French*

Guillaume Blanc[1] and Masahiro Kubo[2]

[1]The University of Manchester
[2]Brown University (**Job Market Paper**)

November 1, 2023

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 and ideology 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 homogenization. Additionally, we explore persistent impacts on identity and ideology in the twentieth century, documenting increased participation in the French Resistance and reduced collaboration with the Nazis, alongside an increase in votes supporting political centralization.

***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

* 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, N. Wolf, and G. Xu. We also thank seminar and conference participants at ASSA, Aix-Marseille, ASREC, Brown, Copenhagen, CREI, Cultural Evolution Society, Econometric Society, EIEF, ENS, Humboldt, Manchester, NES, NEUDC, PSE, Stellenbosch, UC3M, and UCLA Anderson 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.

Revolution, the French language, which was one of more than forty-six dialects and languages 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 the process of homogenization and nation-building. 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 in the formation of a national identity and ideology.

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.[1] The law mandated the building of schools for boys as well as major changes in the curriculum and in the training of schoolteachers in municipalities above five hundred inhabitants, with the teaching of the national language and a national history, the creation and distribution of textbooks approved by the state, and the creation of both a national system of teacher-training colleges and a body of nationally recruited school inspectors. Paul Lorain (1837), a close advisor of Guizot, argued at the time that "each school shall be a colony of French language in a conquered land" (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. In addition, we demonstrate that the regression discontinuity 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. To estimate the lower bound, we rely on the estimated percentage decrease resulting from the policy, relative to the counterfactual. We find 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 ma-

[1] Together with the *Loi du 21 Mars 1831 sur l'Organisation Municipale* (Municipal Law), which established the legal framework for the application of the *Loi Guizot* by expanding voting rights everywhere, and discouraging nepotism in municipalities above the threshold.

nipulation at the threshold and showing that observables are balanced at the threshold. Then, we show the robustness of the estimated coefficients to different methods of estimation, kernels, polynomial order, and bias correction; to using alternative bandwidths; to accounting for individual-level characteristics; and to using different distance metrics and measuring distance across the entire corpus of maps from the *Atlas* instead of only the Swadesh list. 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 that successful nation-building could only occur during the transition from stagnation to growth, as societies demanded education.

Finally, we document a persistent impact of state-sponsored education 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 policy increased the salience of national identity and preferences for centralization. Last but not least, we find suggestive evidence that the nation-building policy decreased nationalism, as captured by votes for the far right.

Our research contributes 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). Using a novel dataset on spoken languages and a range of different outcomes on identity and ideology, we empirically study one of the most extensively debated episodes of nation-building in the literature.

We also contribute to an emerging and vibrant literature on nation-building, which has shed light on the role of various factors such as education (Cinnirella and Schueler, 2018; Clots-Figueras and Masella, 2013), welfare policies (Caprettini and Voth, 2022), conflict (Alesina, Reich and Riboni, 2020; Dell and Querubin, 2018), population resettlement (Bazzi et al., 2019), military conscription (Ronconi and Ramos-Toro, 2023), inter-group contact (Cáceres-Delpiano et al., 2021), charismatic leaders (Assouad, 2021), and propaganda, media, or shared experiences (Blouin and Mukand, 2019; Depetris-Chauvin, Durante and Campante, 2020; Hara, 2022; Kersting and Wolf, 2021; Li, 2022).[2] Additionally, we contribute to a literature on the effects of education on economic outcomes (Duflo, 2001, 2004) and externalities, cultural transmission, or ideology (Algan, Cahuc and Shleifer, 2013; Cantoni et al., 2017; Friedman et al., 2016; Lott, 1999; Wantchekon, Klašnja and Novta, 2015). In the context of the *Loi Guizot*, Montalbo (2021*b*) finds a positive impact on fiscal capacity. Using a natural experiment, we estimate the causal effect of state-sponsored education on the adoption of a common language, the building of a national identity, and political outcomes.

[2]Fouka (2022); Rohner and Zhuravskaya (2023) provide an extensive overview of this growing literature. Additionally, a number of studies focus on state incentives and the role of education (Alesina, Giuliano and Reich, 2021; Paglayan, 2022), military threats (Aghion et al., 2019), and immigration (Bandiera et al., 2019).

Last but not least, a number of studies have illuminated the phenomenon of minority backlash against policies of forced assimilation (Bazzi, Hilmy and Marx, 2021; Carvalho, Koyama and Williams, 2022; Dehdari and Gehring, 2022; Fouka, 2020; Marciante, 2023). We document assimilation and explore the factors that made state-sponsored education successful.[3]

2 Historical Background

Lucien Febvre, founder of the *Annales* school, 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.[4] The *lenga d'òc* (*langues 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.[5]

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 previously regarded as a dialect of *oc*. Although nearly extinct today, this language family hold 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) (Alighieri, 1305, p. 19). Indeed, Occitan is the only foreign language used in Dante's *Divine Comedy*, which established the Florentine dialect as the standard Italian language. 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).[6] The Nobel Prize in Literature was even awarded in 1904 to Occitan poet Frédéric Mistral, in recognition of his contributions to the preservation and promotion of Provençal literature and culture. He

[3] Additionally, there are also a number of literatures, on assimilation, identity, or language, that we contribute to. For assimilation and migrations, 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.

[4] 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 thoughtful discussion of the diversity of norms and cultures in France.

[5] 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).

[6] For example, 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).

remains the only laureate to have been honored for work in a language not officially recognized by any state.[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 (over 80 percent of the total 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 French language.[8] In 1838, Stendhal, the 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 Weber (1976) and Lodge (1993), the French language was not widely adopted until the early twentieth century, with mass state-sponsored education and the development of railways.

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

[7]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 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.

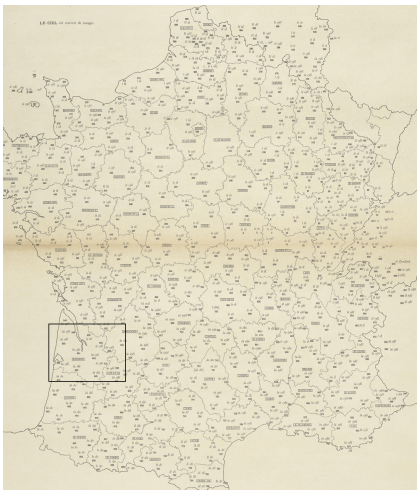
[9]The *Atlas* also contains information on sixty-two additional towns in neighboring countries, including Italy, Switzerland, and Germany. We did not digitize the data for these towns.

carried out by linguist Jules Gilliéron from 1897 to 1901 (henceforth "1900") to study Romance languages in rural parts of the country, where the majority of the population lived and did not speak French.

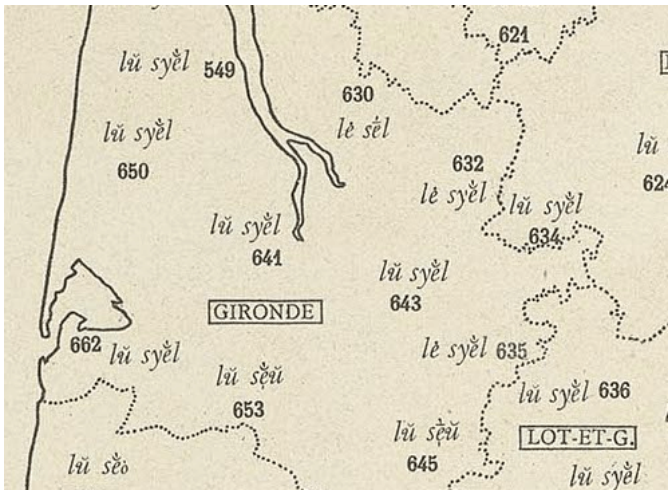
For four years, Gilliéron's assistant Edmond Edmont traveled across the country and asked locals for the standard, common pronunciation of 1,920 words or expressions in their municipality. The survey captures spoken language, including vocabulary and pronunciation, and paints a particularly detailed picture of the words used by ordinary people in their daily life at the time. Indeed, importantly, the Atlas was conducted to capture the common language spoken in the surveyed municipality and not the language spoken by the individuals surveyed (Gilliéron and Edmont, 1902; Le Dû, Le Berre and Brun-Trigaud, 2005).[10]

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

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 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

[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 (Gilliéron and Edmont, 1902; Le Dû, Le Berre and Brun-Trigaud, 2005). Consistent with this, Appendix Tables A7 and A8 find that individual-level characteristics were irrelevant.

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; and the large size of the corpus allows for substantial heterogeneity in their characteristics. 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 100x80 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 common language spoken at the time. We observe substantial heterogeneity, with both persistent 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 mini-

[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**.
[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.

mal 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," "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]

[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.

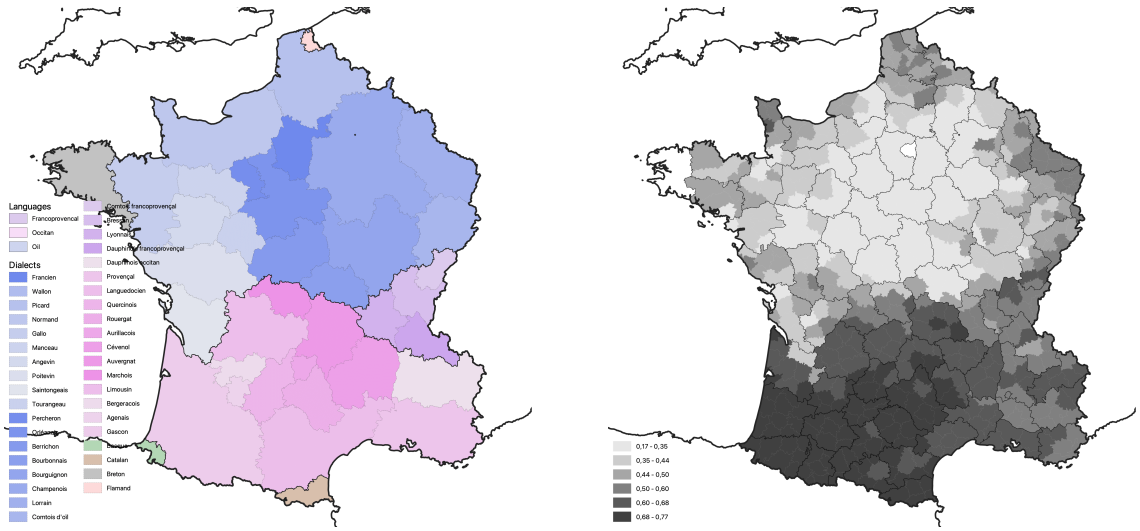
[15] We also measure linguistic distance across the total of 166,176 pairs of municipalities, but we focus in the paper on the linguistic distance from French.

[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] We plot the distribution of linguistic distance from French in Appendix Figure A9. 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 (forthcoming) 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.

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)

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 A17 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]

[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.

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

Finally, we note that, 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-Lorraine. 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 regression discontinuity results to zero because historical population is the running variable. Last, we focus on rural towns because of the better quality and representativeness of the *Atlas* there and because it allows us to bracket the fact that the language spoken and the schools in urban settings were likely very different from those in rural places, both before and after the *Loi Guizot*.[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). Among municipalities with available data at the time, while some level of education was offered in approximately two-thirds of those with around five hundred inhabitants, only one-third had schools, as illustrated in Appendix Figure A13. The classrooms were only open part of the year, sometimes only for a couple of months, with irregular schedules and without trained teachers. According to Meyers (1976), the curriculum placed heavy emphasis on religious and moral education, and 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).[21]

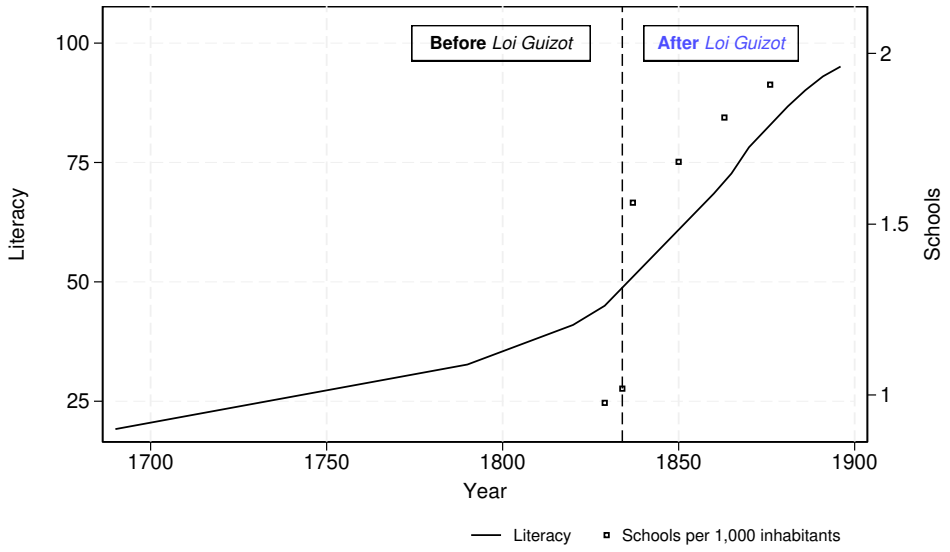
[19] Out of the 577 municipalities in France, 552 are within the parts of mainland France that were not annexed at any point in the nineteenth century; 510 had a non-missing population at the time of the law; 479 were rural, with less than five thousand inhabitants; and 375 have not been merged or split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

[20] In practice, including urban municipalities in the sample does not change our results significantly since we compare municipalities near the threshold. It only (marginally) affects the selection of the optimal bandwidth.

[21] For additional details, the Guizot Survey (1833) used in Montalbo (2021*a*) is available at **inrp.fr/she/guizot/**.

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 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 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).[22] 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 fund its provisions, especially to support the construction of school buildings and hiring of schoolteachers.[23]

[22]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 Guizot Law in Section 5.5.

[23]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

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).[24] Public schools came to account for more than 90 percent of schools on the eve of the *Lois Ferry* of 1881–82 (Appendix Figure A11), and students who could not afford tuition had to be fully funded (Appendix Figure A12). 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 A10). 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).[25] 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.

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 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).[26] 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

(*centimes*).

[24]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 *Loi Falloux* and *Loi Duruy*, which extended the provisions of the *Loi Guizot* to schools for girls at the eight-hundred-inhabitant threshold and five-hundred-inhabitant threshold, respectively.

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

[26]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.

information on recommended teaching methods.[27]

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 ou t 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.[28]

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 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).[29] Their salary improved significantly with the creation of a minimum annual salary that increased sixfold by 1880.[30] According to Meyers (1976), on the eve of the *Lois Ferry*, 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

[27]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).

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

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

[30]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).

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 formation 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 either the linguistic distance from French in 1900 in town 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 Guizot Law and 0 otherwise, capturing state-sponsored education, or state intervention in the provision of education; $f(population_i)$ is a polynomial controlling for smooth 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.[31]

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.

Nature of the treatment. Importantly, we estimate the local, intention-to-treat 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.[32]

[31] We follow Montalbo (2021*b*) and use population in 1836, when the law took effect and most of the schools were built—between 1835 to 1837 (Figure 3). 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 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 A11, we also show that there was no jump in linguistic distance at the five-hundred-inhabitant threshold using the population in other years.

[32] There is no municipality-level data available on schools built in the aftermath of the law. 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,

Although we lack data to empirically evaluate the relative importance of each dimension, we argue that the standardization of the curriculum and of the training of schoolteachers were likely the most important ones (Cantoni et al., 2017). Indeed, the prohibition of nepotism by the Municipal Law is unlikely to directly affect linguistic homogenization, while education was already provided in most municipalities at and around the threshold before the Guizot Law (Appendix Figure A13). Additionally, using data from the Guizot Survey (1833), we also find that, among municipalities with fewer than a thousand inhabitants that offered education prior to the law, 63 percent included religious instruction in their curriculum, while 43 percent included instruction on reading and writing.[33] However, only 3 percent included geography in their curriculum, and a mere 1 percent included history, while the Guizot Law mandated the teaching of French history and geography in all schools.

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. We show that observable geographic, climatic, institutional, cultural, and economic factors vary smoothly at the threshold, 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. Figure 4 plots linguistic distance, predicted linearly using these covariates of interest, against population—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.[34]

Finally, additional threats to identification may come from sorting and migration. Appendix Figure A15 plots the density of the population of municipalities in France and in the *Atlas*. 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.[35] Indeed, municipalities were required to maintain and annually update a population register after 1831. According to Biraben (1963), the lists were then requested by prefects to enable the central

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.

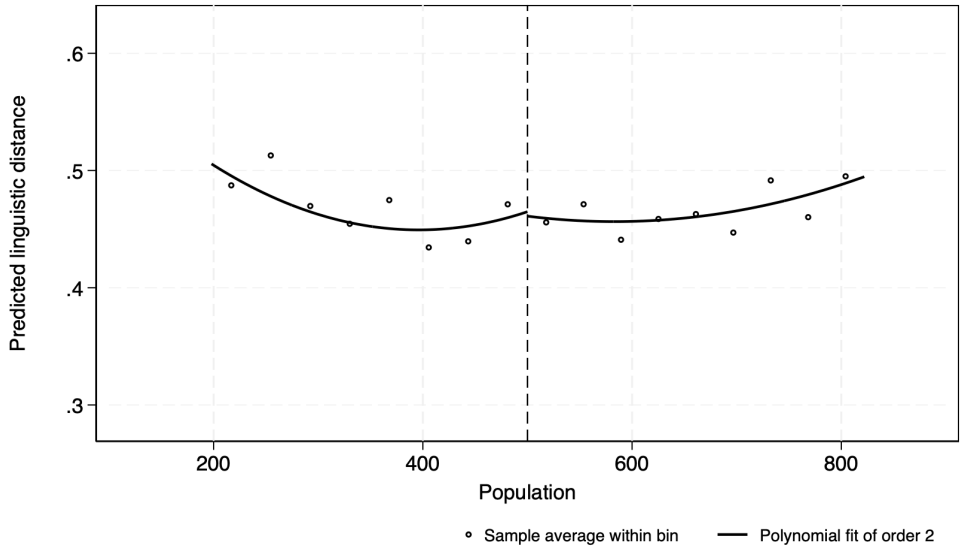
[33] In line with this, Figure 3 does not show a significant or discontinuous increase in literacy in the immediate aftermath of the Guizot Law.

[34] 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 A14, which displays the spatial distribution of towns in the *Atlas* that are near (above or below), or away from the threshold.

[35] 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—instead, negative sorting would bias our results toward zero. 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).

Figure 4: Balance in observables at the threshold: CEF of predicted linguistic distance using a set of geographic, climatic, cultural, and economic covariates

Note: This figure plots predicted linguistic distance from French against population around the discontinuity introduced by the policy. 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 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.



administration to conduct thorough verifications.[36] Migration itself is only an issue if the building of schools led to in-migration, since out-migration would bias our results toward zero. There is no data on migration flows at the municipality level, yet we argue it is very unlikely to be an important factor. Appendix Figure A16 displays evidence supporting this: there was no jump in the population of treated municipalities after the policy. Additionally, because we estimate a local effect, and because of gravity forces, we expect more out-migration than in-migration from municipalities with five hundred inhabitants.

5 The Homogenization of Language

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

[36] In 1836, when most Guizot schools were built, this was incorporated into law (Biraben, 1963, p. 309).

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.[37] 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, hence confounding the effect of the treatment when using OLS. 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, suggesting we might be capturing a causal effect.

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 5. 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 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.[38]

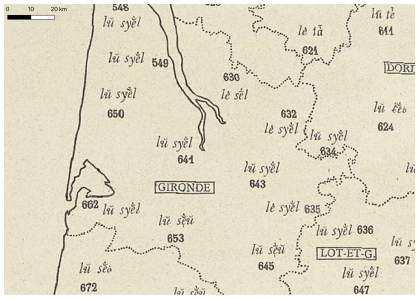
Regression discontinuity results. We rely on the discontinuity introduced by the *Loi Guizot* to study the effect of state-sponsored education on the spread of the French language in Table 2. In column (1), we report results from controlling for a zero-th degree polynomial in population and compare the average linguistic distance with and without state intervention

[37] 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, 2020, 2021). Finally, it is very likely that both representativeness and selection into the sample change at higher levels of population (see Section 3.1).

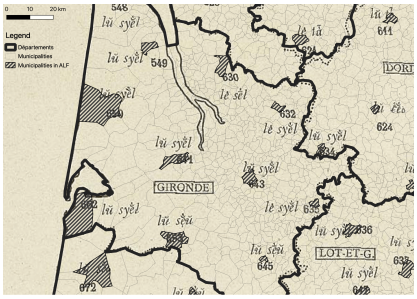
[38] When the *Loi Guizot* was passed, there were 446 inhabitants in Saint-Côme and 801 in Vélines.

Figure 5: Case study in Aquitaine

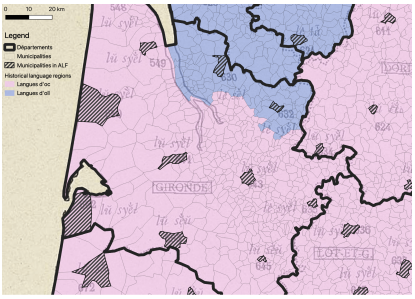
Note: This figure constitutes a case study of the historical region of Aquitaine and is designed to understand our data collection (Panels A and B), the historical framework (Panel C), the provision of state-sponsored education (Panel D), our measure of linguistic distance (Panel E), and our regression 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*.



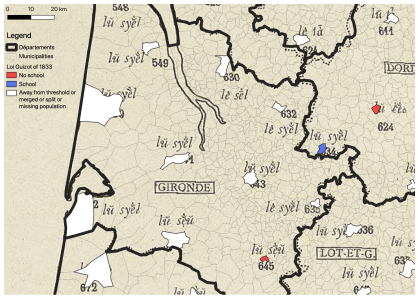
(Panel A) Map of *ciel* ("sky")



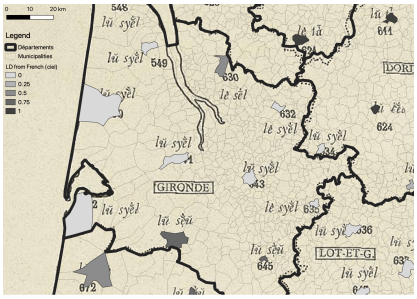
(Panel B) With municipalities



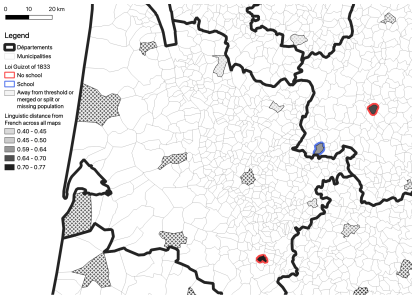
(Panel C) With historical language regions



(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

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, quadratic functions of population. We report the MSE optimal bandwidth and effective number of observations throughout.

We find that the provision of state-sponsored education significantly decreased linguistic distance from French by 0.1 at the threshold. The effect is statistically significant at the 1 percent level across specifications and is particularly large. 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.

We display our results in Figure 6, which plots linguistic distance from French against population after accounting for both distance from Paris and *département* and historical-dialect fixed effects.[39] In order to account for pre-existing differences in linguistic distance, and to capture regional institutional and cultural differences, we present estimates of the effect of state-sponsored education on linguistic distance after accounting for covariates in

[39] We include the full set of controls from Appendix Table A4. Note that we only include fixed effects for *départements* and dialects with more than four observations in the sample.

Table 2: State-sponsored education and 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

Appendix Table A4. We sequentially add covariates. As predicted by the fact that the RD framework allows us to control for these factors, which we expect vary smoothly at the cutoff, the results remain unchanged, while our estimates gain in precision.

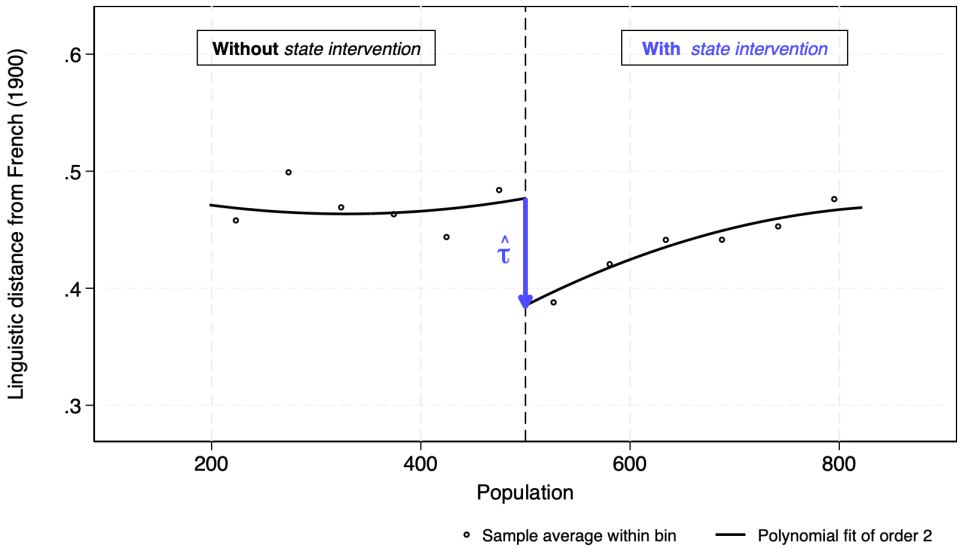
Rate of homogenization. The primary goal of nation-building is to homogenize populations. However, we only observe linguistic distance in a single cross-section and lack temporal variation. Below, we show that our regression discontinuity design allows us to recover key parameters of interests, the change over time and the rate of change over time, under a weak assumption.[40] 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.[41]

[40] 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.

[41] 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

Figure 6: State-sponsored education and linguistic distance from French

Note: This figure plots linguistic distance from French against population 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.



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 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.[42]

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.

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.

[42] 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.

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). 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.[43]

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 (2021); Carvalho, Koyama and Williams (2022); 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.

5.3 Robustness and placebo

We perform a range of standard robustness and placebo analyses in Appendix 4. We show the robustness of the estimated coefficients to (a) different methods of estimation, kernels, and bias correction in Appendix Table A5; (b) alternative, non-optimal bandwidths in Appendix Table A6; (c) accounting for individual-level characteristics in Appendix Tables A7 and A8; and (d) dropping municipalities near the national border to account for bilingualism in Appendix Table A9. We also show, in placebo tests, that (e) only the five-hundred-inhabitant threshold returns statistically significant estimates (Appendix Table A10); and (f) only the five-hundred-inhabitant threshold based on population at the time of the Guizot Law consistently returns a coefficient statistically significant at the 1 percent level across polynomial degrees, and not for population at any other point in time, suggesting that our results predominantly capture the effect of the Guizot Law, and not that of the Municipal Law or other policies such as the Duruy or the Ferry Laws (Appendix Table A11).[44] Finally, our results are robust to the use of different distance metrics and to measuring distance across the entire corpus of maps from the *Atlas* instead of only the 50 words from the Swadesh list (Appendix Table A12).

[43] 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).

[44] When population is measured at the time of the French Revolution (in 1793), after the Guizot Law (in 1841), at the time of the Duruy Law, which also had a five-hundred-inhabitant threshold for schools for girls (in 1886), at the time of the Ferry Laws (in 1881), or at the time of the *Atlas* (in 1901), we do not find any statistically significant discontinuity at the threshold. There is a small negative discontinuity at the threshold based on population at the time of the Municipal Law (in 1831), but this is only statistically significant at the 10 percent level when using a polynomial of degree 0. This could be explained by the fact that, while most schools were built between 1835 and 1837 (Figure 3), some schools were built immediately after the Guizot Law was passed, in 1833 and 1834. Therefore, using the threshold based on population at the time of the Municipal Law also captures the effect of the Guizot Law.

5.4 Discussion and interpretation

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.[45] 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*, 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 A13, 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.

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

[45] 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.

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*. Figure 7 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.

Figure 7: Share of linguistic distance from French accounted for by language

Note: This figure displays the share of linguistic distance accounted for by language (excluding phonetic accents), plotted against linguistic distance from French, including accents. Linguistic distance from French is defined as the Levenshtein distance from Le Plessis-Robinson. Observations are at the municipality level.

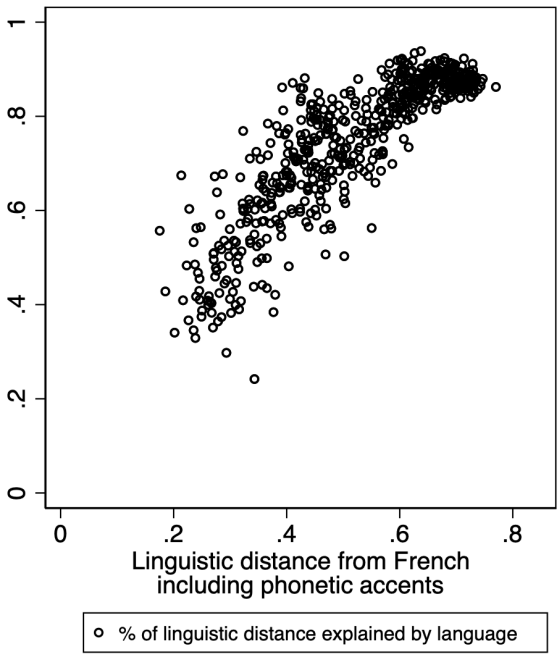


Table 3 reports estimates of the effect of state-sponsored education on linguistic distance including and excluding phonetic accents, in odd and even columns respectively. Our estimates show that about three-fourths of the effect of state-sponsored education on linguistic distance from French comes from 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* region in the north (p. 190). We find supportive empirical evidence in Appendix Table A14, 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 A15, 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.

Table 3: State-sponsored education and linguistic distance from French: 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

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 an initially restricted 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 Guizot Law. 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).[46] Second, local municipal councils voted for the building of schools and arranged for the collection of additional resources in accordance with the law.[47] Below, we exploit variation in the distribution of local elites to understand their role in homogenization.

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).[48] Darnton (1973) (p. 1350) discusses the occupations of 137 subscribers in the city of Besançon. The vast majority were political 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. Yet, one could be concerned that we may not be capturing the effect of elites but rather that of income.[49] Figure 8, Panel B, shows that the effect of state-sponsored education was the same across all levels of development, which alleviates this concern.

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.[50] Appendix Figure A21 shows that state-sponsored education had a larger effect in the historical *langues d'oc* region, consistent with our findings in Section 5.4. 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.

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. The Civil Constitution of the Clergy, passed in July 1790, required clergy members to take an oath of allegiance to the secular state (Tackett,

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

[47] 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 Guizot Law on literacy. See also footnote 35.

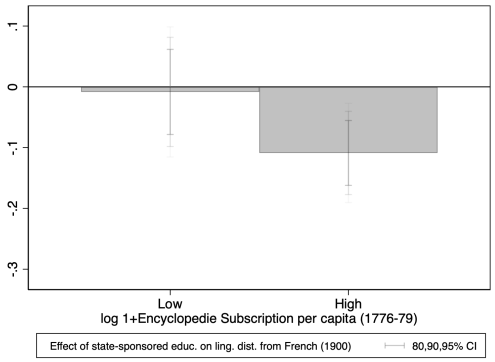
[48] We use a municipality-level measure of all *Encyclopedie* subscriptions per capita within a fifty-kilometer radius of each municipality, while Squicciarini and Voigtländer (2015) use a *département*-level measure and discard rural towns with subscriptions from their sample.

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

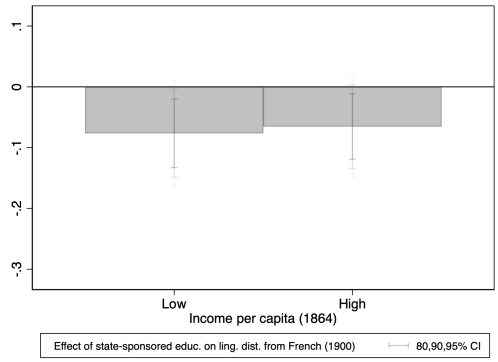
[50] For example, Alesina, Giuliano and Reich (2021) argue that the threat of democratization is the strongest determinant of homogenization. The threat was particularly high at the time, following the French Revolution of 1789 and the July Revolution of 1830, and before the February Revolution of 1848.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

Figure 8: Heterogeneous effect of state-sponsored education

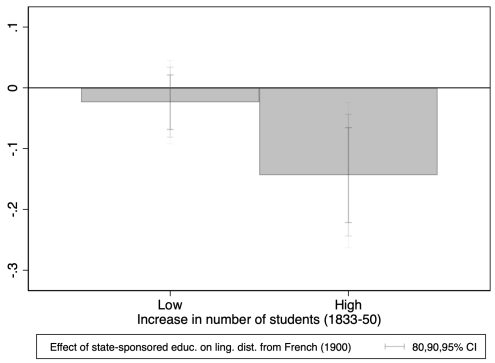
Note: This figure displays the effect of state-sponsored education on linguistic distance from French—at low and high levels of presence of presence of local elites (Panel A), economic development (Panel B), educational demand (Panel C), and returns to education (Panel D)—using the discontinuity introduced 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.



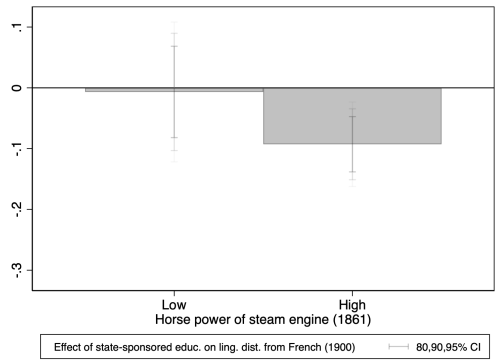
(Panel A) Local elites



(Panel B) Economic development



(Panel C) Educational demand



(Panel D) Returns to education

1986). The presence of refractory clergy, who did not take the oath, is commonly used as a proxy for the devotion of the population at the time of the French Revolution (Blanc, 2021; Squicciarini, 2020). 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).

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 Guizot Law

increased the supply of state-sponsored education, it did not make schooling compulsory. The fact that populations were not coerced into attending schools could have contributed, in itself, to the success of the policy. It also allows us to understand heterogeneity.

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.[51]

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. This is consistent with the hypothesis of Anderson (1983); Gellner (1983).

6 The Persistent Effects of Nation-Building

This section explores the persistent effects of nation-building on society. We provide evidence suggesting that the adoption of a standard language promoted migration and trade, and we document a persistent effect on the salience of national identity and preferences for the centralization of political authority.

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

[51]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.

the district or at the *département* level. We estimate Equation 3 with ordinary least squares.[52]

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

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 of migration or trade flows from administrative unit $d(j)$ to unit $d(i)$.[53] 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

[52] 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.

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

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.[54] Daudin (2010) provides data on 88 receiving districts and 551 supplier districts; less than 20 percent of pairs were linked by trade.[55] We match 67 supplier and 396 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. We explore the persistent effects of state intervention in the provision of education, together with the adoption of the French language, on the salience of national identity and ideology. Henceforth, we continue to use the term 'state-sponsored education' to refer to this bundled treatment. Alternatively, we use the term 'nation-building' to refer to both the nation-building policy and the process of nation-building.

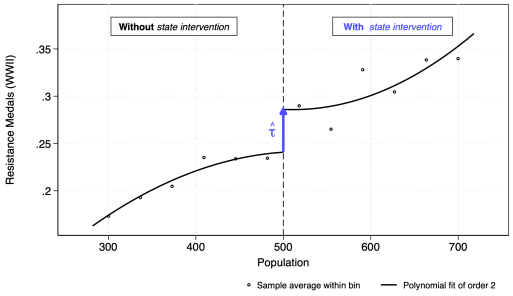
Figure 9, Panel A, shows that state-sponsored education and the adoption of the French language significantly increased the likelihood that individuals born in treated municipalities were heroes of the French Resistance and were awarded a Médaille de la Résistance for documented "remarkable acts of faith 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)

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

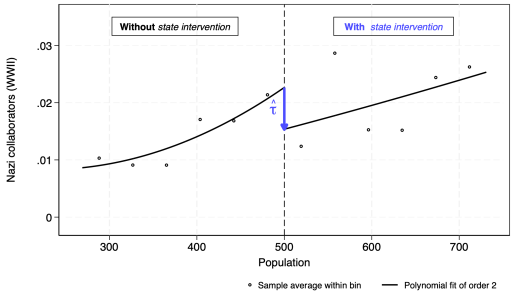
[55] When we aggregate the existence of links across pairs of *départements*, we find that 50 percent were linked by trade. The coefficient on linguistic distance is larger at this level, likely because there are fewer zeros in the data.

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

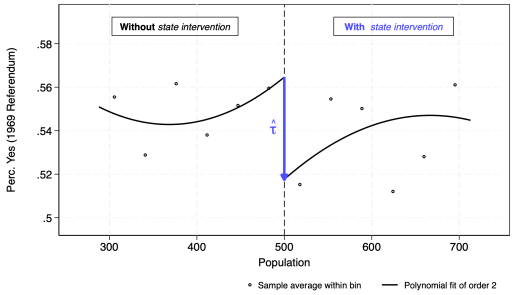
Note: This figure plots measures of the salience of national identity against population around the discontinuity introduced by the policy. Each point plots the average value within a bin, partialled out of the full set of controls used in Appendix Tables A17 (Panel A) and A22 (Panel B). In Panel A, 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 6. 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

fixed effects to account for local institutional and cultural factors.[56]

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.

[56] 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 to account for differences in the reporting of acts of resistance, we improve the precision of the estimator.

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.[57] First, in Appendix Table A24, we find negative effects on votes for the far right, with Jean-Marie and Marine Le Pen in the

[57] 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.

2002 and 2017 presidential elections. While the results are not statistically significant for Jean-Marie Le Pen, they are statistically different from zero 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 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 formation of a national identity in the process of nation-building in a fragmented society. We find that the adoption of the French language was indeed "essential to the concept of France" (Hobsbawm, 1990, p. 60). Our findings resonate with the statements expressed by French President Emmanuel Macron, who emphasized that "The French language builds the unity of the nation," and that "The spread of the French language in our regions and throughout the world, in our colonies, also occurred through coercion" (Discours à l'occasion de l'inauguration de la Cité internationale de la langue française, October 30, 2023, own translation).

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.

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. 2021. "Islam and the State: Religious Education in the Age of Mass Schooling." 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. 2020. "Demographic Change and Development from Crowdsourced Genealogies in Early Modern Europe." Working paper.

Blanc, Guillaume. 2021. "The Cultural Origins of the Demographic Transition in France." Working paper.

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. 2022. "Resisting Education." GMU Working Paper in Economics No. 22-28, 2022.

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."

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.

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.

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.

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. forthcoming. "Fertility and Modernity." *The Economic Journal.*

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.

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

Online appendix to: French

Appendix 1– Additional Details on The Measurement of Linguistic Distance

A1.1 Data construction

As we explain in Section 3, we use the fifty maps/words from the *Atlas* corresponding to the Swadesh list. In the process of digitization, we extract all the phonetic symbols from the fifty maps (Appendix Figure A1) and assign a unique code to a symbol. Each code consists of an English alphabet letter and the number. For example, code starting with *a* is assigned to the phonetic symbol similar with *a*. Every time we find a new symbol from a map, we keep it in the record and assign the next number within a corresponding English alphabet letter. Since we work simultaneously or because of a quality of scans or human careless checks, in the end, we have some duplicates. Therefore, we replace them by a single unique code in the data.

Appendix Figure A2 is an example (*étoile*) in raw data after the above digitization process. In the dataset, we combine the information on the names of *départements* and the identification number of a town to each coded letters, Below, we will explain how to measure Levenshtein Distance with the letters in the dataset.

A1.2 Levenshtein distance (LD)

In this section, we give an example about measuring LD for *étoile*. First,we show an example for a general case and then an corresponding example to our code schemes.

General example for *LD*: *etoile* and *estelo*

To simplify the example, we replace the e-acute (*é*) by *e*. By insertion, substitution and deletion, we need four edits from *etoile* to *estelo*, following the below procedures. To normalize the distance, we divide it by the maximum of the number of letters. In this example, it is six. Hence: LD=4/6.

Procedures

1. etoile → estoile: insertion of *s* before *t*
2. estoile → esteile: substitution of *o* for *e*
3. esteile → estele: deletion of *i*
4. estele → estelo: substitution of *e* for *o*

Example for our coding scheme: /e5/t/w/a11/l/ and /e5/s/t/e10/l/o22/

Next, we will show an corresponding example using our code expressions to *etoile* and *estelo* in the previous example. The phonetic expression *etoile* spoken in Seine is coded as /e5/t/w/a11/l/ and the one *etoile* spoken in Var is coded as /e5/s/t/e10/l/o22/ in our database. We use a slash (/) to separate each code of a letter from another. Hence, the number of /e5/t/w/a11/l/ is five and that of /e5/s/t/e10/l/o22/ is six. Again, by insertion, substitution and deletion, we need four edits from /e5/t/w/a11/l/ to /e5/s/t/e10/l/o22/,

following the below procedures. To normalize the distance, we divide it by the maximum of the number of letters, six. Hence: LD=4/6.

Procedures

1. /e5/t/w/a11/l/ → /e5/s/t/w/a11/l/: insertion of *s* before *t*
2. /e5/s/t/w/a11/l/ → /e5/s/t/w/a11/l/o22/: insertion of *o22* after *l*
3. /e5/s/t/w/a11/l/o22/ → /e5/s/t/a11/l/o22/: deletion of *w*
4. /e5/s/t/a11/l/o22/ → /e5/s/t/e10/l/o22/: substitution of *a11* for *e10*

Appendix 2– Additional Details on July Monarchy Laws

A2.1 Municipal Law

The Municipal Law of March 21, 1831, introduced a series of pivotal provisions that significantly promoted democratization in local governance (Agulhon, 1970; Guionnet, 1996, 1998; Tudesq, 1972, 1982). We provide an overview of the provisions of the law in Section 4.1, and we detail and discuss their impact below.

LOI

Du 21 mars 1831

SUR L'ORGANISATION MUNICIPALE.

TITRE Ier.

DU CORPS MUNICIPAL.

CHAPITRE Ier.

De la composition du corps municipal.

ART. 1er. Le corps municipal de chaque commune se compose du maire, de ses adjoints et des conseillers municipaux.

Les fonctions des maires, des adjoints et des autres membres du corps municipal sont essentiellement gratuites, et ne peuvent donner lieu à aucune indemnité ni frais de représentation.

ART. 2. Il y aura un seul adjoint dans les communes de deux mille cinq cents habitans et au dessous; deux dans celles de deux mille cinq cents à dix mille habitans; et dans les communes d'une population supérieure, un adjoint de plus par chaque excédant de vingt mille habitans. (*Loi du 28 pluviose an VIII, art. 12.*)

Lorsque la mer ou quelque autre obstacle rend difficiles, dangereuses, ou momentanément impossibles, les communications entre le chef-lieu et une portion de commune, un adjoint spécial, pris parmi les habitans de cette fraction, est nommé en sus du nombre ordinaire, et remplit les fonctions d'officier de l'état civil dans cette partie détachée de la commune (1).

ART. 3. Les maires et les adjoints sont nommés par le roi, ou en son nom par le préfet.

Dans les communes qui ont trois mille habitans et au-dessus, ils sont nommés par le roi, ainsi que dans les chefs-lieux d'arrondissement, quelle que soit la population.

Les maires et les adjoints seront choisis parmi les membres du conseil municipal, et ne cesseront pas pour cela d'en faire partie.

Ils peuvent être suspendus par un arrêté du préfet; mais ils ne sont révocables que par une ordonnance du roi.

ART. 4. Les maires et les adjoints sont nommés pour trois ans; ils doivent être âgés de 25 ans accomplis.

Ils doivent avoir leur domicile réel dans la commune.

ART. 5. En cas d'absence ou d'empêchement, le maire est remplacé par l'adjoint disponible, le premier dans l'ordre des nominations.

En cas d'absence ou d'empêchement du maire et des adjoints, le maire est remplacé par le conseiller municipal, le premier dans l'ordre du tableau, lequel sera dressé suivant le nombre des suffrages obtenus.

ART. 6. Ne peuvent être ni maires ni adjoints :

1° Les membres des cours et tribunaux de première instance et des justices de paix;

2° Les ministres des cultes;

3° Les militaires employés des armées de terre et de mer en activité de service ou en disponibilité;

4° Les ingénieurs des ponts et chaussées et des mines en activité de service;

5° Les agens et employés des administrations financières et des forêts;

6° Les fonctionnaires et employés des collèges communaux et les instituteurs primaires;

7° Les commissaires et agens de police.

ART. 7. Néanmoins, les juges suppléans aux tribunaux de première instance et les suppléans des juges de paix peuvent être maires ou adjoints.

Les agens salariés du maire ne peuvent être les adjoints.

ART. 8. Il y a incompatibilité entre les fonctions de maire et d'adjoint et le service de la garde nationale (2).

CHAPITRE II.

Des conseils municipaux.

SECTION PREMIÈRE.—*De la composition des Conseils municipaux.*

ART. 9. Chaque commune a un conseil municipal composé, y compris le maire et adjoints,

De dix membres, dans les communes de cinq cents habitans et au dessous;

De douze dans celles de cinq cents à quinze cents;

De seize, dans celles de quinze cents à deux mille cinq cents;

De vingt-un, dans celles de deux mille cinq cents à trois mille cinq cents;

De vingt-trois, dans celles de trois mille cinq cents à dix mille;

De vingt-sept, dans celles de dix mille à trente mille;

Et de trente-six, dans celles d'une population de trente mille ames et au dessus.

Dans les communes où il y aura plus de trois adjoints, le conseil municipal sera augmenté d'un nombre de membres égal à celui des adjoints au dessus de trois.

Dans celles où il aura été nommé un ou plusieurs adjoints spéciaux et supplémentaires, en vertu du second paragraphe de l'article 2 de la présente loi, le conseil municipal sera également augmenté d'un nombre égal à celui de ses adjoints (3).

ART. 10. Les conseillers municipaux sont élus par l'assemblée des électeurs communaux.

ART. 11. Sont appelés à cette assemblée: 1° les citoyens les plus imposés aux rôles des contributions directes de la commune, âgés de vingt-un ans accomplis, dans les proportions suivantes :

Pour les communes de mille ames et au dessous, un nombre égal au dixième de la population de la commune;

Ce nombre s'accroîtra de cinq par cent habitans en sus de mille jusqu'à cinq mille;

De quatre par cent habitans en sus de cinq mille jusqu'à quinze mille;

De trois par cent habitans au dessus de quinze mille;

2° Les membres des cours et tribunaux, les juges de paix et leurs suppléans;

Les membres des chambres de commerce, des conseils de manufactures, des conseils de prud'hommes;

Les membres des commissions administratives des collèges, des hospices et des bureaux de bienfaisance (4);

Les officiers de la garde nationale (5);

Les membres et correspondans de l'Institut; les membres des sociétés savantes instituées ou autorisées par une loi;

Les docteurs de l'une ou de plusieurs des facultés de droit, de médecine, des sciences, des lettres, après trois ans de domicile réel dans la commune;

Les avocats inscrits au tableau, les avoués près les cours et tribunaux, les notaires, les licenciés de l'une des Facultés de droit, des sciences, des lettres, chargés de l'enseignement de quelqu'une des matières appartenant à la Faculté où ils auront pris leur licence, les uns et les autres après cinq ans d'exercice et de domicile réel dans la commune;

Les anciens fonctionnaires de l'ordre administratif et judiciaire, jouissant d'une pension de retraite;

Les employés des administrations civiles et militaires jouissant d'une pension de retraite de six cents francs et au-dessus;

Les élèves de l'École Polytechnique qui ont été, à leur sortie, déclarés admis ou admissibles dans les services publics, après deux ans de domicile réel dans la commune; toutefois les officiers appelés à jouir du droit électoral en qualité d'anciens élèves de l'École Polytechnique ne pourront l'exercer dans les communes où ils se trouveront en garnison qu'autant qu'ils y auraient acquis leur domicile civil ou politique avant de faire partie de la garnison;

Les officiers de terre et de mer jouissant d'une pension de retraite;

Les citoyens appelés à voter aux élections des membres de la chambre des députés ou des conseils généraux des départemens, quel que soit le taux de leurs contributions dans la commune.

ART. 12. Le nombre des électeurs domiciliés dans la commune ne pourra être moindre de trente, sauf le cas où il ne se trouverait pas un nombre suffisant de citoyens payant une contribution personnelle.

ART. 13. Les citoyens qualifiés pour voter dans l'assemblée des électeurs communaux, conformément au paragraphe 2 de l'article 11, et qui seraient en même temps inscrits sur la liste des plus imposés, voteront en cette dernière qualité.

ART. 14. Le tiers de la contribution du domaine exploité par un fermier à prix d'argent ou à portion de fruits lui est compté pour être inscrit sur la liste des plus imposés de la commune, sans diminution des droits du propriétaire du domaine (6).

ART. 15. Les membres du conseil municipal seront tous choisis sur la liste des électeurs communaux, et les trois quarts, au moins, parmi les électeurs domiciliés dans la commune (7).

ART. 16. Les deux tiers des conseillers municipaux sont nécessairement choisis parmi les électeurs désignés au paragraphe 1er de l'article 11; l'autre tiers peut être choisi parmi tous les citoyens ayant droit de voter dans l'assemblée, en vertu de l'article 11.

ART. 17. Les conseillers municipaux doivent être âgés de vingt-cinq ans accomplis. Ils sont élus pour six ans et toujours rééligibles (8).

Les conseils seront renouvelés par moitié tous les trois ans.

ART. 18. Les préfets, sous-préfets, secrétaires généraux et conseillers de préfecture, les ministres des divers cultes en exercice dans la commune, les comptables des revenus communaux (9) et tout agent salarié par la commune (10), ne peuvent être membres des conseils municipaux. Nul ne peut être membre de deux conseils municipaux (11).

ART. 19. Tout membre d'un conseil municipal dont les droits civiques auraient été suspendus, ou qui en aurait perdu la jouissance, cessera d'en faire partie, et ne pourra être réélu que lorsqu'il aura recouvré les droits dont il aura été privé (12).

ART. 20. Dans les communes de cinq cents ames et au-dessus, les parens au degré de père, de fils, de frère, et les alliés au même degré, ne peuvent être en même temps membres du même conseil municipal (13).

ART. 21. Toutes les dispositions des lois précédentes, concernant les incompatibilités et empêchemens des fonctions municipales, sont abrogées.

ART. 22. En cas de vacance dans l'intervalle des élections triennales, il devra être procédé au remplacement de que le conseil municipal se trouvera réduit aux trois quarts de ses membres.

SECTION II.—*Des assemblées des Conseils municipaux.*

ART. 23. Les conseils municipaux se réunissent quatre fois l'année, au commencement des mois de février, mai, août et novembre. Chaque session peut durer dix jours.

ART. 24. Le préfet ou sous-préfet prescrit la convocation extraordinaire du conseil municipal, ou l'autorise sur la demande du maire, toutes les fois que les intérêts de la commune l'exigent.

Dans les sessions ordinaires, le conseil municipal peut s'occuper de toutes les matières qui rentrent dans ses attributions.

En cas de réunion extraordinaire, il ne peut s'occuper que des objets pour lesquels il a été spécialement convoqué.

La convocation pourra également être autorisée pour un objet spécial et déterminé, sur la demande du tiers des membres du conseil municipal adressée directement au préfet, qui ne pourra la refuser que par un arrêté motivé, qui sera notifié aux réclamans, et dont ils pourront appeler au roi.

Le maire préside le conseil municipal; les fonctions de secrétaire sont remplies par un de ses membres, nommé au scrutin et à la majorité, à l'ouverture de chaque session.

ART. 25. Le conseil municipal ne peut délibérer que lorsque la majorité des membres en exercice assiste au conseil.

Il ne pourra être refusé à aucun des citoyens contribuables de la commune communication, sans déplacement, des délibérations des conseils municipaux.

ART. 26. Le préfet déclarera démissionnaire tout membre d'un conseil municipal qui aura manqué à trois convocations consécutives sans motifs reconnus légitimes par le conseil.

ART. 27. La dissolution des conseils municipaux peut être prononcée par le roi.

L'ordonnance de dissolution fixera l'époque de la réélection.

(*Voir la suite d'autre part.*)

(1) Ord. en conseil d'État du 7 août 1835.
(2) Arr. de la C. de cass. des 9 mars, 1er juin, 24 août et 1er septembre 1832.
(3) Ord. en conseil d'état du 24 octobre 1832.
(4) Arr. de la C. de cass. du 25 février 1832. —Ord. en cons. d'état du 14 août 1837.
(5) Ord. en cons. d'ét. des 14 novembre 1824; 19 octobre 1837; 8 février 1838.
(6) Arr. de la C. de cass. du 25 février 1833.
(7) Ord. en cons. d'ét. des 30 mai 1834 et 23 juillet 1835.
(8) Ord. en cons. d'état du 20 novembre 1832.
(9) Ord. en cons. d'ét. du 19 août 1832.—Arr. de la C. de cass. du 11 avril 1821.
(10) Décision minist. du 29 avril 1833.—Arr. de cass. des 19 août 1835 et 18 mai 1837.
(11) Ord. en cons. d'ét. du 9 mars 1836.
(12) Arr. de cass. des 9 juillet 1822; 25 août 1824; 14 août 1835; 6 août 1836.—Ord. en cons. d'état, 18 septembre 1821.
(13) Ord. en cons. d'état des 26 février et 23 avril 1832; 3 mai 1833; 31 octobre 1835; 8 janvier, 9 mars 1836; 2 janvier, 1838.

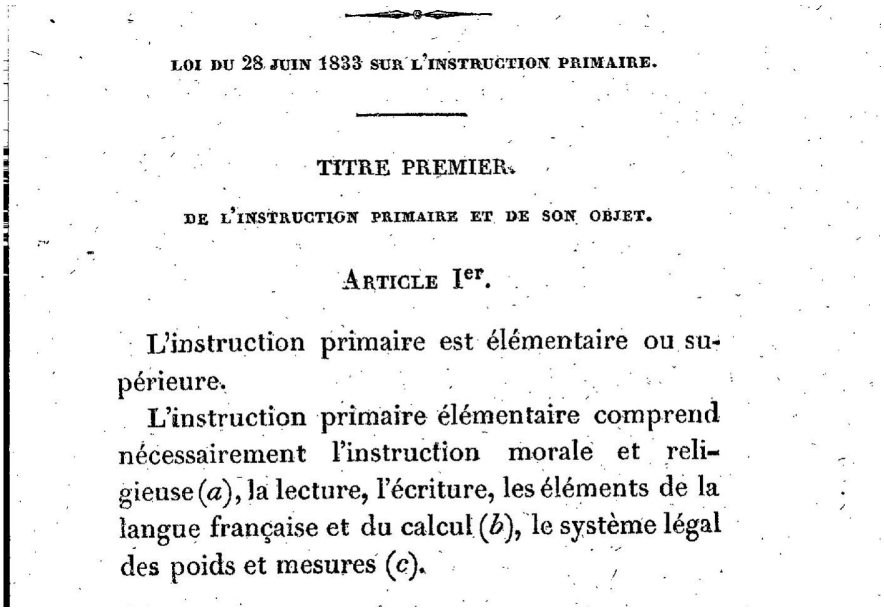
The law broadened political participation and representation at the local level. Before the law, mayors in municipalities below 3,000 inhabitants were appointed by the King. Following the enactment of the law (Article 3), the mayor was to be chosen by the prefect among the members of the municipal council. It called for the creation of municipal councils, composed of 10 members in municipalities with populations below 500 inhabitants and 12 members in

municipalities with populations between 500 and 1,500 (Article 9). Councilors were elected for 6-year terms through censitary suffrage, where only the wealthiest could vote. In municipalities with populations between 300 and 1,000, where there were previously no voting rights, 10 percent of the population was granted the ability to vote (Articles 10, 11, 12, 17). The law also mandated that elections for half of the municipal council members be held every 3 years (Article 17).

Additionally, to prevent nepotism in municipalities with 500 inhabitants and above, the law introduced a provision to prevent the concentration of power within the same family. Article 20 stipulated that "parents to the degree of father, son, brother, and relatives to the same degree cannot be members of the same municipal council simultaneously." According to Tudesq (1982), "On May 20, 1835, the prefect of Aube reported the existence of 300 municipalities with fewer than 500 inhabitants out of the 446 in his department: 'The voters themselves appoint members of the municipal council, rarely do we find those who can read and write; they are all relatives or allies'" (p. 218).

A2.2 *Loi Guizot*

The *Loi Guizot* of June 28, 1833 laid the foundations of mass state-sponsored education in France. Section 4.1 discusses general considerations as well as some of the important provisions and impacts of the law regarding standardization of the curriculum and of the training of schoolteachers. We discuss below the application of the five-hundred-inhabitant threshold, which is not in the text of the law.



LOI DU 28 JUIN 1833 SUR L'INSTRUCTION PRIMAIRE.

TITRE PREMIER.

DE L'INSTRUCTION PRIMAIRE ET DE SON OBJET.

ARTICLE Ier.

L'instruction primaire est élémentaire ou supérieure.

L'instruction primaire élémentaire comprend nécessairement l'instruction morale et religieuse (*a*), la lecture, l'écriture, les éléments de la langue française et du calcul (*b*), le système légal des poids et mesures (*c*).

Furet and Ozouf (1977) argue that the population threshold was introduced in subsequent decrees (p. 136). Additionally, some of the most important textbooks on the history of education in France (e.g. Furet and Ozouf, 1977; Chevallier, Grosperrin and Maillet, 1968), as well as a large number of books and articles (e.g. Blanc and Wacziarg, 2020; Montalbo, 2021*a*,*b*;

Archives Nationales, n.d.; Diebolt, Jaoul-Grammare and Perrin, 2022; French Government, n.d.; García-Peñalosa and Bignon, 2021), mention the application of the threshold.[58] Gevaert (2017) argues that the *Loi Guizot* was largely based on the Ordinance of February 19, 1816 and ultimately on the *Projet de loi sur les écoles civiques inférieures* of 1798 in Switzerland, which also had a five-hundred-inhabitant threshold. Similarly, Chevallier, Grosperrin and Maillet (1968) also argues that Guizot "maintained the regime of the Ordinance of February 19, 1816" (p. 71).

Additionally, the *Loi Guizot*, which mentions "every commune", could have applied to municipalities with a population above five hundred inhabitants for institutional reasons. Indeed, the Municipal Law of March 21, 1831 introduced additional differences at the threshold that influenced the application of the *Loi Guizot*, as discussed in Section 4.1. For example, O'Connor (1955) argues that "it is not always clear when Guizot uses the term commune whether he is referring to the various classes of towns, as he often does, or only to those which possessed charters, for although he frequently speaks of the latter as *communes proprement dites* or *véritables communes*, at times he refers to them simply as communes" (p. 64). Montalbo (2021*b*) provides strong evidence for a discontinuity in the provision of state-sponsored education at the threshold, based on population at the time of the *Loi Guizot*, by showing that local taxes increased following its application.

[58]Such a threshold was commonly used in other countries in Europe at the time. For example, Azar and Espuelas (2021) argue that Spain's Moyano Law of 1857, which had a five-hundred-inhabitant threshold, was largely inspired by the *Loi Guizot*.

Appendix 3– Estimation of the Speed of Homogenization

A3.1 Effect on change in outcome over time

Let us define $y_{i,year}$ the linguistic distance between town i and standard French in some year. By definition, $y_{i,year}$ takes values between 0 and 1. As already defined in the paper, $population_i$ is the population of town i at the time of the law and $state\_educ_i$ is the following deterministic and discontinuous function of $population_i$:

$$state\_educ_i = \begin{cases} 1 & if\ population_i \geq 500 \\ 0 & if\ population_i < 500 \end{cases}$$

The standard RD result follows from the assumption of continuity of conditional regression functions, which states that $\mathbb{E}(y_{i,1900}(1)|population_i)$ and $\mathbb{E}(y_{i,1900}(0)|population_i)$ are continuous in 500. The average treatment effect is:

$$\tau = \mathbb{E}(y_{i,1900}(1) - y_{i,1900}(0)|500) = \mathbb{E}(y_{i,1900}|500) - \mathbb{E}(y_{i,1900}|500^-)$$

Now, let us consider some $t \leq 1833$. Since potential outcomes are, by definition, equal before the policy is realized ($y_{i,t}(1) = y_{i,t}(0) = y_{i,t}$), the average treatment effect on the change in outcome over time is exactly the same as the average treatment effect on outcome levels (with $\Delta y_{i,1900} \equiv y_{i,1900} - y_{i,t}$):

$$\mathbb{E}(\Delta y_{i,1900}(1) - \Delta y_{i,1900}(0)|500) = \mathbb{E}(y_{i,1900}(1) - y_{i,1900}(0)|500) - \underbrace{\mathbb{E}(y_{i,t}(1) - y_{i,t}(0)|500)}_{=0} = \tau$$

Hence, the effect of state-sponsored education on the change in language over time is identified from the cross-sectional regression in 1900.

A3.2 Estimation of rate of homogenization

We want to estimate the parameter $|\boldsymbol{g}|$, the rate of linguistic homogenization from 1833 to 1900, at the threshold, defined as:

$$\boldsymbol{g} \equiv \frac{\mathbb{E}(y_{i,1900}|500)}{\mathbb{E}(y_{i,1833}|500)} - 1 = \frac{\alpha_{1900} + \tau}{\alpha_{1833}} - 1$$

with $\alpha_{1900} \equiv \mathbb{E}(y_{i,1900}|500^-)$ and $\alpha_{1833} \equiv \mathbb{E}(y_{i,1833}|500)$. Yet, we do not observe linguistic distance before the policy at the threshold (α_{1833}).

Proposition 1 in the main text of the paper constructs a lower bound to the rate of homogenization that can be estimated with our cross-sectional data: under Assumptions 1 and 2

below, $|\tau|/\alpha_{1900}$ is a lower bound of $|\boldsymbol{g}|$, and it is itself bounded below by 0:

$$0 \leq |\tau|/\alpha_{1900} \leq |\boldsymbol{g}|$$

Assumption 1. $\tau \leq 0$

Assumption 2. $\alpha_{1833} \geq \alpha_{1900}$

Assumption 1 says that state-sponsored education decreases linguistic distance from French.

Assumption 2 says that, absent the policy, homogenization would still have taken place at the threshold. Because of the continuity of the conditional expectation of counterfactual outcomes in the running variable (main regression discontinuity assumption), this is equivalent to assuming that homogenization took place below the threshold. The assumption is motivated by historical evidence, in particular from Lodge (1993); Weber (1976) who document linguistic homogenization in the long run, starting from the sixteenth century. Moreover, from the start of the French Revolution to the end of the nineteenth century, no policies were put in place to promote diversity in France.

Together with Assumption 1, Assumption 2 therefore assumes that the counterfactual change over time was of the same sign as the effect of treatment in 1900.

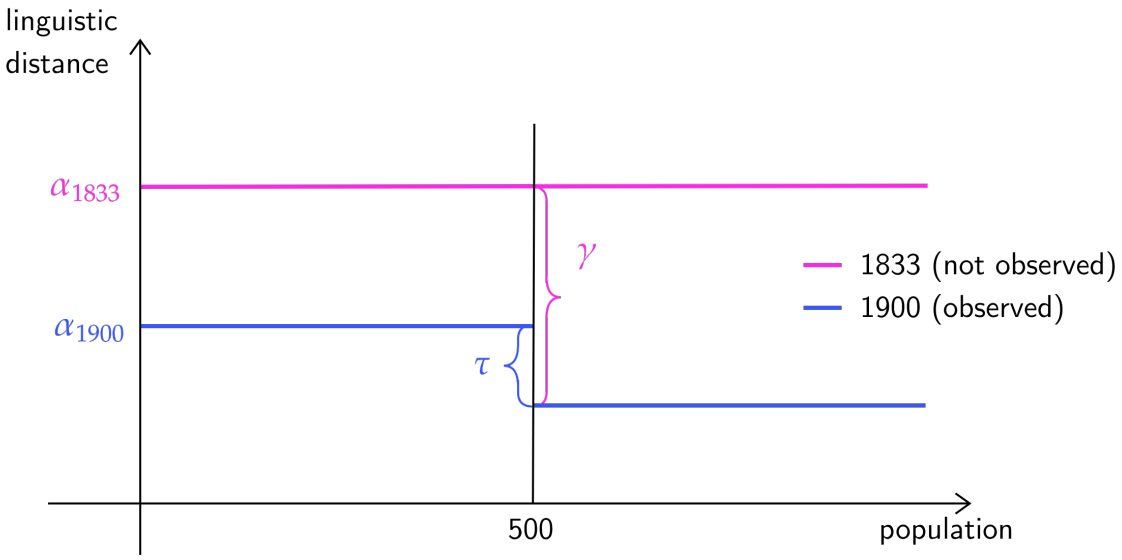
Proof of Proposition 1. The first inequality, $0 \leq |\tau|/\alpha_{1900}$, follows directly from Assumption 1. The second inequality, $|\tau|/\alpha_{1900} \leq |\boldsymbol{g}|$, follows from Assumption 2. In both cases, we also use the fact that linguistic distance takes values between zero and one.

Finally, if there was no secular change at the threshold without the policy (that is, if the inequality in Assumption 2 is binding), then $\boldsymbol{g} = \tau/\alpha_{1900}$ (Corollary 1) and we get an unbiased estimate of $|\boldsymbol{g}|$:

Corollary 1. *If* $\alpha_{1833} = \alpha_{1900}$*, then* $|\tau|/\alpha_{1900} = |\boldsymbol{g}|$

According to Corollary 1, if there was no secular change from 1833 to 1900 at the 500 threshold other than that caused by state-sponsored education, our estimate of $|\tau|/\alpha_{1900}$ will be an unbiased estimate of $|\boldsymbol{g}|$. The assumption that $\alpha_{1833} = \alpha_{1900}$, although unlikely in most contexts, may be relevant in our context. First, cultural traits are usually very persistent. Moreover, Weber (1976) argues that the factors that played the most important role on the homogenization of language in France, besides Guizot, were the *Lois Ferrys* of 1881-2 and the *plan Freycinet* of 1878. The effect of those two policies only materialized much later, in the twentieth century (see Section 2 for more details).

The Figure below provides the intuition for the estimation of the rate of homogenization. Here, $\boldsymbol{g}$ is defined as γ/α_{1833}, with $\gamma \equiv \alpha_{1900} + \tau - \alpha_{1833}$. It is easy to show that, under Assumption 1, $|\tau|/\alpha_{1900} \leq |\tau|/\alpha_{1833} \leq |\gamma|/\alpha_{1833}$.



$$|\mathbf{g}| \equiv |\gamma| \,/\, \alpha_{1833}$$
$$0 \leq |\tau| \,/\, \alpha_{1900} \leq |\mathbf{g}|$$

Appendix 4– Additional Details on Robustness and Placebo Tests

Method of estimation, kernel, and bias-correction. Appendix Table A5 reports estimates of τ using different methods. We first report the baseline coefficient, which relies on a triangular kernel and follows Hahn, Todd and Van der Klaauw (2001). This is the conventional regression discontinuity estimate used throughout the paper. For comparability purposes, the results for the other methods of estimation rely on the optimal bandwidths chosen with the conventional method. Then, we use a uniform kernel, which is equivalent to using ordinary least squares to estimate the RD coefficient. Finally, because mean-squared-error optimal bandwidths are often too large, resulting in biased RD estimates, we report the bias-corrected estimator proposed by Calonico, Cattaneo and Titiunik (2014). We also report robust, bias-corrected standard errors. The estimated coefficients are stable across methods of estimation and, as expected, the bias-corrected coefficients are slightly larger.

Alternative bandwidths. We show that the results are robust to using a range of non-mean-squared-error optimal bandwidths in Appendix Table A6. We consider observations in a 200, 300, 366, 400, and 500 inhabitants bandwidth of the threshold. The least populated town in the *Atlas* has 134 inhabitants, hence the 366 inhabitants-bandwidth is the largest implementable symmetric bandwidth. For comparability purposes, we hold the bandwidth on the left fixed at 366 inhabitants. We find that the zero-th degree polynomial specifications, which are equivalent to mean comparison, yield smaller and less significant results with larger bandwidths. This is likely the result of the impossibility to control for trends and smooth functions of population with a polynomial of order zero. The results are virtually unchanged and the effect of state-sponsored education increases with the bandwidth with linear and cubic specifications.

Alternative (placebo) cutoffs. In Appendix Table A10, we report estimates of τ using placebo thresholds for the school building discontinuity, at 400, 450, 500, 550 and 600 inhabitants. For comparability purposes, we include all municipalities on the left of threshold. The results in column (3) correspond to those in column (3) of Appendix Table A6. We find that only this column, which relies on the five-hundred-inhabitant threshold of the *Loi Guizot*, yields negative and statistically significant results—none of the other (placebo) thresholds yield a statistically significant result.

Alternative (placebo) year of measurement of population. The discontinuity introduced by the *Loi Guizot* was based on the population at the time—as already discussed, the law went into effect from 1835 to 1837 for the most part. Appendix Table A11 reports estimates of τ relying on the 500 inhabitants-discontinuity in population measured at different, placebo points in time: at the time of the French Revolution in 1793, at the time of the law in 1836, at the time of the *Loi Duruy* in 1866, at the time of the *Lois Ferry* in 1881, and at the time of the *Atlas* in 1901. In column (2), we display the baseline coefficient for the effect of state-sponsored education, based on the threshold in 1836. For comparability purposes, the results on population in other years rely on the optimal bandwidths chosen with the threshold

in 1836. We find that only the discontinuity at the time of the law mattered, none of the other (placebo) years of measurement of the population yield significant coefficients. Interestingly, column (3), based on population in 1866, also yields large and negative coefficients, although not significant at conventional levels. We argue that it is likely because of the *Loi Duruy*, which extended the provisions of the Falloux and *Loi Guizot*s to girls by requiring municipalities above 500 inhabitants to open a school for girls as well. The results are likely not significant because the effect could not fully materialize by 1900—the *Atlas* surveyed spoken language too early to capture the effect of that law. Additionally, the Church's influence increased after the Falloux law, potentially accounting for the absence of discernible impact: "clerical influence in French education increased after 1850 because of the large number of Catholic schools which were establish" (Huckaby, 1965, p. 205).

Accounting for individual-level characteristics. The pronunciation in the *Atlas* is supposed to be representative of the language spoken by the population in each municipality. Yet, since Gilliéron and Edmont (1902-1910) collected data on respondents of the survey, we can also estimate the effect of state-sponsored education on linguistic distance from French assuming that the language spoken is representative of the individuals who were surveyed rather than of their municipality. Appendix Tables A7 and A8 report individual-level regressions accounting for age and gender, with standard errors clustered at the municipality level. The results are virtually unaffected.

Accounting for bilingualism. We expect that places near the border would be more likely to be bilingual. This would complicate the interpretation of our results in terms of language as proxy of identity since those around the national border could possibly speak the French language as *lingua franca*. Appendix Table A9 reports the results of the effect of state-sponsored education in the full sample of municipalities and in the restricted sample of municipalities more than 25 kilometers away from the border. The results are virtually unaffected apart from a loss of statistical significant resulting from the lower number of observations.

Distance metrics and words used to measure distance. In the main results of the paper, we use representative words from the Swadesh list and apply a Levenshtein distance algorithm to capture the language spoken in each municipality and its similarity with the language spoken near Paris. Appendix Table A12 shows that our results are robust to the use of different distance metrics (Relative Identity Value and Weighted Identity Value, see Section 3.1) and to measuring distance across the entire corpus of maps from the *Atlas* instead of only the 50 words from the Swadesh list.

Appendix 5– Supplementary Figures

Figure A1: Phonetic transcription

Note: This figure displays the phonetic symbols found in the *Atlas* and their corresponding transcription. Each phonetic symbol is assigned a unique transcription (code). When we find ex post that a code has a duplicate, we flag the duplicates by coloring them and we replace them by a single unique code in the data.

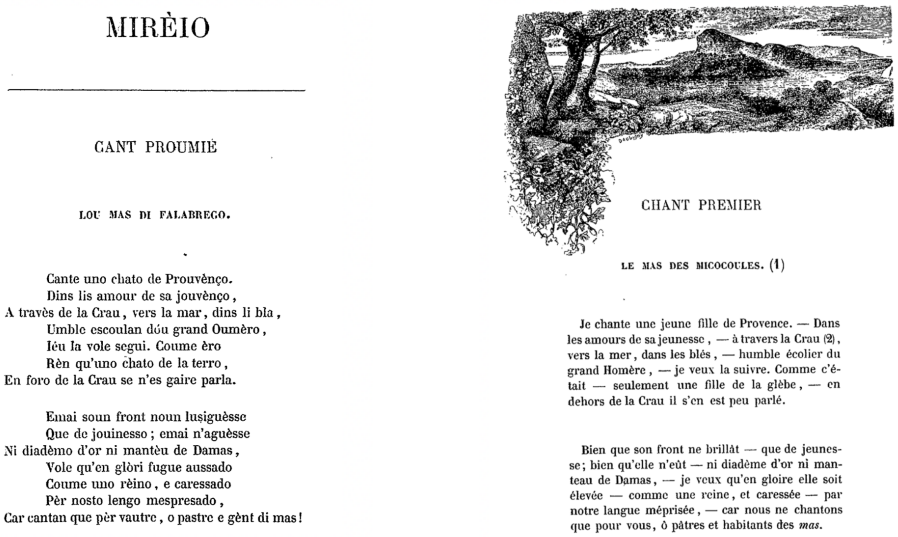
Figure A2: Raw data for *étoile* ("star")

Note: This figure displays an example of our raw data for the map of *étoile* ("star"), with the phonetic transcription of the pronunciation of the word for some of the municipalities in Figure A7.

530	873	Bouches-d	/e5/s/t/e16/l/o22/	
531	882	Bouches-d	/e5/s/t/e10/l/o22/	
532	883	Bouches-d	/e5/s/t/e16/r/o22/	
533	884	Var	/e5/s/t/e10/l/o22/	
534	886	Var	/e5/s/t/e10/l/o22/	
535	893	Var	/e5/s/t/e16/l/o22/	
536	894	Var	/e5/s/t/e10/l/o22/	

Figure A3: Mirèio (1859), Frédéric Mistral

Note: This figure displays the first page of the poem "Mirèio" in Provençal (Panel A) and in French (Panel B). Source: gallica.bnf.fr/ark:/12148/bpt6k7490v.



MIRÈIO

CANT PROUMIÈ

LOU MAS DI FALABREGO.

Cante uno chato de Prouvènço.
Dins lis amour de sa jouvènço,
A travès de la Crau, vers la mar, dins li bla,
Umble escoulan dóu grand Oumèro,
Iéu la vole segui. Coume èro
Rèn qu'uno chato de la terro,
En foro de la Crau se n'es gaire parla.

Emai soun front noun lusiguèsse
Que de jouinesso; emai n'aguèsse
Ni diadèmo d'or ni mantèu de Damas,
Vole qu'en glòri fugue aussado
Coume uno rèino, e caressado
Pèr nosto lengo mespresado,
Car cantan que pèr vautre, o pastre e gènt di mas!

CHANT PREMIER

LE MAS DES MICOCOULES. (1)

Je chante une jeune fille de Provence. — Dans les amours de sa jeunesse, — à travers la Crau (2), vers la mer, dans les blés, — humble écolier du grand Homère, — je veux la suivre. Comme c'était — seulement une fille de la glèbe, — en dehors de la Crau il s'en est peu parlé.

Bien que son front ne brillât — que de jeunesse; bien qu'elle n'eût — ni diadème d'or ni manteau de Damas, — je veux qu'en gloire elle soit élevée — comme une reine, et caressée — par notre langue méprisée, — car nous ne chantons que pour vous, ô pâtres et habitants des *mas*.

(Panel A) Provençal **(Panel B)** French

Figure A4: Street names

Note: This figure displays street names in Alsacien (Panel A) and Provençal (Panel B) dialects. Sources: Wikipedia.

(Panel A) Alsacien **(Panel B)** Provençal

Figure A5: Towns in sample

Note: This figure displays the spatial distribution of municipalities surveyed in the *Atlas Linguistique de la France*.

Figure A6: Distance to closest town in the *Atlas*

Note: This figure displays the distribution of distance to the closest town (in km) across municipalities in the *Atlas Linguistique de la France*.

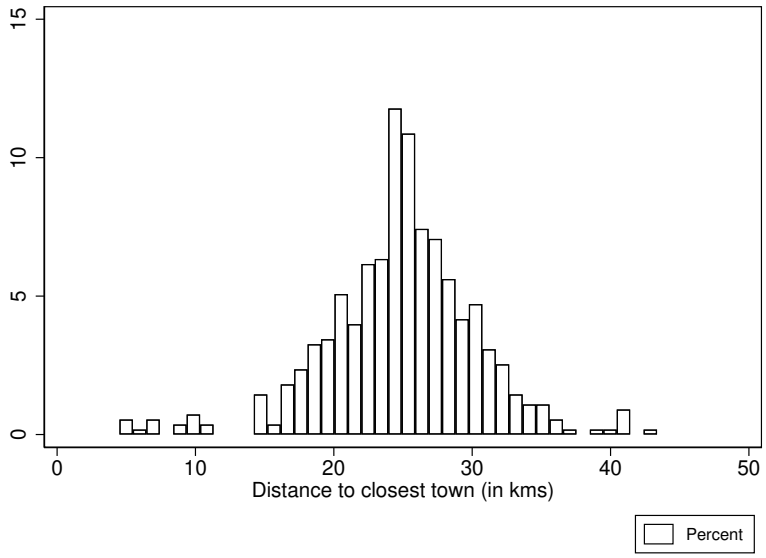
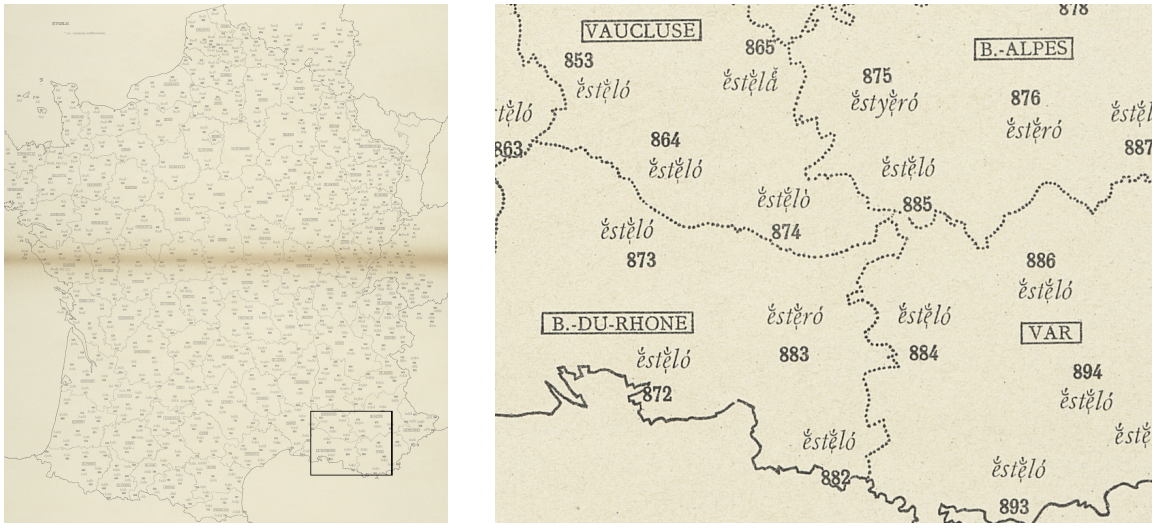


Figure A7: Map of *étoile* ("star")

Note: This figure displays a map from the *Atlas Linguistique de la France* showing the pronunciation of *étoile* ("star") across all municipalities (Panel A) and in the southern part of France only (Panel B). Appendix Figure A2 displays how we digitize the phonetic symbols, using the system of transcription displayed in Appendix Figure A1.



(Panel A) France **(Panel B)** Zoomed in

Figure A8: Cladistic distance (Spolaore and Wacziarg, 2021)

Note: This figure displays cladistic distance from French—the number of nodes separating a language from French, using the classification in the *Ethnologue*—at the *département* level in France. Source: Spolaore and Wacziarg (forthcoming).

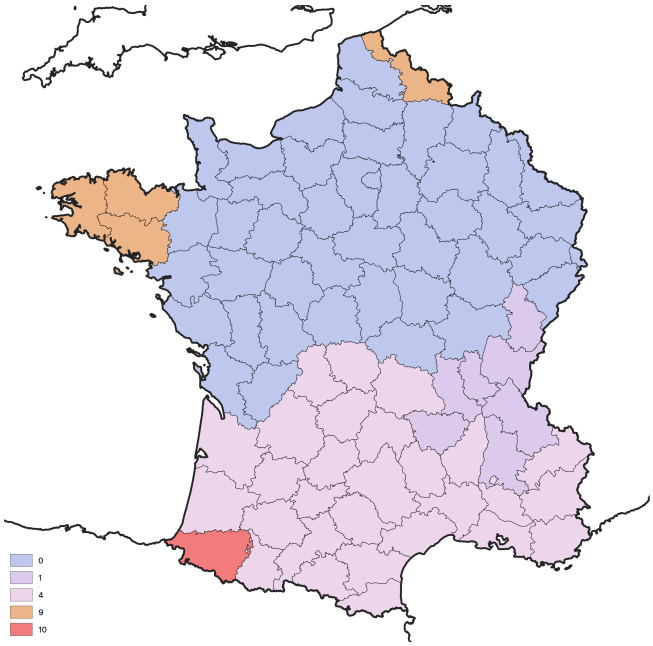


Figure A9: Distribution of linguistic distance from French

Note: This figure displays the distribution of linguistic distance from French across municipalities in the *Atlas Linguistique de la France*. Linguistic distance from French is defined as the Levenshtein distance from Le Plessis-Robinson.

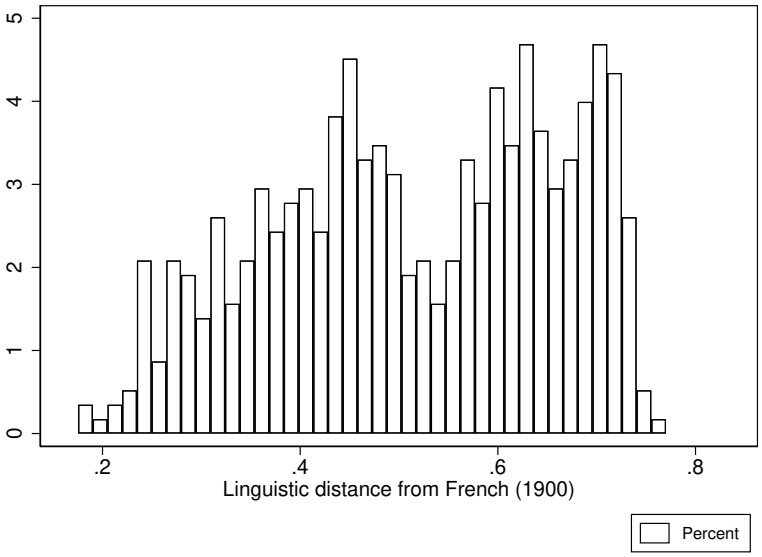


Figure A10: Students per ten thousand inhabitants

Note: This figure displays the average number of students per capita (ten thousand inhabitants) in France over time. Source: Statistique Générale de la France (n.d.*b*).

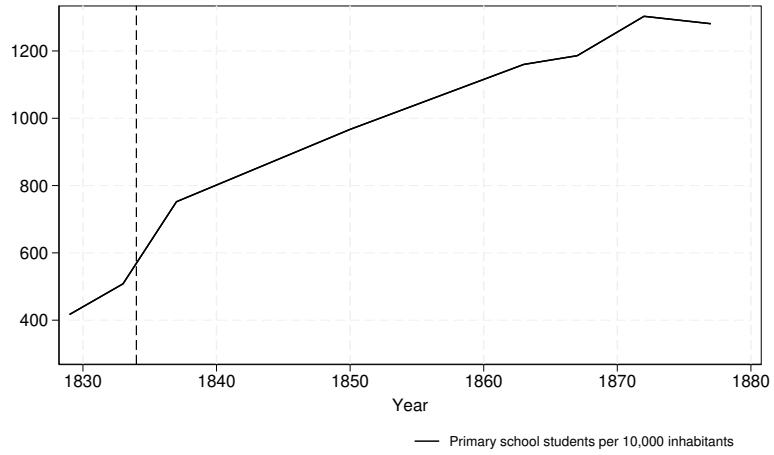


Figure A11: Share of public primary schools

Note: This figure displays the share of public primary schools in France over time. Source: Statistique Générale de la France (n.d.*b*).

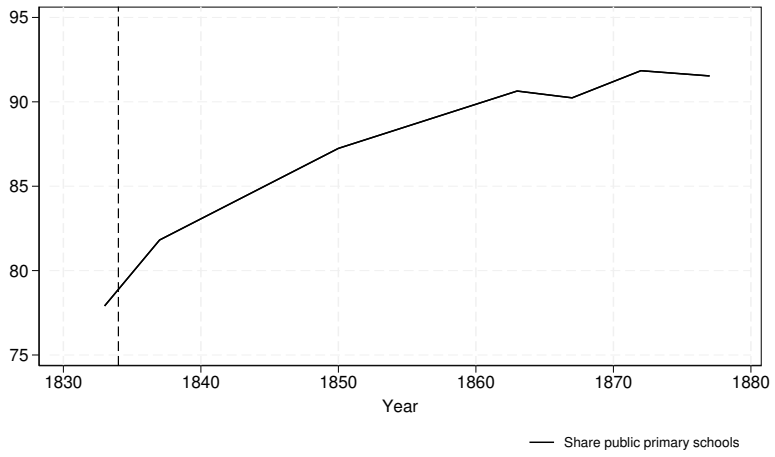


Figure A12: Share of fully funded primary school students

Note: This figure displays the share of fully funded primary school students in France over time. Source: Statistique Générale de la France (n.d.*b*).

Figure A13: Schools before and after the *Loi Guizot*

Note: This figure displays the share of municipalities with primary education and those with a school building (or classroom) before the *Loi Guizot* (at the time of the Guizot survey in 1833), available at **www.inrp.fr/she/guizot**. Observations are at the municipality level, and twenty-two *départements* are included in the data. Source: Montalbo (2021*a*).

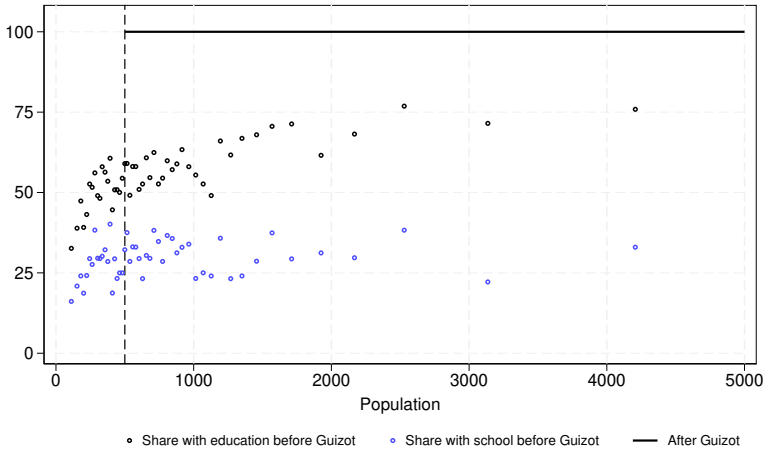


Figure A14: Towns in sample around cutoff

Note: This figure displays the spatial distribution of municipalities surveyed in the *Atlas Linguistique de la France* and used in the regression discontinuity design.

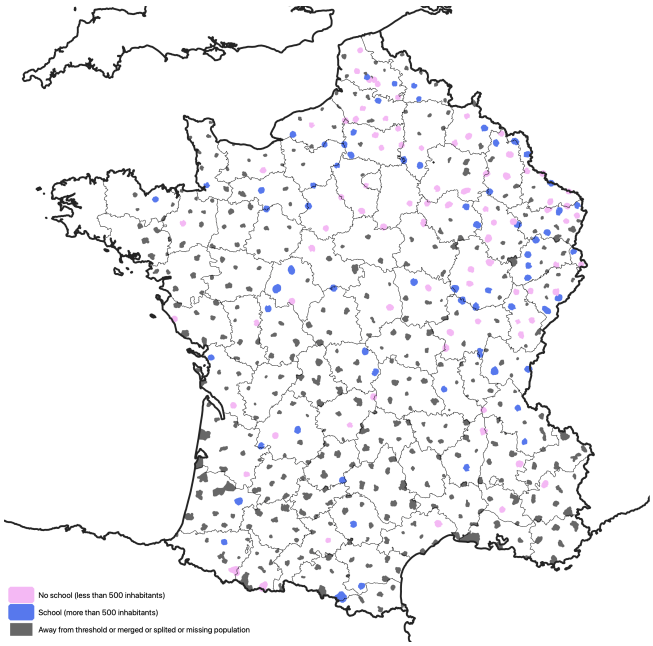


Figure A15: Distribution of population

Note: This figure displays the distribution of population at the municipality level for rural towns in the *Atlas Linguistique de la France* and in all of France. Population is measured in 1836. The p-value for the bias-corrected density test relies on Cattaneo, Jansson and Ma (2018). Source: BDCassini (2017).

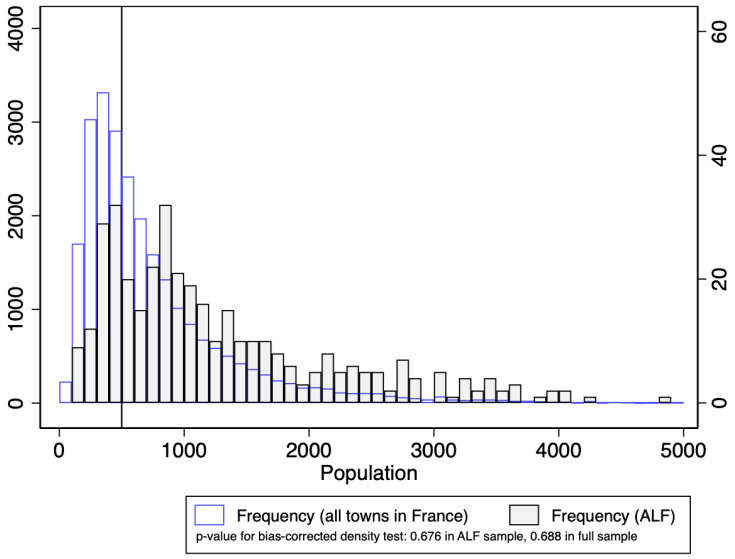


Figure A16: State-sponsored education and population (1841–1999)

Note: This figure plots regression discontinuity estimates of the effect of state-sponsored education on the log of population plotted against the year in which population was measured. We use the discontinuity introduced by the policy. State-sponsored education is defined as a dummy variable that equals 1 if population at the time of the law was above five hundred. We use local-polynomial fits of orders 0, 1, and 2 (mean, linear, quadratic) and mean-squared-error optimal bandwidths for local-polynomial estimation. We report 90 percent confidence intervals.

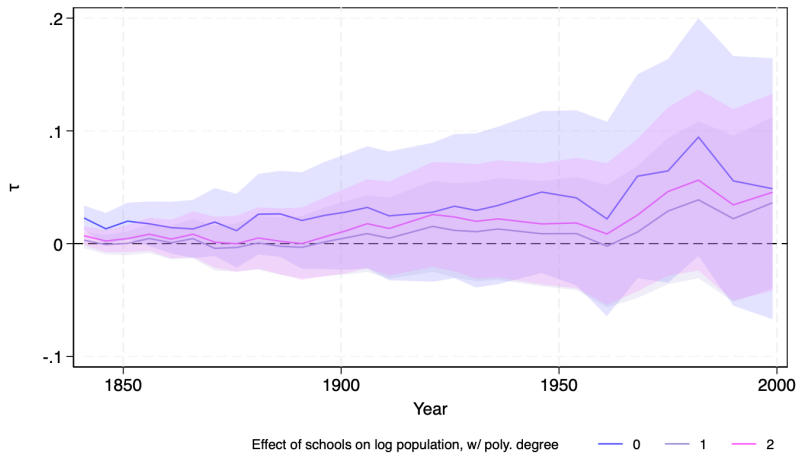
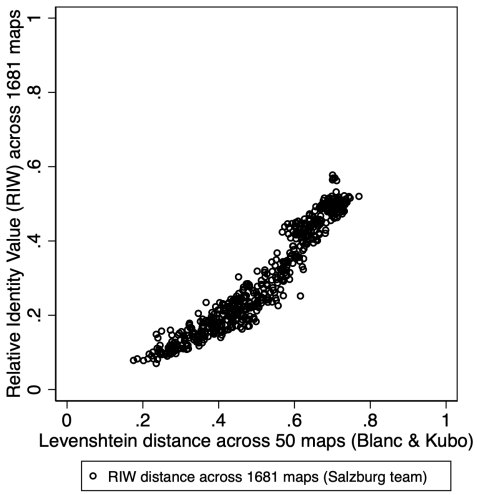
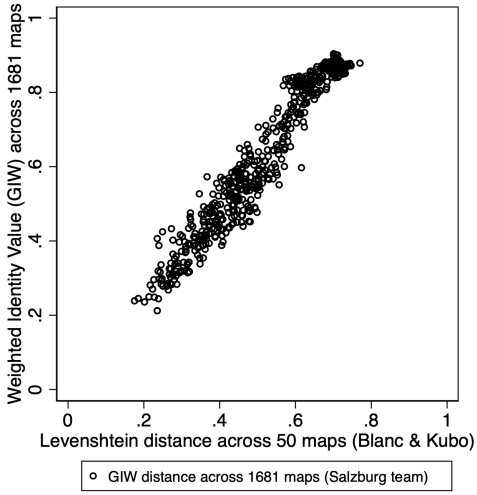


Figure A17: Measures of linguistic distance from French: Levenshtein distance across 50 maps (Blanc and Kubo) against Relative Identity Value and Weighted Identity Value across 1,681 maps (Salzburg dialectometry team) in the Atlas

Note: This figure plots our measure of linguistic distance from French, using the average Levenshtein distance across 50 words from the Swadesh list, against the Relative Identity Value (RIW, Panel A) and Weighted Identity Value (GIW, Panel B) measured across 1,681 words. RIW and GIW are similarity metrics, which we transform into distance metrics $1 - RIW/100$ and $1 - GIW/100$ for consistency and comparison purposes. The data across the entire corpus of maps in the *Atlas* was made available by Goebl et al. (2019).



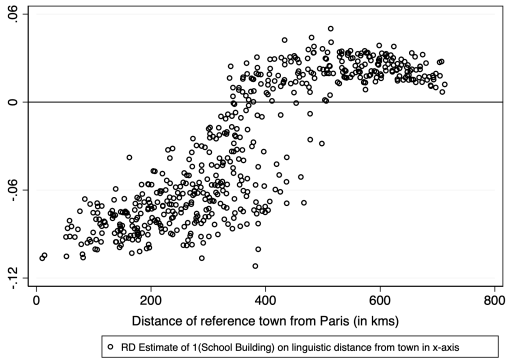
(Panel A) Relative Identity Value (RIW)



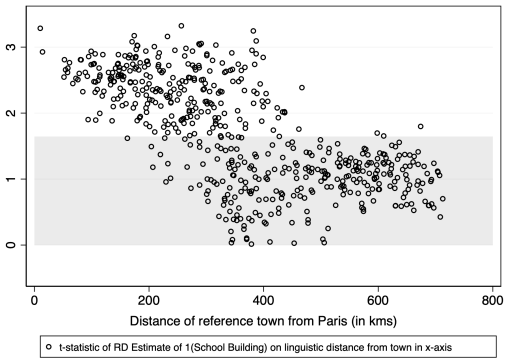
(Panel B) Weighted Identity Value (GIW)

Figure A18: Regression discontinuity estimate and t-statistics of the effect of state-sponsored education on linguistic distance from reference towns across all distances (km) from Paris

Note: This figure displays the regression discontinuity (RD) estimate and t-statistics of the effect of state-sponsored education on linguistic distance from all municipalities in France. In Panel A, the estimated coefficients are plotted against the distance of towns of reference from Paris (in km), hence showing the effect of state-sponsored education on convergence (divergence) toward (from) the language spoken in each municipality depending on how far the municipality is from Paris. Panel B does the same thing with t-statistics (in absolute value). For each regression, we apply a local-polynomial fit of order 1 and a mean-squared-error optimal bandwidth for local-polynomial estimation. Observations are at the municipality level.



(Panel A) RD Estimate



(Panel B) abs(t-statistics)

Figure A19: Tracing out the geographical origins of the French language

Note: This figure displays the spatial distribution of the regression discontinuity estimates of the effect of state-sponsored education on linguistic distance across all potential towns of origin. The estimated coefficients are displayed on the map attached to towns of reference, hence showing the effect of state-sponsored education on convergence (divergence) toward (from) the language spoken in each municipality. Coefficients not statistically significant at the 5 percent level are shaded in blue. For each regression, we apply a local-polynomial fit of order 1 and a mean-squared-error optimal bandwidth for local-polynomial estimation. Observations are at the municipality level, but, to ease readability, we generate Thiessen polygons around each reference town surveyed in the *Atlas Linguistique de la France* to plot our measure.

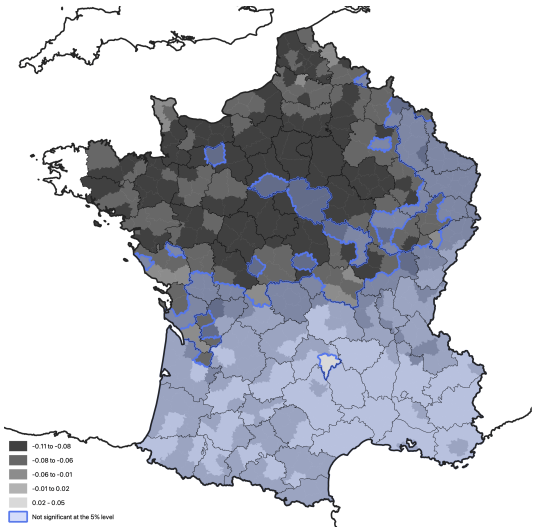


Figure A20: Heterogeneous effect of state-sponsored education: distance from departmental capital

Note: This figure displays the effect of state-sponsored education on linguistic distance from French—at low and high levels of geographical distance from the *département*'s capital—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. Distance from the departmental capital is coded as high if it is above the 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. Source: Delefortrie and Morice (1959) for income per capita.

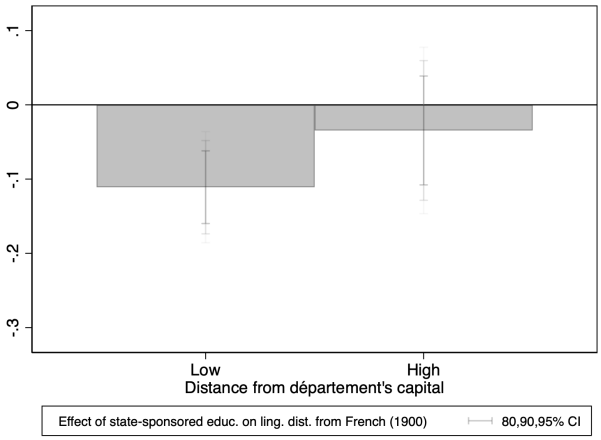


Figure A21: Heterogeneous effect of state-sponsored education: historical language regions

Note: This figure displays the effect of state-sponsored education on linguistic distance from French—in the historical *langues d'oïl* region—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. Historical language region is available at the municipality level. 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 *département* level. Source: Billy, Nadiras and Moufflet (n.d.) for historical language regions.

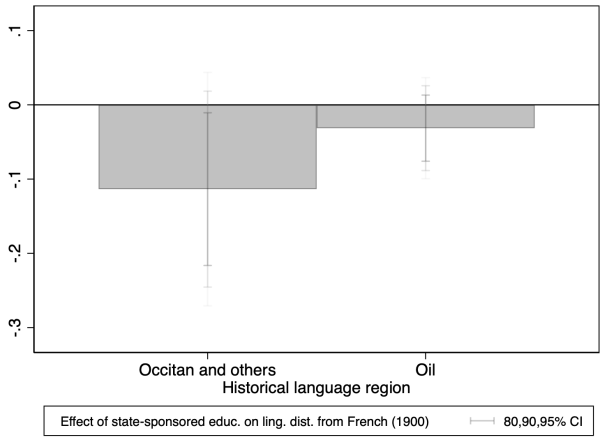


Figure A22: Heterogeneous effect of state-sponsored education: state capacity and Cinq Grosses Fermes

Note: This figure displays the effect of state-sponsored education on linguistic distance from French—at low and high levels of state capacity—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. State capacity is proxied by location inside the Cinq Grosses Fermes (CGF). 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 *département* level. Source: Johnson (2019) for the CGF.

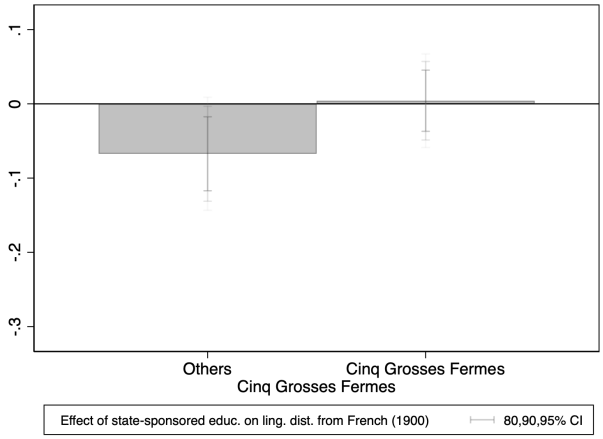


Figure A23: Heterogeneous effect of state-sponsored education: state legitimacy and religiosity

Note: This figure displays the effect of state-sponsored education on linguistic distance from French—at low and high levels of religiosity in 1791—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. Religiosity is proxied by the share of refractory clergy in 1791, available at the district level, and coded as high if above the 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 district level. Source: Tackett (1986) for religiosity.

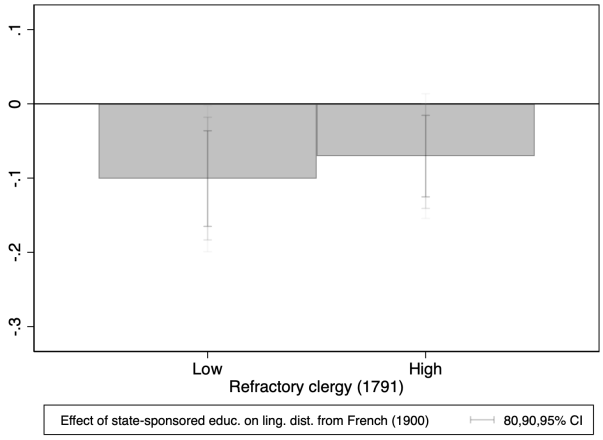


Figure A24: Heterogeneous effect of state-sponsored education: geography (proximity to coalfields)

Note: This figure displays the effect of state-sponsored education on linguistic distance from French—at low and high levels of distance from coalfields in 1931—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 inhabitants. Distance from coalfields is available at the municipality level and coded as high if it is above the 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. Source: Fernihough and O'Rourke (2020) for distance from coalfields.

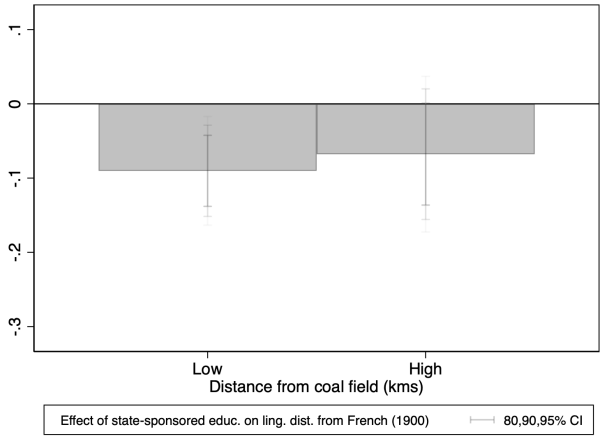
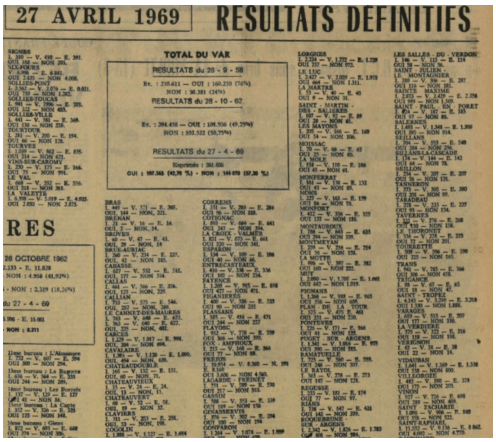


Figure A25: Regional newspapers showing results of 1969 referendum

Note: This figure displays regional newspapers showing the results of the 1969 referendum, in Lorraine (Panel A) and in Provence (Panel B). Sources: Dehdari and Gehring (2022) for Lorraine, Archives départementales du Var (1969) for Provence.

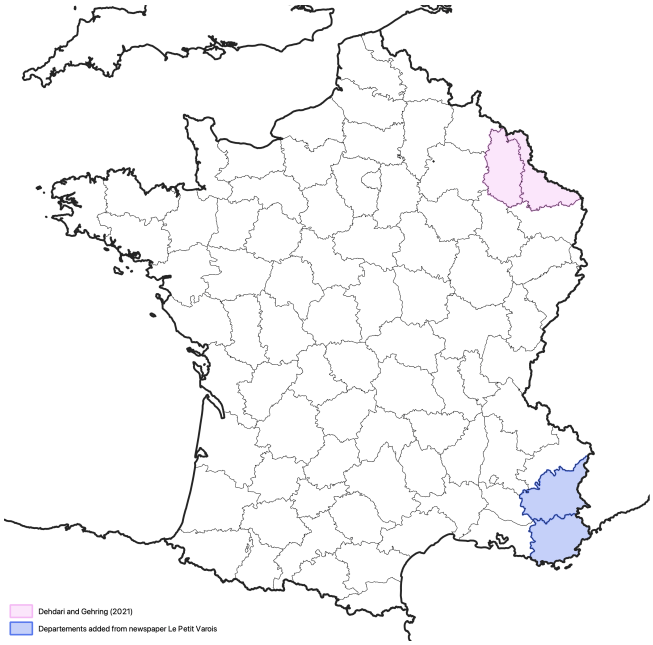
(Panel A) L'Est Republicain



(Panel B) Le Petit Varois

Figure A26: Sample of *départements* with municipality-level results of 1969 referendum

Note: This figure displays the spatial distribution of *départements* with available municipality-level results of the 1969 referendum. We show in blue *départements* added from the digitization of the journal *Le Petit Varois*.



Appendix 6– Supplementary Tables

Table A1: Selection and balance: summary statistics

Note: This table displays tests for selection into the sample and balance in observables at the threshold. Panel A evaluates selection into the ALF sample and displays the mean and standard error for the variable listed in the full sample of all municipalities in France and in the *Atlas Linguistique de la France* (ALF). We also report the difference between columns (2) and (1) and the corresponding p-value. Panel B evaluates the balance in observables at the threshold and displays the differences in the mean of the same variables, above and below the threshold for school building. We report the coefficient in the full sample of all municipalities in France (Panel B1) and in the *Atlas* only (Panel B2), and we vary the bandwidth around the school-building discontinuity using the optimal bandwidths estimated in the paper (Table A4). In Panel B1, we additionally narrow the bandwidth to 10 inhabitants—that is, municipalities between 490 and 510 inhabitants. Observations are at the municipality level, and standard errors are clustered at the *département* level for the historical *langues d'oïl* region and for the salt tax in the Ancien Régime, and clustered at the district level for religiosity. Sources: Fick and Hijmans (2017) for temperature and precipitation, Darnton (1973) for the *Encyclopedie*, Schwartz, Gregory and Thévenin (2011) for railways in 1870. $^{+}$ $p < 0.2$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

	(1) France	(2) ALF	(2)-(1) Difference	p-value
Geography				
Altitude	5.2 (0.01)	5.2 (0.05)	0.1 (0.06)	0.29
Distance from coast (kms)	4.7 (0.01)	4.6 (0.07)	-0.0 (0.07)	0.50
Climate				
Average temperature	10.8 (0.01)	10.8 (0.09)	0.1 (0.09)	0.46
Average precipitation	68.8 (0.09)	69.9 (0.69)	1.1 (0.70)	0.12
Institutional, cultural, and economic factors				
Encyclopedie subscriptions	1.0 (0.00)	1.0 (0.04)	-0.1 (0.04)	0.21
Distance from Paris (kms)	5.5 (0.00)	5.7 (0.03)	0.2 (0.03)	0.00
Distance from railway	1.9 (0.01)	1.9 (0.07)	0.0 (0.07)	0.72
Population				
Population of municipality	17,700.1 (9,711.82)	4,843.4 (905.81)	-13,222.2 (10,020.93)	0.19
Percentage of population urban	24.6 (1.99)	27.1 (5.30)	2.6 (5.67)	0.65

(Panel A) Selection into the sample

	Sample falls within ____ inhabitants of discontinuity						
Bandwidth (inhab.)	(1) 10	(2) 139	(3) 287	(4) 327	(5) 139	(6) 287	(7) 327
Difference between above and below 500 inhabitants							
	Panel B1: All municipalities				**Panel B2**: In the Atlas		
Geography							
Altitude	-0.06	-0.04*	-0.06***	-0.06***	-0.36^{+}	0.06	0.14
	(0.08)	(0.02)	(0.02)	(0.01)	(0.22)	(0.15)	(0.14)
Distance from coast (kms)	0.07	-0.07***	-0.10***	-0.12***	-0.41^{+}	-0.06	0.03
	(0.09)	(0.03)	(0.02)	(0.02)	(0.29)	(0.18)	(0.17)
Climate							
Average temperature	0.17^{+}	0.13***	0.17***	0.18***	0.27	-0.06	-0.10
	(0.12)	(0.03)	(0.02)	(0.02)	(0.25)	(0.21)	(0.22)
Average precipitation	-0.95	0.20	0.00	0.10	-3.81	2.45	4.03*
	(1.11)	(0.33)	(0.24)	(0.22)	(3.10)	(2.57)	(2.34)
Institutional, cultural, and economic factors							
Encyclopedie subscriptions	0.05	-0.02	-0.05***	-0.06***	0.03	0.06	-0.05
	(0.06)	(0.02)	(0.01)	(0.01)	(0.18)	(0.13)	(0.12)
Distance from Paris (kms)	-0.06	0.07***	0.09***	0.10***	-0.15	-0.01	0.13
	(0.06)	(0.02)	(0.01)	(0.01)	(0.16)	(0.11)	(0.11)
Distance from railway	0.06	-0.04^{+}	-0.03^{+}	-0.02	-0.02	-0.09	0.00
	(0.10)	(0.03)	(0.02)	(0.02)	(0.32)	(0.22)	(0.20)
Observations (municipalities)	610	7,371	14,722	16,273	60	123	142

(Panel B) Balance

Table A2: Exposure to *Loi Guizot* and school building

Note: This table displays the results of the regression of the log of the growth rate of the number of schools per capita following the *Loi Guizot*, from 1829 to 1850, on *département*-level exposure to the law, controlling for the log of the number of schools before the law (in 1829). The main independent variable, *département*-level exposure to the law, is defined as the share of municipalities with more than five hundred inhabitants at the time of the law. The first column reports the association estimated with ordinary least squares. The subsequent columns report the results of a two-stage-least-squares estimation, including first stage and reduced form, instrumenting the main independent variable with the share of municipalities above five hundred inhabitants in the sample of municipalities within one hundred, two hundred, three hundred, and four hundred inhabitants of the five-hundred-inhabitant threshold. We also report Kleibergen-Paap F-statistics throughout. Observations are at the *département* level, and we report robust standard errors. Source: Statistique Générale de la France (n.d.*b*). $^{+}$ $p < 0.2$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

	dep var: Growth schools per capita (1829-1850)				
		IV			
Sample for instrument/rf (bandwidth):	(1) OLS	(2) 100	(3) 200	(4) 300	(5) 400
		Panel A: 2SLS			
Perc. of municipalities above 500 inhabitants	0.98***	2.35*	1.16**	1.01***	0.84**
	(0.34)	(1.24)	(0.50)	(0.39)	(0.35)
		Panel B: First stage			
Perc. above 500 inhabitants within ___ inhab. of threshold		0.35***	0.74***	0.62***	0.54***
		(0.11)	(0.08)	(0.03)	(0.02)
Kleibergen-Paap F-stat		11	84	323	715
		Panel C: Reduced form			
Perc. above 500 inhabitants within ___ inhab. of threshold		0.82**	0.86**	0.63**	0.45**
		(0.40)	(0.36)	(0.25)	(0.19)
Schools per capita (1829)	Yes	Yes	Yes	Yes	Yes
Observations	79	79	79	79	79

Table A3: Ordinary least squares

Note: This table displays correlates of linguistic distance from French in 1900. Linguistic distance from French is defined as the Levenshtein distance from Le Plessis-Robinson. In columns (1) and (6), the historical *langues d'oïl* region is the excluded category. We report standardized beta coefficients throughout. Observations are at the municipality level, and standard errors are clustered at the *département* level. $^{+}$ $p < 0.2$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

	dep var: Linguistic distance from French (1900)					
Standardized beta coefficients (clustered se)	(1)	(2)	(3)	(4)	(5)	(6)
Franco-provençal F.E.	0.24***					0.17***
	(0.03)					(0.03)
Occitan F.E.	0.84***					0.48***
	(0.05)					(0.07)
Other languages F.E.	0.12					-0.01
	(0.08)					(0.04)
log distance from Paris (kms)		0.78***				0.41***
		(0.08)				(0.07)
log population (1836)			0.36***			-0.01
			(0.06)			(0.04)
log disp. income per capita (1864)				-0.47***		-0.14**
				(0.09)		(0.06)
Increase in log school buildings (1834 − 50)					0.09	-0.16**
					(0.10)	(0.07)
Observations	501	375	478	499	499	373
Clusters (départements)	81	78	78	80	80	77
Adjusted R^2	0.66	0.61	0.12	0.23	0.01	0.76

Table A4: State-sponsored education and linguistic distance from French, accounting for covariates

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 and accounting for covariates, following Calonico et al. (2019). 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)				
Controls:	(1) No controls	(2) Distance from Paris	(3) F.E. for region	(4) F.E. for language	(5) F.E. for dialect
	Panel A: Polynomial of degree 0				
State-sponsored education	-0.09***	-0.07**	-0.07***	-0.07***	-0.08***
	(0.03)	(0.03)	(0.02)	(0.02)	(0.01)
	Panel B: Polynomial of degree 1				
	-0.11***	-0.07**	-0.08***	-0.09***	-0.08***
	(0.03)	(0.03)	(0.03)	(0.02)	(0.01)
	Panel C: Polynomial of degree 2				
	-0.11***	-0.08*	-0.08**	-0.09***	-0.09***
	(0.04)	(0.04)	(0.03)	(0.03)	(0.02)
Effective observations (and opt. bandwidth) for Panels A, B, C:					
	60 (139 inh.)	57 (129 inh.)	52 (101 inh.)	48 (94 inh.)	46 (87 inh.)
	123 (287 inh.)	118 (260 inh.)	107 (224 inh.)	109 (233 inh.)	109 (234 inh.)
	141 (327 inh.)	169 (393 inh.)	161 (369 inh.)	160 (365 inh.)	141 (322 inh.)
Cutoff	500	500	500	500	500
Bandwidth selection	mse	mse	mse	mse	mse
Observations	375	375	375	375	375

Table A5: Robustness to different methods of estimation for regression discontinuity coefficient

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. 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 report three different estimators: $\widehat{\tau}_{CL}$, the conventional regression discontinuity estimate used throughout the paper with a triangular kernel, as in Hahn, Todd and Van der Klaauw (2001); $\widehat{\tau}_{OLS}$, using a uniform kernel, equivalent to using ordinary least squares to estimate the regression discontinuity coefficient; and $\widehat{\tau}_{BC}$, using the bias-correction methodology provided by Calonico, Cattaneo and Titiunik (2014). Columns (1), (4), and (7) show our baseline estimates. We use local-polynomial fits of orders 0, 1, and 2 (mean, linear, quadratic) and mean-squared-error optimal bandwidths for local-polynomial estimation. We use the baseline mean-squared-error optimal bandwidth of Calonico, Cattaneo and Titiunik (2014) estimated for the conventional regression discontinuity estimate. 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)								
	Poly. degree 0			Poly. degree 1			Poly. degree 2		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
State-sponsored education	-0.09***	-0.07**	-0.13***	-0.11***	-0.11***	-0.13***	-0.11***	-0.10**	-0.14***
	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
Bandwidth selection	mse			mse			mse		
Estimated coefficient	$\widehat{\tau}_{CL}$	$\widehat{\tau}_{OLS}$	$\widehat{\tau}_{BC}$	$\widehat{\tau}_{CL}$	$\widehat{\tau}_{OLS}$	$\widehat{\tau}_{BC}$	$\widehat{\tau}_{CL}$	$\widehat{\tau}_{OLS}$	$\widehat{\tau}_{BC}$
Conventional	Yes			Yes			Yes		
Ordinary Least Squares		Yes			Yes			Yes	
Bias-corrected			Yes			Yes			Yes
Bandwidth (inhab.)	139	139	139	287	287	287	327	327	327
Effective observations	60	60	60	123	123	123	141	141	141

Table A6: Robustness to alternative bandwidths

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 and across different bandwidths. 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 report local-polynomial fits of orders 0, 1, and 2 (mean, linear, quadratic). We include all municipalities in the *Atlas Linguistique de la France* on the left side of the cutoff (hence column (3) in this table is directly comparable to column (3) in Appendix Table A10), and we vary the bandwidth on the right from two hundred to five hundred inhabitants. 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)				
Bandwidth (right):	(1) 200 inh.	(2) 300 inh.	(3) 366 inh.	(4) 400 inh.	(5) 500 inh.
	Panel A: Polynomial of degree 0				
State-sponsored education	-0.07***	-0.05**	-0.04*	-0.03^{+}	-0.02
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
	Panel B: Polynomial of degree 1				
	-0.12***	-0.11***	-0.12***	-0.12***	-0.11***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
	Panel C: Polynomial of degree 2				
	-0.10**	-0.10***	-0.10***	-0.11***	-0.12***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Cutoff	500	500	500	500	500
Effective observations	116	138	161	170	191

Table A7: Robustness to accounting for age and dropping individuals not treated by the *Loi Guizot* (1833) or Ferry Laws (1881–82)

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 and accounting for the age of respondents in the *Atlas Linguistique de la France*. Individuals not treated by Guizot are those older than sixty-five years old in 1900, and those treated by Ferry are less than eighteen years old in 1900. 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, across all specifications, rely on the mean-squared-error optimal bandwidths estimated without controls for local-polynomial estimation. Observations are at the individual level, and standard errors are clustered at the municipality level. $^{+}$ $p < 0.2$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

	dep var: Linguistic distance from French (1900)								
	Poly. degree 0			Poly. degree 1			Poly. degree 2		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
State-sponsored education	-0.11***	-0.11***	-0.12***	-0.12***	-0.14***	-0.13***	-0.13***	-0.16***	-0.10*
	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)
Cutoff	500	500	500	500	500	500	500	500	500
Bandwidth selection	mse			mse			mse		
Age of surveyed individuals (quadratic)		Yes	Yes		Yes	Yes		Yes	Yes
Drop if not treated by Guizot or treated by Ferry			Yes			Yes			Yes
Bandwidth (inhab.)	128	128	128	275	275	275	342	342	342
Effective observations	66	66	56	144	140	119	182	170	147

Table A8: Robustness to accounting for gender and dropping women

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 and accounting for the gender of respondents in the *Atlas Linguistique de la France*. 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, across all specifications, rely on the mean-squared-error optimal bandwidths estimated without controls for local-polynomial estimation. Observations are at the individual level, and standard errors are clustered at the municipality-level. $^{+}$ $p < 0.2$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

	dep var: Linguistic distance from French (1900)								
	Poly. degree 0			Poly. degree 1			Poly. degree 2		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
State-sponsored education	-0.11***	-0.11***	-0.11***	-0.12***	-0.12***	-0.12***	-0.13***	-0.12***	-0.13***
	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)
Cutoff	500	500	500	500	500	500	500	500	500
Bandwidth selection	mse			mse			mse		
Male FE		Yes	Yes		Yes	Yes		Yes	Yes
Only male			Yes			Yes			Yes
Bandwidth (inhab.)	128	128	128	275	275	275	342	342	342
Effective observations	66	66	60	144	140	120	182	170	147

Table A9: Robustness to dropping places near the border

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 and dropping places within twenty-five kilometers of the national border. 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. 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 without controls for local-polynomial estimation. Observations are at the individual level. $^{+}$ $p < 0.2$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

	dep var: Linguistic distance from French (1900)					
	Poly. degree 0		Poly. degree 1		Poly. degree 2	
	(1)	(2)	(3)	(4)	(5)	(6)
State-sponsored education	-0.09***	-0.08**	-0.11***	-0.09*	-0.11***	-0.10
	(0.03)	(0.04)	(0.03)	(0.05)	(0.04)	(0.07)
Cutoff	500	500	500	500	500	500
Bandwidth selection	mse		mse		mse	
Drop if within 25 kms of border		Yes		Yes		Yes
Bandwidth (inhab.)	139	139	287	287	327	327
Effective observations	60	47	123	96	141	110

Table A10: Robustness to alternative (placebo) cutoffs

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 across different (placebo) cutoffs. 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 was above the cutoff listed in the table. We report local-polynomial fits of orders 0, 1, and 2 (mean, linear, quadratic). We include all municipalities in the *Atlas Linguistique de la France* on the left side of the cutoff (hence column (3) in this table is directly comparable to column (3) in Appendix Table A6), and we vary the cutoff from four hundred to six hundred inhabitants. 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)				
Cutoff:	(1) 400 inh.	(2) 450 inh.	(3) 500 inh.	(4) 550 inh.	(5) 600 inh.
	Panel A: Polynomial of degree 0				
State-sponsored education	-0.01	-0.01	-0.04*	0.01	0.03
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
	Panel B: Polynomial of degree 1				
	-0.00	-0.01	-0.12***	-0.03	-0.01
	(0.05)	(0.04)	(0.03)	(0.04)	(0.05)
	Panel C: Polynomial of degree 2				
	0.07	0.07	-0.10***	0.03	-0.01
	(0.08)	(0.06)	(0.04)	(0.05)	(0.08)
Bandwidth (right)	366	366	366	366	366
Effective observations	131	141	161	170	183

Table A11: Robustness to alternative (placebo) year of measurement of population

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 across different (placebo) years of measurement of population. 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 in the year listed in the table. Column (2) shows 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 1836-estimated 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)							
Cutoff based on population in (year):	(1) 1793	(2) 1820	(3) 1831	(4) 1836	(5) 1841	(6) 1866	(7) 1881	(8) 1901
	Panel A: Polynomial of degree 0							
State-sponsored education	0.04	0.01	-0.06*	-0.09***	-0.02	-0.04	-0.01	0.03
	(0.04)	(0.04)	(0.04)	(0.03)	(0.05)	(0.04)	(0.03)	(0.04)
	Panel B: Polynomial of degree 1							
	0.04	0.03	-0.08^{+}	-0.11***	-0.02	-0.05	-0.02	0.03
	(0.05)	(0.05)	(0.05)	(0.03)	(0.06)	(0.05)	(0.04)	(0.05)
	Panel C: Polynomial of degree 2							
	0.03	0.09	-0.08	-0.11***	0.02	-0.04	-0.01	0.02
	(0.08)	(0.08)	(0.07)	(0.04)	(0.07)	(0.06)	(0.06)	(0.07)
Cutoff	500	500	500	500	500	500	500	500
Bandwidth selection				mse				
Observations	386	389	381	375	390	386	394	394

Table A12: State-sponsored education and linguistic distance from French with different measures of distance: Levenshtein distance across 50 maps (Blanc and Kubo) and Relative Identity Value and Weighted Identity Value across 1681 maps (Salzburg dialectometry team) in the Atlas

Note: This table displays regression discontinuity estimates of the effect of state-sponsored education on linguistic distance from French, across different distance metrics measured across different corpus of words, using the discontinuity introduced by the policy. Linguistic distance from French is defined as the Levenshtein distance from Le Plessis-Robinson across 50 words from the Swadesh list in columns (1), (4), and (7) (baseline specification), and as Relative Identity Value (RIW) and Weighted Identity Value (GIW) across 1,681 words in the other columns. RIW and GIW are similarity metrics, which we transform into distance metrics $1 - RIW/100$ and $1 - GIW/100$ for consistency and comparison purposes. 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 in the year listed in the table. 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 without controls for local-polynomial estimation. Observations are at the municipality level. Source: Goebl et al. (2019) for the distance measures across the entire corpus of maps in the *Atlas*. $^{+}$ $p < 0.2$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

	dep var: Linguistic distance from French (1900)								
	Poly. degree 0			Poly. degree 1			Poly. degree 2		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
State-sponsored education	-0.09***	-0.06***	-0.10***	-0.11***	-0.08***	-0.12***	-0.11***	-0.07**	-0.12**
	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)	(0.05)
Cutoff	500	500	500	500	500	500	500	500	500
Bandwidth selection	mse			mse			mse		
Source for measure of distance	Blanc & Kubo	Salzburg dialectometry team		Blanc & Kubo	Salzburg dialectometry team		Blanc & Kubo	Salzburg dialectometry team	
Average linguistic distance across	50 maps	1681 maps		50 maps	1681 maps		50 maps	1681 maps	
Measure of linguistic distance	Levenshtein	RIW	GIW	Levenshtein	RIW	GIW	Levenshtein	RIW	GIW
Standardized beta coefficient	-0.32	-0.27	-0.28	-0.40	-0.36	-0.37	-0.39	-0.32	-0.32
Bandwidth (inhab.)	139	139	139	287	287	287	327	327	327
Effective observations	60	60	60	123	123	123	141	142	142

Table A13: State-sponsored education and linguistic distance from French: towns of reference near Paris

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 and across different towns of reference for standard French language. Linguistic distance from French is defined as the Levenshtein distance from Le Plessis-Robinson (town id 226), Sartrouville (town id 227), Liancourt-Saint-Pierre (town id 238), Ormoy-la-Rivière (town id 217), or Gommecourt (town id 239). State-sponsored education is defined as a dummy variable that equals one if population at the time of the law was above five hundred inhabitants. Column (1) shows our baseline estimates. We use local-polynomial fits of order 0, 1, and 2 (mean, linear, quadratic) and, across all specifications, rely on the baseline mean-squared-error optimal bandwidths estimated in the first column for local-polynomial estimation. Observations are at the individual level and standard errors are clustered at the municipality-level. $^{+}$ $p < 0.2$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

	dep var: Linguistic distance from French (1900)					
Reference town (id):	(1) 226	(2) 227	(3) 238	(4) 217	(5) 219	(6) 239
	Panel A: Polynomial of degree 0					
State-sponsored education	-0.09***	-0.09***	-0.08***	-0.09***	-0.07***	-0.07***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
	Panel B: Polynomial of degree 1					
	-0.11***	-0.10***	-0.09**	-0.11***	-0.09***	-0.09***
	(0.03)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)
	Panel C: Polynomial of degree 2					
	-0.11***	-0.09**	-0.08*	-0.10**	-0.08**	-0.09**
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Cutoff	500	500	500	500	500	500
Bandwidth selection	mse					
Distance from Paris	10 kms	14 kms	51 kms	52 kms	54 kms	58 kms
Observations	375	375	375	375	375	375

Table A14: Language convergence: schools, historical language regions, and phonetic distance from French

Note: This table displays regression discontinuity estimates of the effect of state-sponsored education on phonetic distance from French—and the heterogeneous effect in the historical *langues d'oïl* region versus other regions—using the discontinuity introduced by the policy. Phonetic distance from French is defined as the linguistic distance from French including accent minus the distance excluding accent. By excluding accents, we can compute linguistic distance without taking into account the different ways of pronouncing letters. 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 interact it with a dummy variable that equals 1 if the municipality is located in the historical *langues d'oïl* region. We use local-polynomial fits of orders 0, 1, and 2 (mean, linear, quadratic) and, across all specifications, rely on the mean-squared-error optimal bandwidths estimated in the baseline regression (with linguistic distance including accent as an outcome variable, controlling for a dummy for the historical *langues d'oïl* region) for local-polynomial estimation. 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 *département* level. $^{+}$ $p < 0.2$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

	dep var: Phonetic dist. from French (1900)		
	(1) Poly. 0	(2) Poly. 1	(3) Poly. 2
State-sponsored education	0.01^{+}	-0.01	-0.00
	(0.01)	(0.01)	(0.02)
_____ × Historical *langues d'oïl* region	-0.04**	-0.03**	-0.03***
	(0.01)	(0.01)	(0.01)
Bandwidth	105	252	333
Observations	53	116	144

Table A15: Language shift: schools, historical language regions, and linguistic distance from French excluding accents

Note: This table displays regression discontinuity estimates of the effect of state-sponsored education on linguistic distance from French excluding accents—and the heterogeneous effect of the historical *langues d'oïl* region versus other regions—using the discontinuity introduced by the policy. Linguistic distance from French excluding accents is defined as the Levenshtein distance from Le Plessis-Robinson excluding the different ways of pronouncing letters. State-sponsored education is defined as a dummy variable that equals 1 if population at the time of the law was above five hundred. We interact it with a dummy variable that equals 1 if the municipality is located in the historical *langues d'oïl* region. We use local-polynomial fits of orders 0, 1, and 2 (mean, linear, quadratic) and, across all specifications, rely on the mean-squared-error optimal bandwidths estimated in the baseline regression (with linguistic distance including accent as the outcome variable, controlling for a dummy for the historical *langues d'oïl* region) for local-polynomial estimation. 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 *département* level. $^{+}$ $p < 0.2$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

	dep var: Linguistic distance from French—without accents		
	(1) Poly. 0	(2) Poly. 1	(3) Poly. 2
State-sponsored education	-0.12^{+}	-0.06	-0.05
	(0.08)	(0.05)	(0.06)
_____ × Historical *langues d'oïl* region	0.12^{+}	0.04	0.03
	(0.09)	(0.04)	(0.03)
Bandwidth	105	252	333
Observations	53	116	144

Table A16: Heterogeneous effect of state-sponsored education

Note: This table displays regression discontinuity estimates of the effect of state-sponsored education on linguistic distance from French—and the heterogeneous effect across different variables—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. x-var corresponds to the variable listed in the columns of the table and is standardized. We use local-polynomial fits of orders 0, 1, and 2 (mean, linear, quadratic) and, across all specifications, rely on the mean-squared-error optimal bandwidths estimated in the baseline (without controls). 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 appropriate level, either municipality, district, or *département* depending on the variable (see notes in the corresponding Appendix Figures). $^{+}$ $p < 0.2$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

	dep var: Linguistic distance from French (1900)								
	Local elites			State capacity, legitimacy			Return to education		
x-variable:	(1) Encycl.	(2) Dist. prefecture	(3) Disp. income	(4) CGF	(5) Religiosity	(6) Oil Dummy	(7) Growth students	(8) Steam engine	(9) Dist. coal
	Panel A: Polynomial of degree 0								
State-sponsored education	-0.01	-0.11***	-0.08*	-0.07*	-0.10**	-0.11^{+}	-0.02	-0.01	-0.09**
	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)	(0.08)	(0.03)	(0.06)	(0.04)
_____ * x-var above median	-0.10^{+}	0.08	0.01	0.07^{+}	0.03	0.08	-0.12*	-0.09	0.02
	(0.07)	(0.07)	(0.06)	(0.05)	(0.07)	(0.09)	(0.07)	(0.07)	(0.07)
x-var above median	0.01	-0.03	-0.11**	-0.21***	0.06	-0.23***	0.13***	0.02	-0.10**
	(0.05)	(0.05)	(0.05)	(0.03)	(0.05)	(0.03)	(0.05)	(0.05)	(0.04)
	Panel B: Polynomial of degree 1								
State-sponsored education	-0.06	-0.13***	-0.12**	-0.09**	-0.13**	-0.07^{+}	-0.11**	-0.10*	-0.08*
	(0.05)	(0.04)	(0.05)	(0.04)	(0.05)	(0.05)	(0.04)	(0.06)	(0.05)
_____ * x-var above median	-0.09*	0.01	-0.01	0.04	-0.03	0.01	-0.09**	-0.06	-0.01
	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)
x-var above median	-0.01	-0.06**	-0.11***	-0.19***	0.10***	-0.21***	0.12***	0.03	-0.07**
	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)
	Panel C: Polynomial of degree 2								
State-sponsored education	-0.07	-0.09*	-0.16***	-0.07^{+}	-0.10*	-0.06	-0.07	-0.09^{+}	-0.06
	(0.06)	(0.05)	(0.06)	(0.05)	(0.06)	(0.05)	(0.06)	(0.06)	(0.05)
_____ * x-var above median	-0.06	0.00	0.01	0.02	-0.02	-0.00	-0.07*	-0.06^{+}	-0.01
	(0.05)	(0.04)	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)
x-var above median	-0.03	-0.05*	-0.13***	-0.19***	0.09***	-0.21***	0.13***	0.02	-0.08***
	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)
Effective observations (and bandwidth)									
Panel A	60 (138 inh.)	60 (135 inh.)	65 (157 inh.)	68 (163 inh.)	58 (134 inh.)	53 (105 inh.)	69 (165 inh.)	71 (167 inh.)	60 (141 inh.)
Panel B	123 (288 inh.)	125 (290 inh.)	126 (294 inh.)	106 (220 inh.)	124 (294 inh.)	116 (252 inh.)	128 (297 inh.)	133 (308 inh.)	116 (259 inh.)
Panel C	143 (330 inh.)	143 (329 inh.)	164 (375 inh.)	152 (347 inh.)	149 (345 inh.)	144 (333 inh.)	152 (348 inh.)	159 (364 inh.)	141 (322 inh.)

Table A17: State-sponsored education and French Resistance (WWII)

Note: This table displays regression discontinuity estimates of the effect of state-sponsored education on the French Resistance during World War II using the discontinuity introduced by the policy. French Resistance during World War II is defined, in columns (1) and (2), as 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," or, in columns (3) and (4), as the (standardized) inverse hyperbolic sine of the number of medals per capita. 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. The full set of controls and fixed effects include the log of distance from Paris as well as *département*, historical-dialect, and Vichy-regime fixed effects. Observations are at the municipality level. Source: Site Mémoire des hommes (n.d.) for French Resistance, Ferwerda and Miller (2014) for demarcation line during Vichy regime. $^{+}$ $p < 0.2$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

	dep var: Resistance Medals (WWII)			
	Dummy=1 if Y-var>0		asinh Y-var pc (std.)	
	(1)	(2)	(3)	(4)
	Panel A: Polynomial of degree 0			
State-sponsored education	0.05***	0.05***	0.01	0.02
	(0.02)	(0.02)	(0.03)	(0.03)
	Panel B: Polynomial of degree 1			
	0.05**	0.06***	0.02	0.03
	(0.02)	(0.02)	(0.03)	(0.03)
	Panel C: Polynomial of degree 2			
	0.05**	0.06**	0.03	0.05
	(0.03)	(0.03)	(0.04)	(0.04)
Effective obs. and opt. bandwidth (right) for Panels A, B, C:				
	3,245 (62 inh.)	3,132 (59 inh.)	3,975 (75 inh.)	3,975 (76 inh.)
	5,855 (111 inh.)	5,434 (102 inh.)	8,507 (161 inh.)	7,509 (143 inh.)
	11,842 (226 inh.)	11,471 (218 inh.)	10,142 (192 inh.)	9,887 (187 inh.)
Cutoff	500	500	500	500
Bandwidth selection	mse	mse	mse	mse
Full set of controls and FEs		Yes		Yes
Observations	25,190	25,190	25,190	25,190

Table A18: State-sponsored education and unverified acts of resistance (WWII)

Note: This table displays regression discontinuity estimates of the effect of state-sponsored education on reported acts of French resistance during World War II using the discontinuity introduced by the policy. Unverified acts of resistance during World War II is defined, in columns (1) and (2), as a dummy variable that equals 1 if at least one individual born in the municipality was reported to be a resistant fighter in the database "Titres, homologations et services pour faits de résistance," or, in columns (3) and (4), as the (standardized) inverse hyperbolic sine of the number of unverified resistant fighters per capita. This database lists nearly six hundred thousand men and women who are reported (not necessarily confirmed) to have participated in some way in the Resistance. 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. The full set of controls and fixed effects include the log of distance from Paris as well as *département*, historical-dialect, and Vichy-regime fixed effects. Observations are at the municipality level. Source: Site Mémoire des hommes (n.d.) for French Resistance, Ferwerda and Miller (2014) for demarcation line during Vichy regime. $^{+}$ $p < 0.2$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

	dep var: Resistance acts (unverified) (WWII)			
	Dummy=1 if Y-var>0		asinh Y-var pc (std.)	
	(1)	(2)	(3)	(4)
	Panel A: Polynomial of degree 0			
State-sponsored education	0.01	0.02	0.00	0.03
	(0.02)	(0.02)	(0.03)	(0.03)
	Panel B: Polynomial of degree 1			
	-0.00	0.00	0.01	0.05^{+}
	(0.02)	(0.02)	(0.04)	(0.04)
	Panel C: Polynomial of degree 2			
	-0.00	0.00	0.01	0.06
	(0.02)	(0.02)	(0.05)	(0.04)
Effective obs. and opt. bandwidth (right) for Panels A, B, C:				
	2,004 (38 inh.)	1,956 (37 inh.)	3,586 (68 inh.)	3,245 (62 inh.)
	8,203 (156 inh.)	8,134 (155 inh.)	6,763 (128 inh.)	5,803 (110 inh.)
	11,709 (223 inh.)	11,522 (220 inh.)	10,523 (200 inh.)	9,999 (190 inh.)
Cutoff	500	500	500	500
Bandwidth selection	mse	mse	mse	mse
Full set of controls and FEs		Yes		Yes
Observations	25,190	25,190	25,190	25,190

Table A19: State-sponsored education and Nazi collaborators (WWII)

Note: This table displays regression discontinuity estimates of the effect of state-sponsored education on collaboration with the Nazis during World War II using the discontinuity introduced by the policy. Nazi collaborators during World War II is defined, in columns (1) and (2), as a dummy variable that equals 1 if at least one individual born in the municipality joined the Gestapo, the German Intelligence Service, or the Waffen-SS, or, in columns (3) and (4), as the (standardized) inverse hyperbolic sine of the number of collaborators per capita. 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. The full set of controls and fixed effects include the log of distance from Paris, as well as *département*, historical-dialect, and Vichy-regime fixed effects. We also control for Verdun and Verdun-under-Pétain, following Cagé et al. (2023). Observations are at the municipality level. Source: Cagé et al. (2023) for collaboration with the Nazis, Ferwerda and Miller (2014) for demarcation line during Vichy regime. $^{+}$ $p < 0.2$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

	dep var: Nazi collaborators (WWII)			
	Dummy=1 if Y-var>0		asinh Y-var pc (std.)	
	(1)	(2)	(3)	(4)
	Panel A: Polynomial of degree 0			
State-sponsored education	-0.00	-0.00	-0.06*	-0.05*
	(0.00)	(0.00)	(0.03)	(0.03)
	Panel B: Polynomial of degree 1			
	-0.01	-0.01	-0.08*	-0.08*
	(0.01)	(0.01)	(0.05)	(0.05)
	Panel C: Polynomial of degree 2			
	-0.01^{+}	-0.01^{+}	-0.09*	-0.09^{+}
	(0.01)	(0.01)	(0.05)	(0.05)
Effective obs. and opt. bandwidth (right) for Panels A, B, C:				
	4,230 (80 inh.)	4,382 (83 inh.)	3,817 (72 inh.)	4,116 (78 inh.)
	8,032 (153 inh.)	8,032 (152 inh.)	6,484 (122 inh.)	6,218 (118 inh.)
	12,341 (237 inh.)	12,103 (231 inh.)	11,166 (211 inh.)	10,761 (203 inh.)
Cutoff	500	500	500	500
Bandwidth selection	mse	mse	mse	mse
Full set of controls and FEs		Yes		Yes
Observations	25,190	25,190	25,190	25,190

Table A20: State-sponsored education and Nazi collaborators (WWII)

Note: This table displays regression discontinuity estimates of the effect of state-sponsored education on collaboration during World War II using the discontinuity introduced by the policy. We explore Nazi collaboration, political collaboration, paramilitary collaboration, all types of collaboration together, and all types of collaborations for non-migrants only, following Table 7 of Cagé et al. (2023). In Panel A, collaboration is defined as a dummy variable that equals 1 if at least one individual born in the municipality was a collaborator. In Panel B, collaboration is defined as the (standardized) inverse hyperbolic sine of the number of collaborators per capita. 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. The full set of controls and fixed effects include the log of distance from Paris, as well as *département*, historical-dialect, and Vichy-regime fixed effects. We also control for Verdun and Verdun-under-Pétain, following Cagé et al. (2023). Observations are at the municipality level. Source: Cagé et al. (2023) for collaboration with the Nazis, Ferwerda and Miller (2014) for demarcation line during Vichy regime. $^{+}$ $p < 0.2$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

	dep var: Dummy=1 if Collaborators (WWII)>0									
	Nazi		Political		Paramilitary		All		Non-movers	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Panel A: Polynomial of degree 0									
State-sponsored education	-0.00	-0.00	0.01	0.00	0.01	0.01	0.00	0.00	0.00	-0.00
	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)
	Panel B: Polynomial of degree 1									
	-0.01	-0.01	0.00	-0.01	0.01	0.01	-0.01	-0.01	-0.01	-0.01
	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
	Panel C: Polynomial of degree 2									
	-0.01^{+}	-0.01^{+}	0.00	-0.01	0.01	0.01	-0.00	-0.01	-0.01	-0.01
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Effective obs. and opt. bandwidth (right) for Panels A, B, C:										
	4,230 (80 inh.)	4,382 (83 inh.)	3,192 (61 inh.)	2,753 (51 inh.)	4,327 (82 inh.)	4,013 (76 inh.)	3,132 (60 inh.)	2,705 (51 inh.)	3,547 (67 inh.)	2,840 (54 inh.)
	8,032 (153 inh.)	8,032 (152 inh.)	6,709 (127 inh.)	6,328 (119 inh.)	10,065 (190 inh.)	6,963 (132 inh.)	7,785 (147 inh.)	7,712 (146 inh.)	7,112 (134 inh.)	6,963 (132 inh.)
	12,341 (237 inh.)	12,103 (231 inh.)	11,749 (225 inh.)	11,429 (218 inh.)	12,667 (244 inh.)	11,943 (229 inh.)	12,476 (240 inh.)	12,157 (232 inh.)	11,842 (227 inh.)	11,709 (224 inh.)
Cutoff	500	500	500	500	500	500	500	500	500	500
Bandwidth selection	mse	mse	mse	mse	mse	mse	mse	mse	mse	mse
Full set of controls and FEs		Yes		Yes		Yes		Yes		Yes
Observations	25,190	25,190	25,190	25,190	25,190	25,190	25,190	25,190	25,190	25,190

(Panel A) Dummy variable

	dep var: asinh Collaborators (WWII) pc. (std.)									
	Nazi		Political		Paramilitary		All		Non-movers	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Panel A: Polynomial of degree 0									
State-sponsored education	-0.06*	-0.05*	0.01	0.00	-0.02	-0.01	-0.04	-0.04^{+}	-0.04	-0.04^{+}
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)	(0.04)	(0.03)
	Panel B: Polynomial of degree 1									
	-0.08*	-0.08*	-0.02	-0.03	-0.01	-0.01	-0.05	-0.06	-0.05	-0.06
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)
	Panel C: Polynomial of degree 2									
	-0.09*	-0.09^{+}	-0.02	-0.04	-0.01	0.00	-0.05	-0.06	-0.07	-0.07
	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.06)
Effective obs. and opt. bandwidth (right) for Panels A, B, C:										
	3,817 (72 inh.)	4,116 (78 inh.)	3,547 (67 inh.)	3,864 (73 inh.)	3,586 (69 inh.)	3,398 (65 inh.)	3,975 (76 inh.)	4,166 (79 inh.)	3,647 (69 inh.)	3,647 (70 inh.)
	6,484 (122 inh.)	6,218 (118 inh.)	4,612 (88 inh.)	4,327 (82 inh.)	7,306 (139 inh.)	6,380 (120 inh.)	5,645 (107 inh.)	5,645 (107 inh.)	5,394 (101 inh.)	5,434 (102 inh.)
	11,166 (211 inh.)	10,761 (203 inh.)	8,255 (156 inh.)	8,134 (155 inh.)	11,256 (214 inh.)	10,523 (199 inh.)	9,730 (185 inh.)	9,631 (182 inh.)	9,363 (177 inh.)	9,256 (176 inh.)
Cutoff	500	500	500	500	500	500	500	500	500	500
Bandwidth selection	mse	mse	mse	mse	mse	mse	mse	mse	mse	mse
Full set of controls and FEs		Yes		Yes		Yes		Yes		Yes
Observations	25,190	25,190	25,190	25,190	25,190	25,190	25,190	25,190	25,190	25,190

(Panel B) Inverse hyperbolic sine

Table A21: State-sponsored education and *Morts pour la France* (WWI and II)

Note: This table displays regression discontinuity estimates of the effect of state-sponsored education on the numbers of deaths during World War I (in columns (1) and (2)) and World War II (in columns (3), (4), (5), and (6)) using the discontinuity introduced by the policy. Deaths in World Wars I are defined as the log number of individuals born in the municipality who died in action or from an injury or an illness contracted during service. Deaths in World Wars II are defined as a dummy variable that equals 1 if at least one individual born in the municipality died in action or from an injury or an illness contracted during service (in columns (3) and (4)), or as the (standardized) inverse hyperbolic sine of the number of death (in columns (5) and (6)). 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. The full set of controls and fixed effects include the log of distance from Paris as well as *département*, historical-dialect, and (for WWII only) Vichy-regime fixed effects. Observations are at the municipality level. Source: Site Mémoire des hommes (n.d.) for French Resistance, Ferwerda and Miller (2014) for demarcation line during Vichy regime. $^{+}$ $p < 0.2$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

	dep var: *Morts pour la France*					
	World War 1		World War 2			
	log Y-var pc		Dummy=1 if Y-var>0		asinh Y-var pc (std.)	
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Polynomial of degree 0					
State-sponsored education	0.02	0.02	-0.00	-0.01	0.02	0.01
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
	Panel B: Polynomial of degree 1					
	0.02	0.02	-0.01	-0.02	0.03	0.04
	(0.03)	(0.02)	(0.02)	(0.02)	(0.04)	(0.04)
	Panel C: Polynomial of degree 2					
	0.03	0.02	-0.02	-0.02	0.04	0.05
	(0.03)	(0.03)	(0.03)	(0.03)	(0.05)	(0.05)
Effective obs. and opt. bandwidth (right) for Panels A, B, C:						
	3,582 (68 inh.)	3,582 (68 inh.)	2,594 (48 inh.)	2,387 (45 inh.)	3,493 (67 inh.)	3,453 (65 inh.)
	7,150 (135 inh.)	7,105 (135 inh.)	7,560 (144 inh.)	7,600 (144 inh.)	8,356 (158 inh.)	7,985 (152 inh.)
	11,362 (216 inh.)	11,230 (213 inh.)	12,667 (244 inh.)	12,476 (240 inh.)	11,429 (218 inh.)	11,215 (212 inh.)
Cutoff	500	500	500	500	500	500
Bandwidth selection	mse	mse	mse	mse	mse	mse
Full set of controls and FEs		Yes		Yes		Yes
Observations	24,971	24,971	25,190	25,190	25,190	25,190

Table A22: State-sponsored education and 1969 constitutional referendum on regionalization

Note: This table displays regression discontinuity estimates of the effect of state-sponsored education on votes in favor of the 1969 constitutional referendum on regionalization using the discontinuity introduced by the policy. Votes during the 1969 constitutional referendum on regionalization are defined as the vote share in favor of the referendum. 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. The full set of controls and fixed effects include the log of distance from Paris as well as region and historical-language fixed effects. Observations are at the municipality level. Sources: Dehdari and Gehring (2022) and Archives départementales du Var (1969). $^{+}$ $p < 0.2$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

	dep var: Perc. Yes (1969 Referendum)					
	Poly. degree 0		Poly. degree 1		Poly. degree 2	
	(1)	(2)	(3)	(4)	(5)	(6)
State-sponsored education	-0.02^{+}	-0.03*	-0.03	-0.04*	-0.03	-0.06**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
Cutoff	500	500	500	500	500	500
Bandwidth selection	mse	mse	mse	mse	mse	mse
Bandwidth (inhab.)	119	89	160	156	217	214
Full set of controls and FEs		Yes		Yes		Yes
Observations	1,002	1,002	1,002	1,002	1,002	1,002
Effective observations	288	215	404	392	550	542

Table A23: State-sponsored education and 1969 referendum (by region)

Note: This table displays regression discontinuity estimates of the effect (by region) of state-sponsored education on votes in favor of the 1969 constitutional referendum on regionalization using the discontinuity introduced by the policy. Votes during the 1969 constitutional referendum on regionalization are defined as the vote share in favor of the referendum. We separate the sample into towns in the Lorraine region, in columns (1) and (2), and those in Provence, in columns (3) and (4). State-sponsored education is defined as a dummy variable that equals 1 if population at the time of the law was above five hundred. We use local-polynomial fits of orders 0, 1, and 2 (mean, linear, quadratic) and mean-squared-error optimal bandwidths for local-polynomial estimation. The full set of controls and fixed effects include the log of distance from Paris as well as historical-language fixed effects. Observations are at the municipality level. Sources: Dehdari and Gehring (2022) and Archives départementales du Var (1969). $^{+}$ $p < 0.2$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

	dep var: Votes 1969 Referendum			
	Yes (Provence)		Yes (Lorraine)	
	(1)	(2)	(3)	(4)
	Panel A: Polynomial of degree 0			
State-sponsored education	-0.03^{+}	-0.03^{+}	-0.04	-0.03
	(0.02)	(0.02)	(0.04)	(0.04)
	Panel B: Polynomial of degree 1			
	-0.05*	-0.05*	-0.07	-0.07
	(0.03)	(0.03)	(0.05)	(0.05)
	Panel C: Polynomial of degree 2			
	-0.05^{+}	-0.05^{+}	-0.08	-0.08
	(0.03)	(0.03)	(0.06)	(0.06)
Effective obs. and opt. bandwidth (right) for Panels A, B, C:				
	175 (84 inh.)	181 (87 inh.)	39 (134 inh.)	39 (131 inh.)
	262 (124 inh.)	255 (122 inh.)	66 (218 inh.)	63 (209 inh.)
	419 (193 inh.)	396 (181 inh.)	98 (321 inh.)	95 (314 inh.)
Cutoff	500	500	500	500
Bandwidth selection	mse	mse	mse	mse
Full set of controls and FEs		Yes		Yes
Observations	810	810	192	192

Table A24: State-sponsored education and votes for the Far Right

Note: This table displays regression discontinuity estimates of the effect of state-sponsored education on votes for the Far Right using the discontinuity introduced by the policy. Votes for the far right is defined as the vote share in favor of Jean-Marie Le Pen in the first round of the 2002 presidential election, in columns (1) and (2), or in favor of Marine Le Pen in the first round of the 2017 presidential election, in columns (3) and (4). State-sponsored education is defined as a dummy variable that equals one if population at the time of the law was above five hundred inhabitants. We use local-polynomial fits of order 0, 1, and 2 (mean, linear, quadratic) and mean-squared-error optimal bandwidths for local-polynomial estimation. The full set of controls and fixed effects include the log of distance from Paris as well as *département* and historical-dialect fixed effects. Observations are at the municipality level. $^{+}$ $p < 0.2$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

	dep var: Votes Far Right			
	Le Pen (2002)		Le Pen (2017)	
	(1)	(2)	(3)	(4)
	Panel A: Polynomial of degree 0			
State-sponsored education	-0.00	-0.00	-0.01**	-0.01**
	(0.00)	(0.00)	(0.00)	(0.00)
	Panel B: Polynomial of degree 1			
	-0.00	-0.00	-0.01*	-0.01*
	(0.00)	(0.00)	(0.00)	(0.00)
	Panel C: Polynomial of degree 2			
	-0.00	-0.00	-0.01	-0.01^{+}
	(0.00)	(0.00)	(0.01)	(0.00)
Effective obs. and opt. bandwidth (right) for Panels A, B, C:				
	3,995 (78 inh.)	3,433 (66 inh.)	3,189 (62 inh.)	3,077 (60 inh.)
	9,283 (179 inh.)	8,006 (155 inh.)	6,601 (127 inh.)	6,074 (116 inh.)
	11,767 (228 inh.)	11,412 (221 inh.)	11,573 (224 inh.)	10,942 (210 inh.)
Cutoff	500	500	500	500
Bandwidth selection	mse	mse	mse	mse
Full set of controls and FEs		Yes		Yes
Observations	24,527	24,527	24,524	24,524

Table A25: State-sponsored education and European Referenda of 1992 and 2005

Note: This table displays regression discontinuity estimates of the effect of state-sponsored education on votes in favor of the European Referenda of 1992 (in columns (1) and (2)) and 2005 (in columns (3) and (4)) using the discontinuity introduced by the policy. Votes during the European Referenda of 1992 and 2005 are defined at the vote share in favor of the referendum (yes). State-sponsored education is defined as a dummy variable that equals one if population at the time of the law was above five hundred inhabitants. We use local-polynomial fits of order 0, 1, and 2 (mean, linear, quadratic) and mean-squared-error optimal bandwidths for local-polynomial estimation. The full set of controls and fixed effects include the log of distance from Paris as well as *département* and historical-dialect fixed effects. Observations are at the municipality level. Sources: Dehdari and Gehring (2022) and Archives départementales du Var (1969). $^{+}$ $p < 0.2$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

	dep var: Votes European Referenda			
	Yes (1992)		Yes (2005)	
	(1)	(2)	(3)	(4)
	Panel A: Polynomial of degree 0			
State-sponsored education	0.00^{+}	0.00^{+}	0.01	0.01^{+}
	(0.00)	(0.00)	(0.00)	(0.00)
	Panel B: Polynomial of degree 1			
	0.00	0.00	0.01^{+}	0.01^{+}
	(0.01)	(0.00)	(0.00)	(0.00)
	Panel C: Polynomial of degree 2			
	-0.00	0.00	0.01^{+}	0.01^{+}
	(0.01)	(0.01)	(0.01)	(0.01)
Effective obs. and opt. bandwidth (right) for Panels A, B, C:				
	5,124 (99 inh.)	5,021 (97 inh.)	2,542 (49 inh.)	2,830 (54 inh.)
	5,351 (103 inh.)	5,879 (112 inh.)	7,541 (146 inh.)	7,147 (138 inh.)
	10,003 (192 inh.)	10,456 (200 inh.)	11,574 (225 inh.)	11,175 (216 inh.)
Cutoff	500	500	500	500
Bandwidth selection	mse	mse	mse	mse
Full set of controls and FEs		Yes		Yes
Observations	24,577	24,577	24,525	24,525

Appendix 7– Appendix references

Agulhon, Maurice. 1970. *La République au village.* Paris:Pion.

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

Archives Nationales. n.d.. *Fonds Guizot (1560-1878).* France Archives. Available at **https://www.siv.archives-nationales.culture.gouv.fr/siv/IR/FRAN$_IR_0$03444**.

Azar, Paola, and Sergio Espuelas. 2021. "Democracy and Primary Education Spending in Spain, 1902-22." UB Economics Working Papers E21/409.

BDCassini. 2017. *Geohistoricaldata.* Paris:LaDéHiS, CRH, EHESS.

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**)."

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.

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.

Calonico, Sebastian, Matias D. Cattaneo, Max H. Farrell, and Rocío Titiunik. 2019. "Regression Discontinuity Designs Using Covariates." *The Review of Economics and Statistics*, 101(3): 442–451.

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

Chevallier, Pierre, Bernard Grosperrin, and Jean Maillet. 1968. *L'enseignement Francais De La Revolution a Nos Jours.* Paris:Editions Mouton.

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

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.

Diebolt, Claude, Magali Jaoul-Grammare, and Faustine Perrin. 2022. "A Cliometric Reading of the Development of Primary Education in France in the Nineteenth Century." Working paper.

Fernihough, Alan, and Kevin Hjortshøj O'Rourke. 2020. "Coal and the European Industrial Revolution." *The Economic Journal*, 131(635): 1135–1149.

Ferwerda, Jeremy, and Nicholas L. Miller. 2014. "Political Devolution and Resistance to Foreign Rule: A Natural Experiment." *The American Political Science Review*, 108(3): 642–660.

Fick, Stephen E., and Robert J. Hijmans. 2017. "WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas." *International Journal of Climatology*, 37(12): 4302–4315.

French Government. n.d.. *Jules Ferry rend l'enseignement primaire obligatoire.* Available at **https://www.gouvernement.fr/partage/10037-jules-ferry-rendait-l-enseignement-primaire-obligatoire**.

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

García-Peñalosa, Cecilia, and Vincent Bignon. 2021. "The Toll of Tariffs: Protectionism, Education and Fertility in Late 19th Century France." Working paper.

Gevaert, Herve. 2017. "La loi Guizot du 28 juin 1833, une sortie de l'ancien régime scolaire ? : recherche sur l'organisation pédagogique des classes et les écoles primaires supérieures jusqu'à l'enseignement secondaire spécial de Victor Duruy." PhD diss. Normandie Université.

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

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."

Guionnet, Christine. 1996. "Élections et apprentissage de la politique. Les élections municipales sous la Monarchie de Juillet." *Revue française de science politique*, 46(4): 555–579.

Guionnet, Christine. 1998. "La politique au village. Une révolution silencieuse." *Revue d'Histoire Moderne Contemporaine*, 45(4): 775–788.

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.

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

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

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

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.

O'Connor, Mary Consolata. 1955. *The historical thought of François Guizot.* Washington:Catholic University of America Press.

Schwartz, Robert, Ian Gregory, and Thomas Thévenin. 2011. "Spatial History: Railways, Uneven Development, and Population Change in France and Great Britain, 1850–1914." *The Journal of Interdisciplinary History*, 42(1): 53–88.

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."

Spolaore, Enrico, and Romain Wacziarg. forthcoming. "Fertility and Modernity." *The Economic Journal.*

Statistique Générale de la France. n.d.. *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.

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

Tudesq, André-Jean. 1972. "L'administration municipale dans le Sud-Ouest sous la Monarchie de Juillet." *Annales du Midi*, 84(109): 483–492. Included in a thematic issue : L'administration des communautés d'habitants avant et après la Révolution.

Tudesq, André-Jean. 1982. "Le monde paysan dans le système politique censitaire : un absent ou un enjeu ?" *Annales de Bretagne et des pays de l'Ouest*, 89(2): 215–228. Included in a thematic issue : Les paysans et la politique (1750-1850).

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