From *Pluribus* to *Unum*? The Civil War and Imagined Sovereignty in 19th Century America

Dataverse Appendix

Melissa M. Lee
University of Pennsylvania
mmlee@upenn.edu

Nan Zhang
Mannheim Center for European Social Research
n.zhang@uni-mannheim.de

Tilmann Herchenröder
University of Oxford
tilmann.herchenroder@wolfson.ox.ac.uk

Contents

D1 Details on Variable Coding **D1**

D2 Information about Textual Sources and Data Processing **D5**

D3 Modernization as a Potential Confounding Factor **D8**

D4 Full Regression Results from the Main Text **D11**

D5 Robustness: Including Border States with the North **D15**

D6 Robustness: City and State Fixed Effects **D19**

D7 Robustness: States as Areal Units **D22**

D8 Full Regression Results from the Dataverse Appendix **D25**

D1 Details on Variable Coding

Dependent variable

Singular usage

This dichotomous variable takes 1 if the noun "United States" is singular and 0 otherewise (i.e., plural). The noun must appear in the grammatical subject position of a sentence. Source data are from Gale 2021, Readex 2021, Phillips 2015 and Gentzkow, Shapiro, and Taddy 2018.

Main predictors: Newspaper Corpus

Year

This variable captures the year of publication for each grammatical subject mention. We center this variable at 1865 in our tests of the abandonment hypothesis and at 1860 in all other analyses.

Period: Southern Defeat

This dichotomous indicator takes 1 if the year of publication after 1865 and 0 otherwise.

Period: War Onset

This dichotomous indicator takes 1 if the year of publication after 1860 and 0 otherwise.

North

This dichotomous indicator takes 1 if the place of publication was in a Northernn state and 0 otherwise. North is defined as the states (and their territorial predecessors) that fought for the Union: California, Connecticut, Illinois, Indiana, Iowa, Kansas, Maine, Massachusetts, Michigan, Minnesota, Nevada, New Hampshire, New Jersey, New York, Ohio, Oregon, Pennsylvania, Rhode Island, Vermont, Washington, D.C., and Wisconsin. This measure excludes the four "border" states of Delaware, Kentucky, Maryland, and Missouri, as well as West Virginia, which seceded from Virginia.

Lincoln County

This dichotomous indicator takes 1 if the county (state) voted for Lincoln in the 1864 election and 0 otherwise. Data are from Clubb, Flanigan, and Zingale 2006.

Main predictors: Speech Corpus

Year of Speech This variable captures the year of speech for each grammatical subject mention. We center this variable at 1860.

Period: War Onset This dichotomous indicator takes 1 if the year of speech after 1860 and 0 otherwise.

Whig/Republican This dichotomous indicator takes 1 if the Congressmember was a member of the Whig or Republican parties and 0 if he was a member of the Democratic party.

Covariates: Newspaper Corpus

We calculate several covariates related to modernization. We match newspaper content by publication year to data from the nearest decade, using county (state) boundaries from that decade.

% Urban.
This variable measures the percentage of the urban population in the county (state) using census data from Manson, Schroeder, Van Riper et al. 2020. We standardize this measure to have mean 0 and standard deviation 1.

Post Office Density
This variable is defined as the count of post offices in a county (state) divided by the area of the newspaper's county (state) of publication in square kilometers. We standardize this measure to have mean 0 and standard deviation 1.

We manually transcribed archival post office data from Osgood 1790 (1790), Granger 1802 (1800), United States Post Office Department 1842 (1840), United States Post Office Department 1862 (1860), United States Post Office Department 1881 (1880), United States Post Office Department 1892 (1890), and United States Post Office Department 1900 (1900).

Post office data for decades 1810, 1820, 1830, 1850, and 1870 are from Acemoglu, Moscona, and Robinson 2016.

We calculate county (state) areas from GIS shapefiles from Manson, Schroeder, Van Riper et al. 2020.

Terrain Ruggedness

This variable measures the difference in average elevation in any given square kilometer area and the average elevations of each of its neighboring square kilometer areas. We calculate ruggedness at the county (state) level using GIS data from Shaver, Carter, and Shawa 2019. We standardize this measure to have mean 0 and standard deviation 1.

County on River

This dichotomous indicator is coded as 1 if the county (state) contains a steamboat-navigated river and 0 otherwise. Data are from Atack 2015.

County on Canal

This dichotomous indicator is coded as 1 if the county (state) contains a canal and 0 otherwise. Data are from Atack 2017.

Covariates: Speech Corpus

We calculate several covariates related to modernization. For all birth variables, we match the Congressmember's year of birth to data from the nearest decade, using county (state) boundaries from that decade. Birth locations are coded from United States Congress 2020.

County of Birth: % Urban

This variable measures the percentage of the urban population in the Congressmember's county (state) of birth using census data from Manson, Schroeder, Van Riper et al. 2020. We standardize this measure to have mean 0 and standard deviation 1.

County of Birth: Post Office Density

This variable is defined as the count of post offices in the Congressmember's county (state) of birth divided by the area of the county (state) in square kilometers. We standardize this measure to have mean 0 and standard deviation 1.

We manually transcribed archival post office data from Osgood 1790 (1790), Granger 1802 (1800), United States Post Office Department 1842 (1840), United States Post Office Department 1862 (1860), United States Post Office Department 1881 (1880), United States Post Office Department 1892 (1890), and United States Post Office Department 1900 (1900).

Post office data for decades 1810, 1820, 1830, 1850, and 1870 are from Acemoglu, Moscona, and Robinson 2016.

We calculate county (state) areas from GIS shapefiles from Manson, Schroeder, Van Riper et al. 2020.

County of Birth: Terrain Ruggedness

This variable measures the difference in average elevation in any given square kilometer area and the average elevations of each of its neighboring square kilometer areas. We calculate ruggedness of the Congressmember's county (state) of birth using GIS data from Shaver, Carter, and Shawa 2019. We standardize this measure to have mean 0 and standard deviation 1.

Born on River

This dichotomous indicator is coded as 1 if the Congressmember's county (state) of birth contains a steamboat-navigated river and 0 otherwise. Data are from Atack 2015.

Born on Canal

This dichotomous indicator is coded as 1 if the Congressmember's county (state) of birth contains a canal and 0 otherwise. Data are from Atack 2017.

Year of Birth

This variable records the Congressmember's year of birth. We standardize this measure to have mean 0 and standard deviation 1. Data are from Lewis, Poole, Rosenthal et al. 2021.

Attended College

This dichotomous variable takes 1 if the Congressmember attended college and 0 otherwise. Data are from Swift, Brookshire, Canon et al. 2009.

Served in U.S. Military

This dichotomous variable takes 1 if the Congressmember served in the U.S. Military and 0 otherwise. Military service excludes service in the Confederate Army. Data are from Swift, Brookshire, Canon et al. 2009.

D2 Information about Textual Sources and Data Processing

Identifying Singular Usage

We identify singular usage by examining subject-verb agreement.[1] That is, a noun in the singular form must take a verb in singular form, and a noun in the plural form must take a verb in plural form. In each case, the term "United States" must appear in the subject position of a sentence. We identify instances of grammatical subjecthood and examine three common verb pairs that are unambiguously singular or plural. These pairs are: is/are, has/have, and was/were.

We use automated text mining processes to capture plural/singular usage in our textual sources. However, our approach goes beyond conducting simple keyword searches for phrases such as "the United States [is/are/has/have/was/were]." In particular, we confront two methodological challenges. First, we must isolate mentions of the United States as a grammatical subject, which are relatively uncommon. This task requires excluding object mentions, nouns such as the "Supreme Court of the United States," and compound subjects such as "Mexico and the United States." Second, our corpuses vary considerably in terms of OCR quality, a problem known to computer scientists and digital humanities scholars.[2]

To address the first issue, we discard all hits in which a preposition follows the search phrase; the words "of the" precede the search phrase; or the search phrase appears as part of pre-defined list of titles, names, and compound subjects. To address the second issue, we program flexible search patterns using regular expressions to allow for common typographic errors in the OCR conversion process. As an additional check, we manually review all remaining results and discarded mentions that were not in the grammatical subject position.[3]

Harmonizing Newspaper Names

To our knowledge, the proprietary AHP and NCNP datasets contain mutually exclusive newspaper content. However, the corpuses are not mutually exclusive in terms of publication coverage. For example, both datasets include content from different years of the *Daily Inter Ocean*, a paper headquartered in Chicago. To ensure that our publication random effects and fixed effects accurately

[1] Santin, Murphy, and Wilkens 2016; Myers 2008.

[2] van Strien, Beelen, Ardanuy et al. 2020; Hill and Hengchen 2019.

[3] We follow the same procedure when searching for mentions of the Confederacy.

capture publications, we harmonize newspapers names and generate unique publication-specific numeric identifiers.

We caution readers that the harmonization process is imperfect. Straightforward cases involve minor differences in the string identifiers (e.g., *Daily Inter Ocean* versus *Daily Inter Ocean (Chicago, IL)*). We also treat different editions of the same newspaper as a single publication (e.g., daily versus weekly editions). Whenever possible, we account for name changes when the publication itself remains unchanged. We do not, however, combine newspapers under a single publication ID due to mergers between separate papers.

Descriptive Statistics by Data Source

The tables below report descriptive statistics for the four textual corpuses that appear in our analysis: *America's Historical Newspapers* (AHP), *Nineteenth Century Newspapers* (NCNP), the *Congressional Globe*, and *Congressional Record*. Both tables report descriptives using the full set of observations available in the newspaper and speech datasets.

Table D1: Descriptive Statistics for Newspaper Data

	AHP	**NCNP**	**Total**
Coverage	1800–1899	1800–1899	1800–1899
Publications	320	116	411
Documents	1,755	2,299	4,054
Singular mentions	580	1,272	1,852
Total mentions	2,005	2,586	4,591

Note: Newspaper data include Northern and Southern states only. Border states and Washington, D.C. are excluded. Because some publications appear in both databases, the total number of publications does not equal the sum of the number that appear in each database.

Table D2: Descriptive Statistics for Congressional Speech Data

	Globe	***Record***	**Total**
Coverage	1851–1873	1873–1899	1851–1899
Congresses	32^{nd}–42^{nd}	43^{rd}–55^{th}	32^{rd}–55^{th}
Congressmembers	215	392	556
Singular mentions	312	1,684	1,996
Total mentions	722	2,178	2,900

Note: Congressional data include Northern delegations only. Because Congressmembers can appear in multiple Congresses, the total number of unique Congressmembers does not equal the sum of the number that appear in each database.

D3 Modernization as a Potential Confounding Factor

Our covariates W attempt to account for potentially confounding variables that are plausibly correlated with political geography, time, and the adoption of the grammatical singular. In the 19th century, the most likely confounders relate to processes of economic and political modernization. From a temporal perspective, scholars have argued that modernization plays a critical role in weakening local ties and paving the way for citizens to "imagine" a larger, nationally-based political community. For example, industrialization and urbanization brought individuals out of their agrarian milieu and exposed them to the homogenizing forces of the city.[4] Similarly, the advent and spread of print capitalism helped to create a sense of shared national community,[5] while infrastructure – often federally-provided – served a similar purpose by fostering physical access and connectedness on a national scale.[6] Finally, the increased contact with central (as opposed to local) institutions that comes with political modernization may facilitate an ideational reorientation of the citizen toward a larger political unit.[7] In these also ways, the "experience" of modernization may shape how individuals view the proper locus of sovereign authority.[8]

In the American context, approaches towards modernization also frequently took on sectional dimensions. For example, one of the most contentious issues of the time was national tariff policy: protectionism benefited the manufacturing interests concentrated in the East at the expense of agrarian interests in the South and Midwest. Similarly, Americans fiercely debated whether the federal government should fund and construct "internal improvements" (i.e., transportation infrastructure), which advanced the interests of eastern capitalism and western yeoman agriculture.[9] These issues were intimately bound up in questions about the authority of the national government to regulate policies and override the individual interests of the several states. These debates also frequently pitted the the more nationally-oriented Whigs and Republicans against their Democratic rivals.

---

[4] Gellner 1983.

[5] Holt 2019; Loughran 2007; Anderson 1983.

[6] Cermeño, Enflo, and Lindvall 2021; Herbst 2000.

[7] Zhang and Lee 2020; Englebert 2009; Gennaioli and Rainer 2007; Bockstette, Chanda, and Putterman 2002; Hechter 2000.

[8] Bensel 1990.

[9] Bensel 1990, 66.

We operationalize modernization in several different ways using historical and geospatial data, aggregated to the county level. URBANIZATION measures the percentage of the urban population in the county using census data.[10] To capture variation in support for internal improvements, we code variables using geospatial data that capture whether the county contains a steamboat-navigated RIVER or a CANAL as well as the county's RUGGEDNESS as a proxy for physical accessibility.[11] Because the presence of state administrative infrastructure increases state power, we combine archival data with data from Acemoglu, Moscona, and Robinson (2016) to capture the density of POST OFFICES, the government agency Americans were most likely to encounter in their daily lives.[12]

We note that many of our covariates come from the U.S. census. For our period, counties are the smallest geographic unit for which census data are reported. Regarding our newspaper data, we expect that the individuals contributing content to newspapers reside either in the headquarters city or in the surrounding locality, and that singular usage therefore reflects the characteristics of that locality. Although we lack information on the precise spatial extent of each publication's readership, we believe that counties offer a reasonable approximation. We therefore code newspaper content to the county that existed at the time of publication, based on the city in which the newspaper is headquartered. In our robustness checks, we also aggregate newspaper data and covariates to the state level (Appendix Section D7). Results are unchanged.

To draw valid inferences from our speech corpus, we must address another issue with our data: the entry and exit of individual speakers from our dataset. This issue arises from the high level of turnover in Congress during this period in combination with our need to restrict attention to U.S. mentions in the grammatical subject position.[13] One may worry that Congressmembers elected prior to the Civil War are not comparable to the more professionalized, careerist Congressmembers in office during the late 19th century. In order to pool mentions from individuals who appear in

[10] Manson, Schroeder, Van Riper et al. 2020.

[11] Atack 2015, Atack 2017, Shaver, Carter, and Shawa 2019, respectively. Although we would prefer to account for manufacturing directly (it is correlated with urbanization), the U.S. census did not start collecting this in information in a comparable format until 1850.

[12] Blevins 2021; Rogowski, Gerring, Maguire et al. 2021; Jensen and Ramey 2020; Acemoglu, Moscona, and Robinson 2016.

[13] Swain, Borrelli, Reed et al. 2000; Kernell 1977.

only one Congress with those from individuals who contribute mentions in multiple Congresses, we must account for differences that could influence their usage of the singular.

We focus in particular on variables related to a Congressmember's life experiences and the social milieu of the area where he was born. We identify each Congressmember's place and year of birth using the official biographies contained in the *Biographical Directory of Congress* and use this information to assign birthplaces to the counties that existed at the time of birth.[14] We then code the same set of variables from the newspaper analysis at the county-decade level (matching birth year to the nearest decennial census year).[15] To be clear, these variables reflect birthplace characteristics, not contemporaneous characteristics at the time a statement is observed. In addition, we also account for biographical experiences that could influence imagined sovereignty by coding individual-level controls to account for whether the Congressmember was COLLEGE EDUCATED; whether he served in the U.S. MILITARY, defined as military service other than Confederate service; and his BIRTH YEAR.[16]

As with the newspaper covariates, the locality of birth is not a precise areal unit. As before, we assume that the county of birth offers a reasonable approximation. Appendix Section D7 replicates our speech analyses using birthplace characteristics aggregated to the state level. Results are unchanged.

[14] United States Congress 2020.

[15] Unfortunately, we lack county-level information on individuals born outside the United States. These are excluded from the analysis.

[16] United States Congress 2020; Swift, Brookshire, Canon et al. 2009.

D4 Full Regression Results from the Main Text

Table D3: Singular Usage in Northern Newspapers before and after 1860

	(1)	(2)	(3)	(4)
Year	0.006***	0.004*	0.006***	0.003+
	(0.001)	(0.002)	(0.001)	(0.002)
After 1860	0.090	0.096	0.082	0.078
	(0.062)	(0.082)	(0.061)	(0.079)
North	-0.020	0.000	0.000	0.000
	(0.034)	(.)	(0.044)	(.)
Year × After 1860	0.001	0.003	0.002	0.003
	(0.003)	(0.005)	(0.003)	(0.004)
Year × North	-0.000	0.001	0.000	-0.000
	(0.001)	(0.002)	(0.001)	(0.002)
After 1860 × North	-0.086	-0.133	-0.086	-0.110
	(0.072)	(0.090)	(0.072)	(0.085)
Year × After 1860 × North	0.007*	0.005	0.006+	0.006
	(0.003)	(0.005)	(0.003)	(0.005)
County % Urban (std)			0.010	0.092+
			(0.011)	(0.053)
Post Office Density (std)			-0.013	0.028
			(0.012)	(0.019)
Terrain Ruggedness (std)			0.009	0.021
			(0.010)	(0.124)
County on River			0.021	0.167**
			(0.023)	(0.064)
County on Canal			-0.007	-0.054
			(0.029)	(0.040)
Constant	0.358***	0.354***	0.346***	0.314***
	(0.029)	(0.030)	(0.039)	(0.035)
N	4578	4578	4578	4578
Fixed Effects	No	Publication	No	Publication

Note: Full table for results reported in Figure 3. Dependent variable is SINGULAR usage. Linear probability model with standard errors clustered within newspapers. Year is centered such that 0 corresponds to the year 1860.
$^{+}p < 0.10$,* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table D4: Singular Usage in Southern Newspapers before and after 1865

	(1)	(2)	(3)	(4)
Year	0.007***	0.005**	0.007***	0.004**
	(0.001)	(0.002)	(0.001)	(0.002)
After 1865	0.104	0.071	0.100	0.045
	(0.071)	(0.093)	(0.072)	(0.088)
North	-0.045	0.000	-0.026	0.000
	(0.044)	(.)	(0.054)	(.)
Year × After 1865	0.000	0.003	0.000	0.003
	(0.003)	(0.006)	(0.003)	(0.005)
Year × North	-0.001	-0.001	-0.001	-0.002
	(0.001)	(0.002)	(0.001)	(0.002)
After 1865 × North	-0.004	-0.019	-0.007	0.019
	(0.079)	(0.106)	(0.081)	(0.101)
Year × After 1865 × North	0.007+	0.005	0.006	0.005
	(0.004)	(0.006)	(0.004)	(0.006)
County % Urban (std)			0.011	0.093+
			(0.011)	(0.054)
Post Office Density (std)			-0.012	0.030
			(0.012)	(0.020)
Terrain Ruggedness (std)			0.009	0.008
			(0.010)	(0.119)
County on River			0.022	0.167**
			(0.023)	(0.064)
County on Canal			-0.006	-0.052
			(0.029)	(0.038)
Constant	0.407***	0.369***	0.394***	0.321***
	(0.034)	(0.028)	(0.043)	(0.037)
N	4578	4578	4578	4578
Fixed Effects	No	Publication	No	Publication

Note: Full table for results reported in Figure 5. Dependent variable is SINGULAR usage. Linear probability model with standard errors clustered within newspapers. Year is centered such that 0 corresponds to the year 1865.
$^{+}p < 0.10$,* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table D5: Singular Usage among Northern Congressmen, by Party

	(1)	(2)
Year of Speech	0.015	0.016
	(0.030)	(0.029)
After 1860	0.248*	0.248*
	(0.109)	(0.109)
Democrat	0.188	0.173
	(0.154)	(0.154)
After 1860 × Year of Speech	-0.002	-0.003
	(0.030)	(0.030)
Democrat × Year of Speech	0.012	0.011
	(0.035)	(0.035)
After 1860 × Democrat	-0.220	-0.220
	(0.170)	(0.170)
After 1860 × Democrat × Year of Speech	-0.013	-0.011
	(0.035)	(0.035)
County of Birth: % Urban (std)		-0.021
		(0.015)
County of Birth: Post Office Density (std)		-0.010
		(0.016)
County of Birth: Terrain Ruggedness (std)		-0.020
		(0.013)
Born on River		0.041
		(0.028)
Born on Canal		0.028
		(0.029)
Year of Birth (std)		-0.004
		(0.016)
Attended College		-0.043+
		(0.025)
Served in US Military		0.006
		(0.027)
Constant	0.218*	0.221*
	(0.106)	(0.111)
N	2705	2705

Note: Full table for results reported in Figure 6. Dependent variable is SINGULAR usage. Hierarchical linear model with cross-nested random effects (year and speaker). Year is centered such that 0 corresponds to the year 1860.
$^{+}p < 0.10$,* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table D6: Singular Usage in Northern Newspapers, by 1864 Election Results

	(1)	(2)	(3)	(4)
Year	0.006***	0.006**	0.007***	0.005*
	(0.000)	(0.002)	(0.001)	(0.003)
After 1860	-0.002	-0.058	-0.005	-0.062
	(0.048)	(0.046)	(0.044)	(0.046)
Lincoln County	-0.013	0.000	-0.027	0.000
	(0.033)	(.)	(0.041)	(.)
Year × After 1860	0.006***	0.003	0.005*	0.004
	(0.002)	(0.002)	(0.002)	(0.002)
Year × Lincoln County	-0.001	-0.003	-0.001	-0.003
	(0.001)	(0.002)	(0.001)	(0.003)
After 1860 × Lincoln County	0.008	0.058	0.005	0.064
	(0.065)	(0.055)	(0.062)	(0.057)
Year × After 1860 × Lincoln County	0.004	0.009**	0.005+	0.008*
	(0.003)	(0.003)	(0.003)	(0.004)
County % Urban (std)			0.014	0.066
			(0.012)	(0.069)
Post Office Density (std)			-0.016	0.015
			(0.013)	(0.016)
Terrain Ruggedness (std)			0.010	-0.019
			(0.011)	(0.117)
County on River			-0.024	0.034
			(0.030)	(0.111)
County on Canal			-0.002	-0.021
			(0.032)	(0.053)
Constant	0.346***	0.345***	0.368***	0.344***
	(0.020)	(0.023)	(0.025)	(0.037)
N	3579	3579	3579	3579
Fixed Effects	No	Publication	No	Publication

Note: Full table for results reported in Figures 7 and S5. Dependent variable is SINGULAR usage. Linear probability model with standard errors clustered within newspapers. Year is centered such that 0 corresponds to the year 1860.
$^{+}p < 0.10$,* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

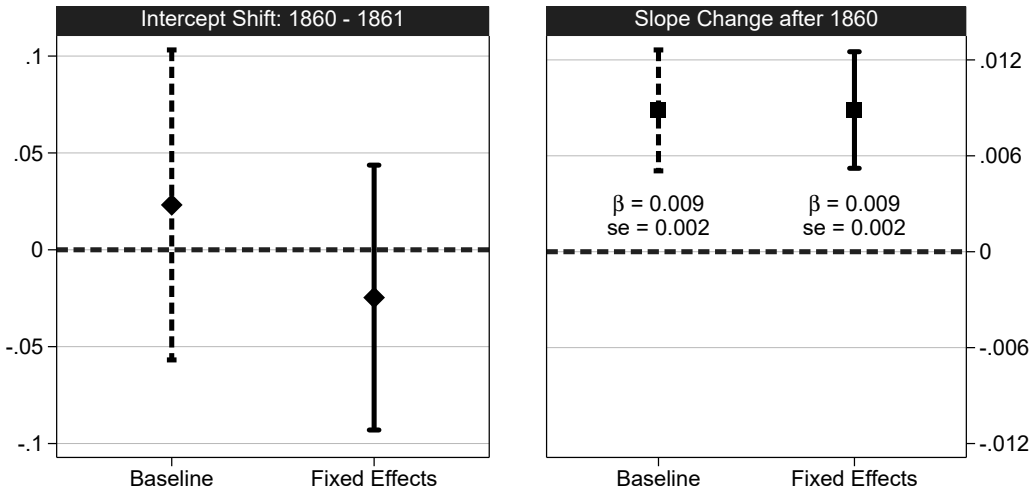
D5 Robustness: Including Border States with the North

Our sample includes newspaper mentions and Congressional speech mentions from the five border states: Delaware, Maryland, Missouri, Kentucky, and West Virginia. These states permitted the institution of slavery but did not – or were unable to – participate in the Civil War on the side of the Confederacy.[17] We also code Washington, D.C. as a "border" state given its geographic proximity to Maryland (a border state) and Virginia (a Confederate state).

To ensure that our sampling decisions do not drive our results, we replicate our analyses with an expanded definition of the "North" that includes the border states and Washington, D.C., Figures D1, D2, and D3 show that our results are substantively unchanged.

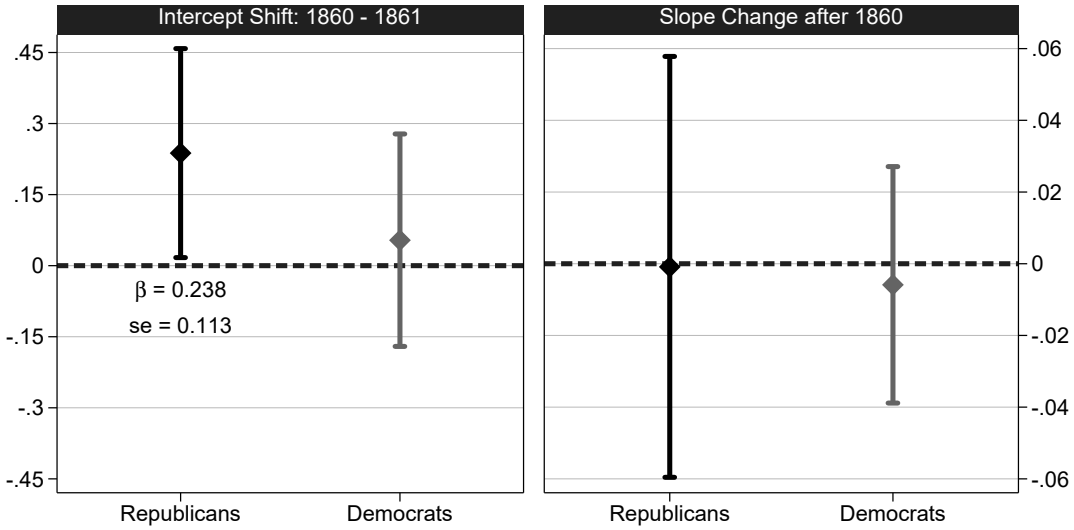
[17] West Virginia seceded from Virginia after Virginia seceded from the Union.

Figure D1: Replication of Results Reported in Figure 3 including Border States in the North



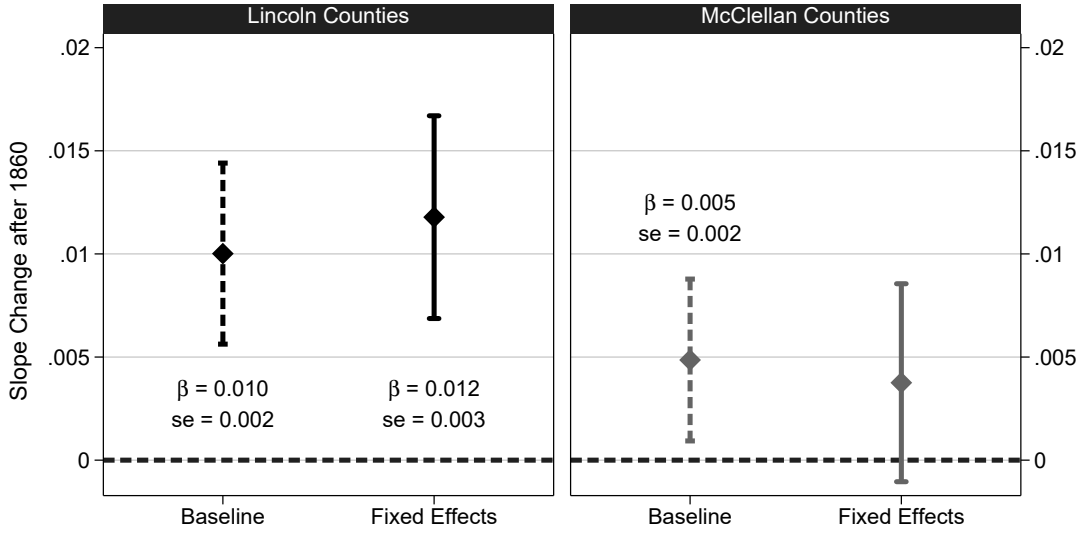
Note: The figure displays coefficients, linear combinations, and 95% confidence intervals from Equation 1, centering Y at the year 1860 and setting $P = 1$ if $Y > 0$. Results are displayed for Northern and Border states only. Full output is provided in Table D7.

Figure D2: Replication of Results Reported in Figure 6 including Border States in the North



Note: The figure displays coefficients, linear combinations, and 95% confidence intervals from Equation 2, centering Y at the year 1860 and setting $P = 1$ if $Y > 0$. Results are for Congressmen from Northern and Border states only. Full output is provided in Table D8.

Figure D3: Replication of Results Reported in Figure 7 including Border States in the North

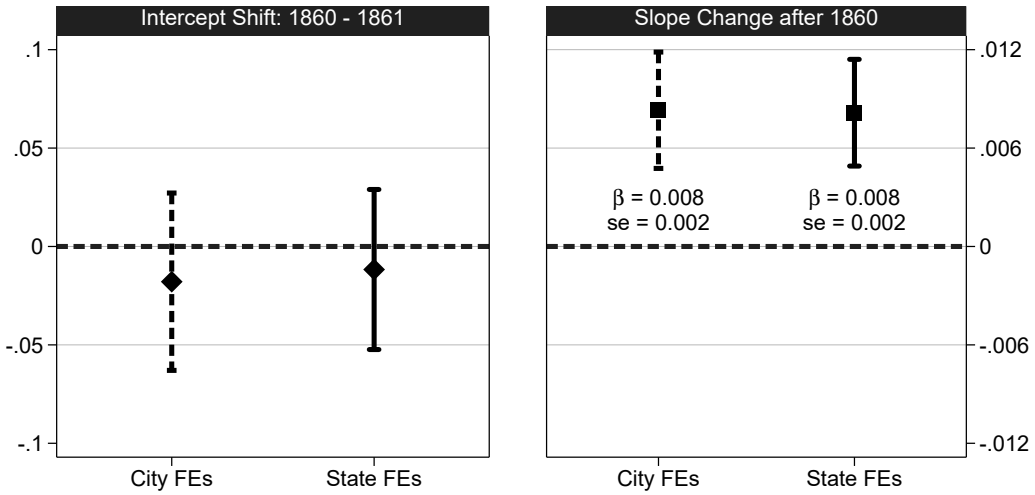


Note: The figure displays coefficients, linear combinations, and 95% confidence intervals from modified Equation 1, where the Northern newspaper dummy variable (N) is replaced with an indicator ($LINCOLN$) for whether Lincoln won the county in which the newspaper is headquartered. Y is centered at the year 1860 and $P = 1$ if $Y > 0$. Results are for counties located in Northern and Border states only. Full output is provided in Table D9.

D6 Robustness: City and State Fixed Effects

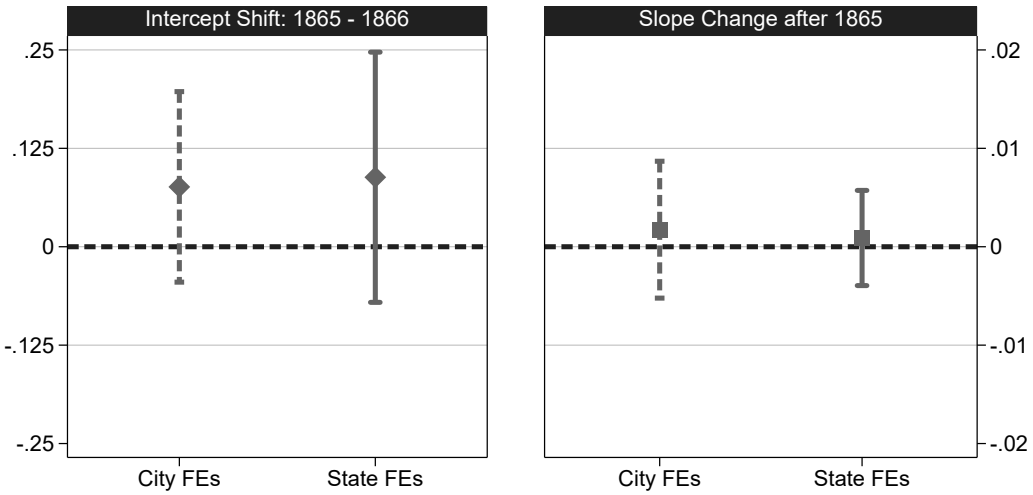
Our baseline models include standard errors clustered by publication, and our secondary models introduce publication fixed effects. However, readers may be concerned about unobserved time-invariant heterogeneity across cities or states rather than publications. We examine this possibility by replication our secondary models with city fixed effects or state fixed effects. Figures D4 – D6 display the results, which are substantively similar to those reported in main text Figures 3, 5, and 7.

Figure D4: Replication with City and State FEs of Results Reported in Figure 3



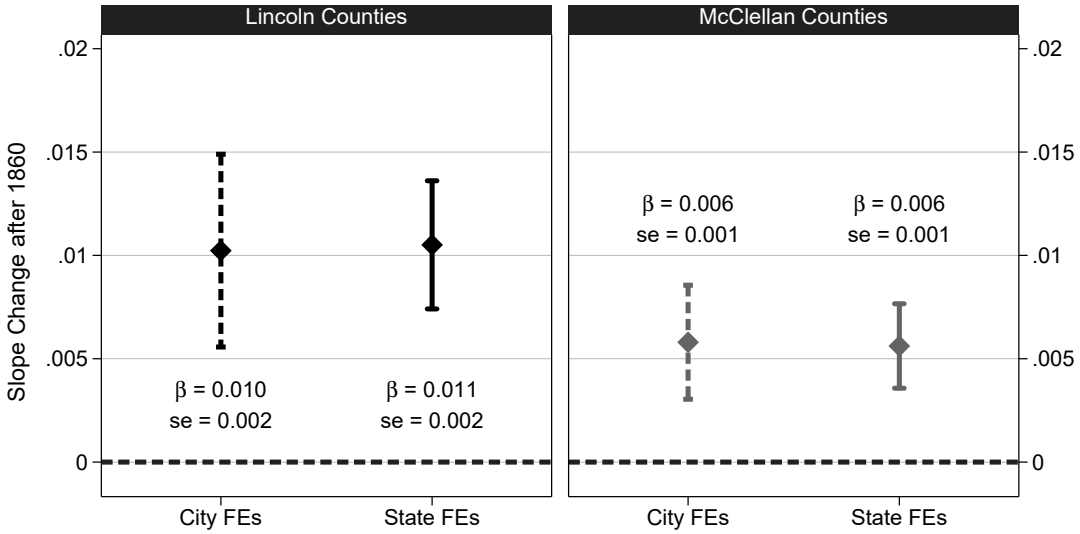
Note: The figure displays coefficients, linear combinations, and 95% confidence intervals from Equation 1, centering Y at the year 1860 and setting $P = 1$ if $Y > 0$. Results are displayed for Northern states only. Full output is provided in Table D10.

Figure D5: Replication with City and State FEs of Results Reported in Figure 5



Note: The figure displays coefficients, linear combinations, and 95% confidence intervals from Equation 1, centering Y at the year 1865 and setting $P = 1$ if $Y > 0$. Results are displayed for Southern states only. Full output is provided in Table D11.

Figure D6: Replication with City and State FEs of Results Reported in Figure 7



Note: The figure displays coefficients, linear combinations, and 95% confidence intervals from modified Equation 1, where the Northern newspaper dummy variable (N) is replaced with an indicator ($LINCOLN$) for whether Lincoln won the county in which the newspaper is headquartered. Y is centered at the year 1860 and $P = 1$ if $Y > 0$. Results are for Northern counties only. Full output is provided in Table D12.

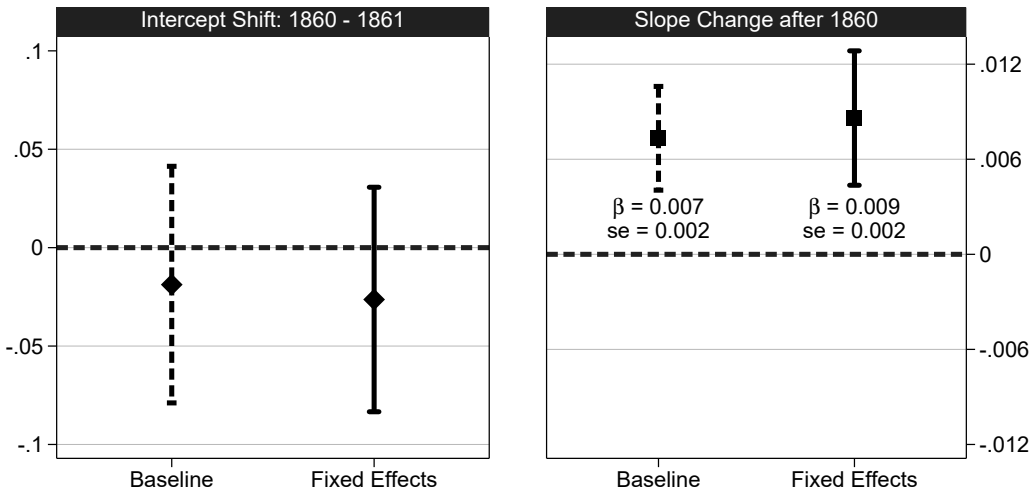
D7 Robustness: States as Areal Units

As we describe in Appendix Section D3, we assume that singular usage may be a function of locality characteristics related to economic and political modernization. In the newspaper data, the locality can include the city in which the newspaper is headquartered or its surrounding area; in the speech data, the locality refers to the area in which a Congressmember was born. In neither case is locality a precise areal unit.

We treat counties as a reasonable approximation of locality and operationalize our covariates at the county level. The county is typically the smallest geographic unit at which census data (an important data source for our covariates) are reported. Yet when there is no theoretical reason to prefer a particular areal unit, regression results may be sensitive to the choice of that unit – an issue known as the modifiable areal unit problem.[18]

To ensure that our results are not driven by the selection of the county as the spatial unit at which our covariates are operationalized, we rerun our analyses using covariates calculated at the state level. Figures D7, D8, D9, and D10 report the results. Our conclusions are unchanged.

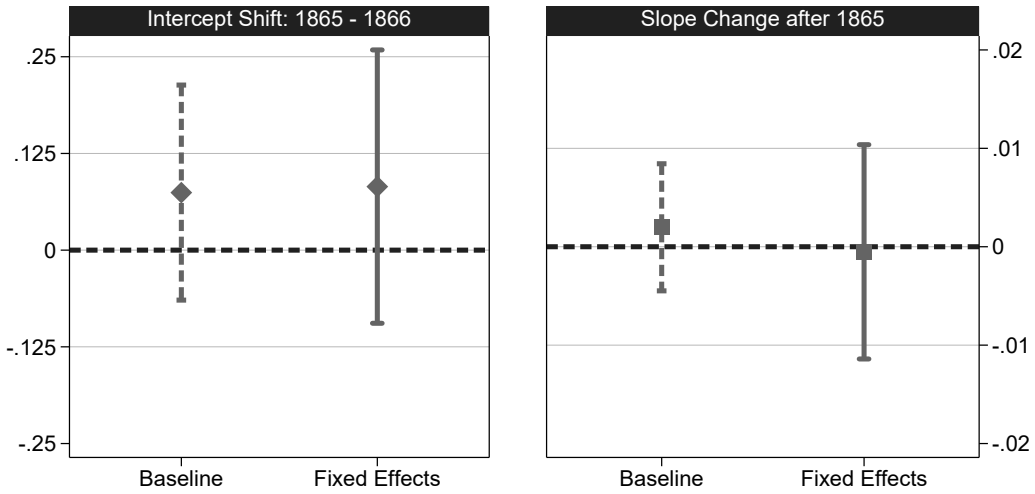
Figure D7: Replication with State-Level Covariates of Results Reported in Figure 3



Note: The figure displays coefficients, linear combinations, and 95% confidence intervals from Equation 1, centering Y at the year 1860 and setting $P = 1$ if $Y > 0$. Results are displayed for Northern states only. Full output is provided in Table D13.

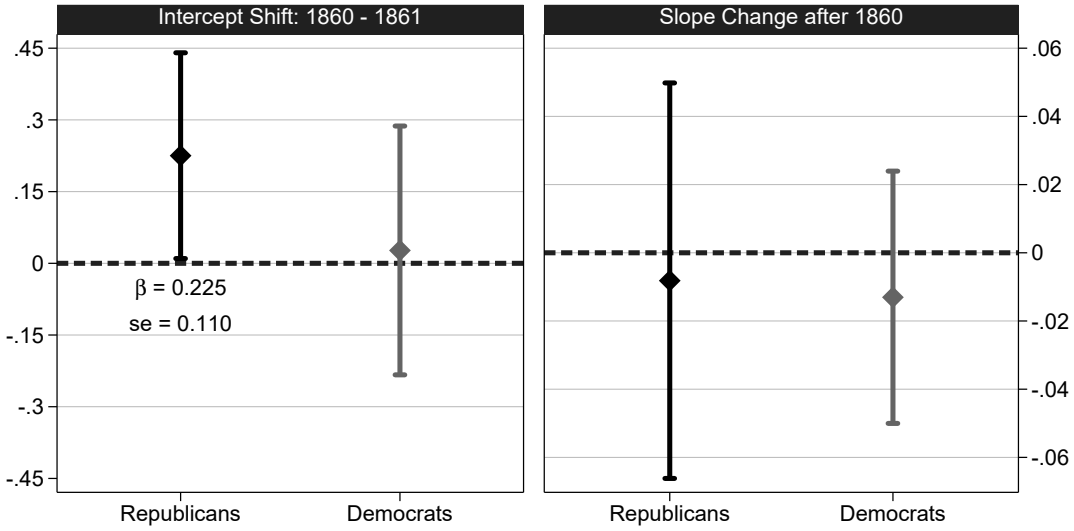
[18] Lee, Rogers, and Soifer 2022.

Figure D8: Replication with State-Level Covariates of Results Reported in Figure 5



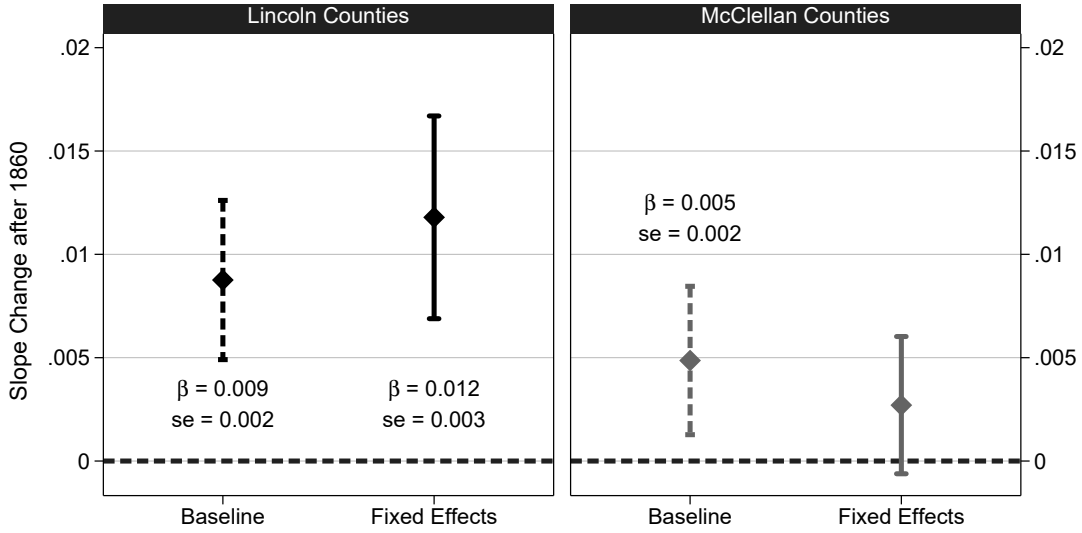
Note: The figure displays coefficients, linear combinations, and 95% confidence intervals from Equation 1, centering Y at the year 1865 and setting $P = 1$ if $Y > 0$. Results are displayed for Southern states only. Full output is provided in Table D14.

Figure D9: Replication with State-Level Covariates of Results Reported in Figure 6



Note: The figure displays coefficients, linear combinations, and 95% confidence intervals from Equation 2, centering Y at the year 1860 and setting $P = 1$ if $Y > 0$. Results are for Northern Congressmen only. Full output is provided in Table D15.

Figure D10: Replication with State-Level Covariates of Results Reported in Figure 7



Note: The figure displays coefficients, linear combinations, and 95% confidence intervals from modified Equation 1, where the Northern newspaper dummy variable (N) is replaced with an indicator ($LINCOLN$) for whether Lincoln won the county in which the newspaper is headquartered. Y is centered at the year 1860 and $P = 1$ if $Y > 0$. Results are for Northern counties only. Full output is provided in Table D16.

D8 Full Regression Results from the Dataverse Appendix

Table D7: Regression Output for Results Reported in Figure D1

	(1)	(2)
Year	0.006***	0.003+
	(0.001)	(0.002)
After 1860	0.087	0.080
	(0.062)	(0.079)
North	-0.034	0.000
	(0.047)	(.)
Year × After 1860	0.002	0.003
	(0.003)	(0.004)
Year × North	-0.001	-0.000
	(0.001)	(0.002)
After 1860 × North	-0.064	-0.105
	(0.074)	(0.085)
Year × After 1860 × North	0.007*	0.006
	(0.003)	(0.005)
County % Urban (std)	0.004	0.102*
	(0.009)	(0.048)
Post Office Density (std)	-0.003	0.027+
	(0.011)	(0.016)
Terrain Ruggedness (std)	0.009	-0.016
	(0.009)	(0.098)
County on River	0.022	0.130**
	(0.022)	(0.043)
County on Canal	0.004	-0.036
	(0.027)	(0.034)
Constant	0.349***	0.299***
	(0.038)	(0.023)
N	5444	5444
Fixed Effects	No	Publication

Note: Full table for results reported in Figure D1. Dependent variable is SINGULAR usage. Border states are included with the North. Linear probability model with standard errors clustered within newspapers. Year is centered such that 0 corresponds to the year 1860.
$^{+}p < 0.10$,* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table D8: Regression Output for Results Reported in Figure D2

	(1)
Year of Speech	0.014
	(0.030)
After 1860	0.238*
	(0.113)
Democrat	0.135
	(0.143)
After 1860 × Year of Speech	-0.001
	(0.030)
Democrat × Year of Speech	0.004
	(0.033)
After 1860 × Democrat	-0.184
	(0.154)
After 1860 × Democrat × Year of Speech	-0.005
	(0.034)
County of Birth: % Urban (std)	0.003
	(0.013)
County of Birth: Post Office Density (std)	-0.027+
	(0.014)
County of Birth: Terrain Ruggedness (std)	-0.018
	(0.012)
Born on River	0.039
	(0.027)
Born on Canal	0.023
	(0.027)
Year of Birth (std)	0.012
	(0.016)
Attended College	-0.024
	(0.024)
Served in US Military	-0.004
	(0.026)
Constant	0.222*
	(0.113)
N	3081

Note: Full table for results reported in Figure D2. Dependent variable is SINGULAR usage. Hierarchical linear model with cross-nested random effects (year and speaker). Year is centered such that 0 corresponds to the year 1860. Sample includes Congressmen from Northern and Border states only.

$^{+}p < 0.10$,* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table D9: Regression Output for Results Reported in Figure D3

	(1)	(2)
Year	0.007***	0.005*
	(0.001)	(0.003)
After 1860	-0.005	-0.064
	(0.044)	(0.046)
Lincoln County	-0.028	0.000
	(0.041)	(.)
Year × After 1860	0.005*	0.004
	(0.002)	(0.002)
Year × Lincoln County	-0.001	-0.003
	(0.001)	(0.003)
After 1860 × Lincoln County	0.005	0.066
	(0.062)	(0.057)
Year × After 1860 × Lincoln County	0.005+	0.008*
	(0.003)	(0.004)
County % Urban (std)	0.014	0.083
	(0.011)	(0.066)
Post Office Density (std)	-0.019	0.014
	(0.013)	(0.016)
Terrain Ruggedness (std)	0.010	-0.021
	(0.009)	(0.118)
County on River	-0.028	-0.020
	(0.028)	(0.108)
County on Canal	0.000	-0.014
	(0.030)	(0.052)
Constant	0.370***	0.353***
	(0.026)	(0.037)
N	3717	3717
Fixed Effects	No	Publication

Note: Full table for results reported in Figure D3. Dependent variable is SINGULAR usage. Border states are included with the North. Linear probability model with standard errors clustered within newspapers. Year is centered such that 0 corresponds to the year 1860. Sample includes Northern and Border counties only.
$^{+}p < 0.10$,* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table D10: Regression Output for Results Reported in Figure D4

	(1)	(2)
Year	0.004***	0.005***
	(0.001)	(0.001)
After 1860	0.099	0.095
	(0.062)	(0.067)
North	0.000	0.000
	(.)	(.)
Year × After 1860	0.002	0.002
	(0.003)	(0.003)
Year × North	-0.000	-0.000
	(0.001)	(0.001)
After 1860 × North	-0.117+	-0.107
	(0.067)	(0.070)
Year × After 1860 × North	0.006	0.006*
	(0.004)	(0.003)
County % Urban (std)	0.061+	0.029+
	(0.037)	(0.015)
Post Office Density (std)	0.020	0.017
	(0.014)	(0.011)
Terrain Ruggedness (std)	0.135	0.021
	(0.085)	(0.018)
County on River	0.119**	0.032
	(0.045)	(0.023)
County on Canal	-0.056	-0.037
	(0.048)	(0.025)
Constant	0.327***	0.347***
	(0.027)	(0.023)
N	4578	4578
Fixed Effects	City	State

Note: Full table for results reported in Figure D4. Dependent variable is SINGULAR usage. Linear probability model with standard errors clustered within cities and states, respectively. Year is centered such that 0 corresponds to the year 1860.
$^{+}p < 0.10$,* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table D11: Regression Output for Results Reported in Figure D5

	(1)	(2)
Year	0.005***	0.006***
	(0.001)	(0.001)
After 1865	0.076	0.088
	(0.062)	(0.081)
North	0.000	0.000
	(.)	(.)
Year × After 1865	0.002	0.001
	(0.004)	(0.002)
Year × North	-0.001	-0.001
	(0.001)	(0.001)
After 1865 × North	0.010	0.010
	(0.075)	(0.087)
Year × After 1865 × North	0.005	0.005+
	(0.004)	(0.003)
County % Urban (std)	0.063+	0.026
	(0.038)	(0.016)
Post Office Density (std)	0.023	0.018
	(0.016)	(0.012)
Terrain Ruggedness (std)	0.103	0.015
	(0.071)	(0.018)
County on River	0.117**	0.028
	(0.043)	(0.022)
County on Canal	-0.048	-0.033
	(0.046)	(0.025)
Constant	0.342***	0.367***
	(0.032)	(0.027)
N	4578	4578
Fixed Effects	City	State

Note: Full table for results reported in Figure D5. Dependent variable is SINGULAR usage. Linear probability model with standard errors clustered within cities and states, respectively. Year is centered such that 0 corresponds to the year 1865.
$^{+}p < 0.10$,* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table D12: Regression Output for Results Reported in Figure D6

	(1)	(2)
Year	0.005***	0.005***
	(0.001)	(0.001)
After 1860	-0.008	-0.002
	(0.029)	(0.022)
Lincoln County	0.000	0.011
	(.)	(0.046)
Year × After 1860	0.006***	0.006***
	(0.001)	(0.001)
Year × Lincoln County	-0.001	-0.001
	(0.001)	(0.001)
After 1860 × Lincoln County	-0.015	-0.015
	(0.036)	(0.028)
Year × After 1860 × Lincoln County	0.004	0.005*
	(0.003)	(0.002)
County % Urban (std)	0.046	0.026
	(0.052)	(0.019)
Post Office Density (std)	0.012	0.009
	(0.016)	(0.012)
Terrain Ruggedness (std)	0.092	0.009
	(0.085)	(0.024)
County on River	0.036	0.009
	(0.060)	(0.034)
County on Canal	-0.043	-0.027
	(0.056)	(0.026)
Constant	0.348***	0.338***
	(0.026)	(0.028)
N	3579	3579
Fixed Effects	City	State

Note: Full table for results reported in Figure D6. Dependent variable is SINGULAR usage. Linear probability model with standard errors clustered within cities and states, respectively. Year is centered such that 0 corresponds to the year 1860. Sample includes Northern counties only.

$^{+}p < 0.10$,* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table D13: Regression Output for Results Reported in Figure D7

	(1)	(2)
Year	0.006***	0.005*
	(0.001)	(0.002)
After 1860	0.070	0.107
	(0.060)	(0.082)
North	0.057	0.000
	(0.044)	(.)
Year × After 1860	0.003	0.001
	(0.003)	(0.004)
Year × North	0.001	-0.002
	(0.001)	(0.002)
After 1860 × North	-0.089	-0.134
	(0.068)	(0.089)
Year × After 1860 × North	0.005	0.008
	(0.003)	(0.005)
State % Urban (std)	-0.004	-0.018
	(0.012)	(0.058)
State Post Office Density (std)	-0.040**	0.067
	(0.013)	(0.058)
State Terrain Ruggedness (std)	-0.001	-0.172
	(0.010)	(0.145)
State on River	-0.024	0.006
	(0.023)	(0.049)
State on Canal	-0.028	-0.003
	(0.018)	(0.046)
Constant	0.358***	0.342***
	(0.043)	(0.063)
N	4587	4587
Fixed Effects	No	Publication

Note: Full table for results reported in Figure D7. Dependent variable is SINGULAR usage. Linear probability model with standard errors clustered within newspapers. Year is centered such that 0 corresponds to the year 1860.
$^{+}p < 0.10$,* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table D14: Regression Output for Results Reported in Figure D8

	(1)	(2)
Year	0.007***	0.005**
	(0.001)	(0.002)
After 1865	0.074	0.082
	(0.071)	(0.090)
North	0.025	0.000
	(0.049)	(.)
Year × After 1865	0.002	-0.001
	(0.003)	(0.006)
Year × North	0.000	-0.003
	(0.001)	(0.002)
After 1865 × North	0.012	-0.007
	(0.079)	(0.102)
Year × After 1865 × North	0.004	0.008
	(0.004)	(0.006)
State % Urban (std)	-0.007	-0.028
	(0.013)	(0.056)
State Post Office Density (std)	-0.037**	0.081
	(0.013)	(0.060)
State Terrain Ruggedness (std)	-0.000	-0.207
	(0.009)	(0.140)
State on River	-0.022	0.004
	(0.023)	(0.048)
State on Canal	-0.023	-0.003
	(0.017)	(0.046)
Constant	0.407***	0.352***
	(0.047)	(0.067)
N	4587	4587
Fixed Effects	No	Publication

Note: Full table for results reported in Figure D8. Dependent variable is SINGULAR usage. Linear probability model with standard errors clustered within newspapers. Year is centered such that 0 corresponds to the year 1865.
$^{+}p < 0.10$,* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table D15: Regression Output for Results Reported in Figure D9

	(1)
Year of Speech	0.021
	(0.030)
After 1860	0.225*
	(0.110)
Democrat	0.149
	(0.154)
After 1860 × Year of Speech	-0.008
	(0.030)
Democrat × Year of Speech	0.004
	(0.035)
After 1860 × Democrat	-0.198
	(0.170)
After 1860 × Democrat × Year of Speech	-0.005
	(0.035)
State of Birth: % Urban (std)	0.008
	(0.021)
State of Birth: Post Office Density (std)	-0.006
	(0.024)
State of Birth: Terrain Ruggedness (std)	-0.025
	(0.015)
Born on River	0.056
	(0.053)
Born on Canal	-0.005
	(0.031)
Year of Birth (std)	-0.009
	(0.019)
Attended College	-0.047+
	(0.025)
Served in US Military	0.015
	(0.027)
Constant	0.214+
	(0.117)
N	2705

Note: Full table for results reported in Figure D9. Dependent variable is SINGULAR usage. Hierarchical linear model with cross-nested random effects (year and speaker). Year is centered such that 0 corresponds to the year 1860. Sample includes Northern Congressmen only.

$^{+}p < 0.10$,* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table D16: Regression Output for Results Reported in Figure D10

	(1)	(2)
Year	0.008***	0.006*
	(0.001)	(0.002)
After 1860	-0.015	-0.054
	(0.050)	(0.039)
Lincoln County	-0.006	0.000
	(0.030)	(.)
Year × After 1860	0.005**	0.003
	(0.002)	(0.002)
Year × Lincoln County	-0.000	-0.003
	(0.001)	(0.003)
After 1860 × Lincoln County	-0.003	0.050
	(0.061)	(0.049)
Year × After 1860 × Lincoln County	0.004	0.009**
	(0.002)	(0.003)
State % Urban (std)	-0.004	0.003
	(0.022)	(0.057)
State Post Office Density (std)	-0.039*	-0.001
	(0.017)	(0.045)
State Terrain Ruggedness (std)	-0.001	-0.041
	(0.010)	(0.214)
State on River	-0.032	-0.030
	(0.028)	(0.063)
State on Canal	-0.018	0.025
	(0.022)	(0.058)
Constant	0.410***	0.361***
	(0.045)	(0.085)
N	3588	3588
Fixed Effects	No	Publication

Note: Full table for results reported in Figure D10. Dependent variable is SINGULAR usage. Linear probability model with standard errors clustered within newspapers. Year is centered such that 0 corresponds to the year 1860. Sample includes Northern counties only.

$^{+}p < 0.10$,* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Supplementary References

Acemoglu, Daron, Jacob Moscona, and James A Robinson. 2016. State capacity and American technology: evidence from the nineteenth century. *American Economic Review* 106 (5):61–67.

Atack, Jeremy. 2015. Steamboat-Navigated Rivers During the Nineteenth Century in the United States. https://my.vanderbilt.edu/jeremyatack/data-downloads/.

———. 2017. Historical Geographic Information Systems (GIS) database of Nineteenth Century U.S. Canals. https://my.vanderbilt.edu/jeremyatack/data-downloads/.

Clubb, Jerome M., William H. Flanigan, and Nancy H. Zingale. 2006. *Electoral Data for Counties in the United States: Presidential and Congressional Races, 1840-1972*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/ICPSR08611.v1.

Gale. 2021. *Nineteenth Century U.S Newspapers*. Farmington Hills, MI: Cengage. https://www.gale.com/c/nineteenth-century-us-newspapers.

Gentzkow, Matthew, Jesse M. Shapiro, and Matt Taddy. 2018. *Congressional Record for the 43rd-114th Congresses: Parsed Speeches and Phrase Counts*. Palo Alto, CA: Stanford Libraries. https://data.stanford.edu/congress_text.

Granger, Gideon. 1802. Role of Civil, Military, and Naval Officers. In *American State Papers: Class X., Miscellaneous, Volume I*, edited by Walter Lowrie and Walter S. Franklin, 8–12. Washington: Gales and Seaton.

Lewis, Jeffrey B., Keith Poole, Howard Rosenthal, Adam Boche, Aaron Rudkin, and Luke Sonnet. 2021. Voteview: Congressional Roll-Call Votes Database. https://voteview.com/.

Manson, Steven, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles. 2020. *IPUMS National Historical Geographic Information System: Version 15.0 [dataset]*. Minneapolis, MN: IPUMS. http://doi.org/10.18128/D050.V15.0.

Osgood, Samuel. 1790. List of Post Offices, and the Receipts and Expenditures to January 5, 1790. In *American State Papers: Post Office Department*, edited by Walter Lowrie and Walter S. Franklin, 289–300. Washington: Gales and Seaton.

Phillips, Mark Edward. 2015. *Congressional Globe OCR Dataset*. Denton: University of North Texas Libraries. https://digital.library.unt.edu/ark:/67531/metadc824861/.

Readex. 2021. *America's Historical Newspapers*. Naples, FL: Newsbank. https://www.readex.com/products/americas-historical-newspapers.

Shaver, Andrew, David B Carter, and Tsering Wangyal Shawa. 2019. Terrain ruggedness and land cover: Improved data for most research designs. *Conflict Management and Peace Science* 36 (2):191–218.

Swift, Elaine K., Robert G. Brookshire, David T. Canon, Evelyn C. Fink, John R. Hibbing, Brian D. Humes, Michael J. Malbin, and Kenneth C. Martis. 2009. *Database of [United States] Congressional Historical Statistics, 1789-1989*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/ICPSR03371.v2.

United States Congress. 2020. *Biographical directory of the United States Congress*. Washington, D.C.: United States Congress. https://bioguide.congress.gov/.

United States Post Office Department. 1842. *Table of the Post Offices in the United States*. Washington: J. & G.S. Gideon.

———. 1862. *List of Post Offices in the United States*. Washington: Government Printing Office.

———. 1881. *United States Official Postal Guide*. Boston: Houghton, Mifflin, and Company.

———. 1892. *United States Official Postal Guide*. Philadelphia: Geo F. Lasher.

———. 1900. *United States Official Postal Guide*. Philadelphia: George F. Lasher.