity, Education, and Economic Progress in Nineteenth-Century France." *American Economic Review*, 110(11): 3454–91.

Squicciarini, Mara P., and Nico Voigtländer. 2015. "Human Capital and Industrialization: Evidence from the Age of Enlightenment." *The Quarterly Journal of Economics*, 130(4): 1825–1883.

Squicciarini, Mara P., and Nico Voigtländer. 2016. "Knowledge Elites and Modernization: Evidence from Revolutionary France." Working paper.

Statistique Générale de la France. n.d.*a. La statistique industrielle de 1861 à 1896.* Atelier du Centre de recherches historiques, données traitées par l'INSEE, mis en ligne le 06 février 2011, consulté le 21 juin 2021.

Statistique Générale de la France. n.d.*b. L'enseignement primaire de 1829 à 1897.* Atelier du Centre de recherches historiques, données traitées par l'INSEE, mis en ligne le 07 février 2011, consulté le 21 juin 2021.

Statistique Générale de la France. n.d.*c. Les recensements de 1851 à 1921.* Atelier du Centre de recherches historiques, données traitées par l'INSEE, mis en ligne le 06 février 2011, consulté le 21 juin 2021.

Stendhal. 1838. *Mémoires d'un Touriste.* Paris:Ambroise Dupont.

Swadesh, Morris. 1952. "Lexico-statistic dating of prehistoric ethnic contacts: with special reference to North American Indians and Eskimos." *Proceedings of the American philosophical society*, 96(4): 452–463.

Tackett, Timothy. 1986. *Religion, Revolution, and Regional Culture in Eighteenth-Century France: The Ecclesiastical Oath of 1791.* Princeton:Princeton University Press.

Ther, Philipp. 2014. *The Dark Side of Nation-States: Ethnic Cleansing in Modern Europe.* New York City:Berghahn Books.

Tilly, Charles. 1975. *The Formation of National States in Western Europe.* Princeton:Princeton University Press.

Todd, Emmanuel. 1990. *L'invention de l'Europe.* Paris:Seuil.

Todd, Emmanuel, and Hervé Le Bras. 1981. *L'invention de la France. Atlas anthropologique et politique.* Paris:Gallimard.

Wantchekon, Leonard, Marko Klašnja, and Natalija Novta. 2015. "Education and Human Capital Externalities: Evidence from Colonial Benin *." *The Quarterly Journal of Economics*, 130(2): 703–757.

Weber, Eugen. 1976. *Peasants into Frenchmen: The Modernization of Rural France, 1870–1914.* Stanford:Stanford University Press.

Weinreich, Max. 1945. "Der YIVO un di problemen fun undzer tsayt." *YIVO Bleter*, 25(1): 1–13.

You, Yang. 2018. "Language Unification, Labor and Ideology." Working paper.

The online appendix for this paper can be found **here**.