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

Supplemental Materials

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

S1 Letter Writing in the 19th Century **S1**

S2 Figure 1 under Alternative Bandwidths **S3**

S3 Simulation: Slope Changes at All Time Points, 1820-1880 **S5**

S4 Additional Newspaper Results **S8**

S5 Additional Congressional Corpus Results **S12**

S1 Letter Writing in the 19th Century

Our analysis draws on two types of textual sources: Congressional speeches and newspapers (specifically, editorials and letters to the editor). We argue that the contributors to newspapers were non-elites or less elite than members of Congress. Here, we consider evidence to support that claim.

Ideally, we would examine data on the identities of letter writers along with their biographic and demographic characteristics, such as level of education or occupation. Unfortunately, these data do not exist for the vast majority of letters that appear in our dataset.

Instead, we draw evidence from secondary sources, which suggest that letter writing in general (as opposed to the specific practice of writing to newspapers) was unlikely to be an elite-only phenomenon. Consider two types of potential barriers to letter-writing: literacy and economic cost. With respect to the first factor, literacy in the middle of the 19th century was "virtually universal" among the free population.[1] Data from the 1850 decennial census indicate that 90% of the white population aged 20 or older was literate.[2] Thus, the ability to read and write would not have posed a major barrier to letter writing in this period.

With respect to the second factor, the barrier to letter writing would have been the cost to send the letter in the mail. Prior to 1845, letter writing was expensive because the post office charged high rates on letters in order to subsidize the sending of newspapers.[3] After 1845, however, rate reductions "made it possible for virtually anyone, regardless of financial circumstances, to carry on a correspondence with distant family and friends."[4]

---

[1] Goldin 2006, 2-387.

[2] Ruggles, Flood, Foster et al. 2021.

[3] John 1995, 156, 318fn189.

[4] John 1995, 161.

Furthermore, although sending letters in the pre-1845 era would have been expensive, we do not think that cost would have discouraged the sending of mail to businesses such as newspaper organizations. Postage fees in the pre-1845 period were collected from the recipient unless the sender paid the fee in advance; prepayment was uncommon, however, because senders were "unwilling to pay for a service until they could be sure that it had been properly performed."[5] While individuals may have been reluctant to impose such costs on family members or friends, we think it is less likely they would have had such qualms about saddling impersonal organizations with the expense.

[5] John 1995, 160.

S2 Figure 1 under Alternative Bandwidths

Figure 1 displayed smoothed temporal trends in singular usage in from the newspaper corpus. The data were fitted using lowess and a 60% bandwidth, and suggested that the Northern trend begins to diverge from the Southern trend around the time of the Civil War. Here, we examine this impression across a range of bandwidths. Figure S1 displays the results. We see that, no matter what bandwidth used, the first-differences in the North become more positive than in the South around the time of the Civil War, indicating an acceleration of singular usage in the North around this period.

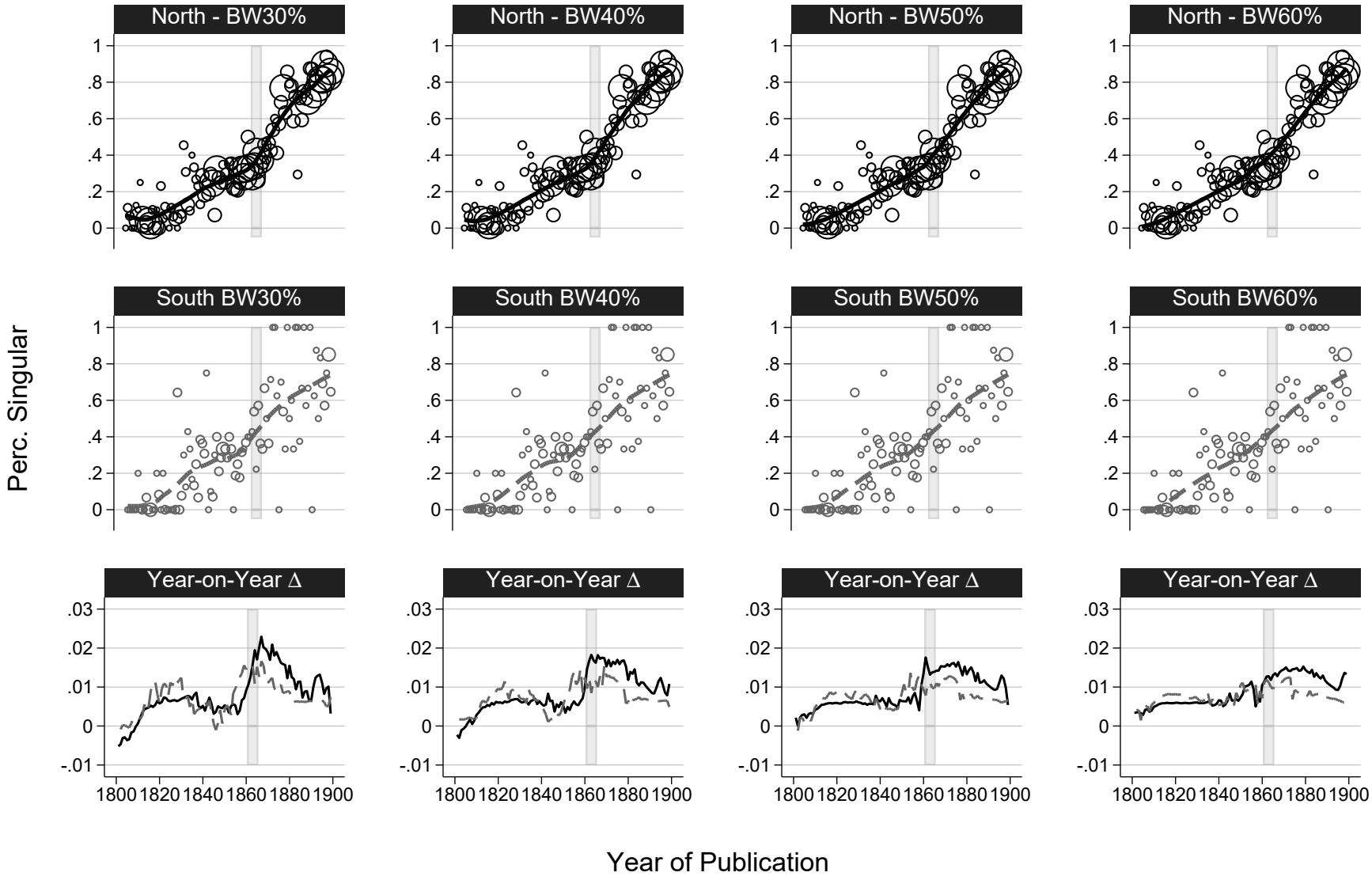


Figure S1: The top and middle panels show fitted lowess lines using different bandwidths (e.g. line fit using a moving window comprising 30% of the data). The bottom panel shows the first differences of the lowess lines (i.e. estimated singular in year t - estimated singular in year t-1).

S3 Simulation: Slope Changes at All Time Points, 1820-1880

This section reports on diagnostics we conducted to determine whether the Civil War period represents a "true" turning point in terms of an acceleration in the time trend of singular usage in the North. Importantly, we note that underlying shape of the data is convex and upward sloping (i.e. taking the form of an elongated "J"). As a result, traditional placebo tests (i.e. employing alternative cutpoints besides 1860) will "over-detect" statistically significant slope changes simply as an artifact of the data curvature.

We employ simulated data to illustrate this issue. In particular, we generate a dataset of 4500 observations, distributed even over 100 years (45 observations per year) in the period 1800-1899. We chose N = 4500 as this approximates the size of our newspaper dataset. The probability that an observation is coded as "singular" is then given by the following formulas:

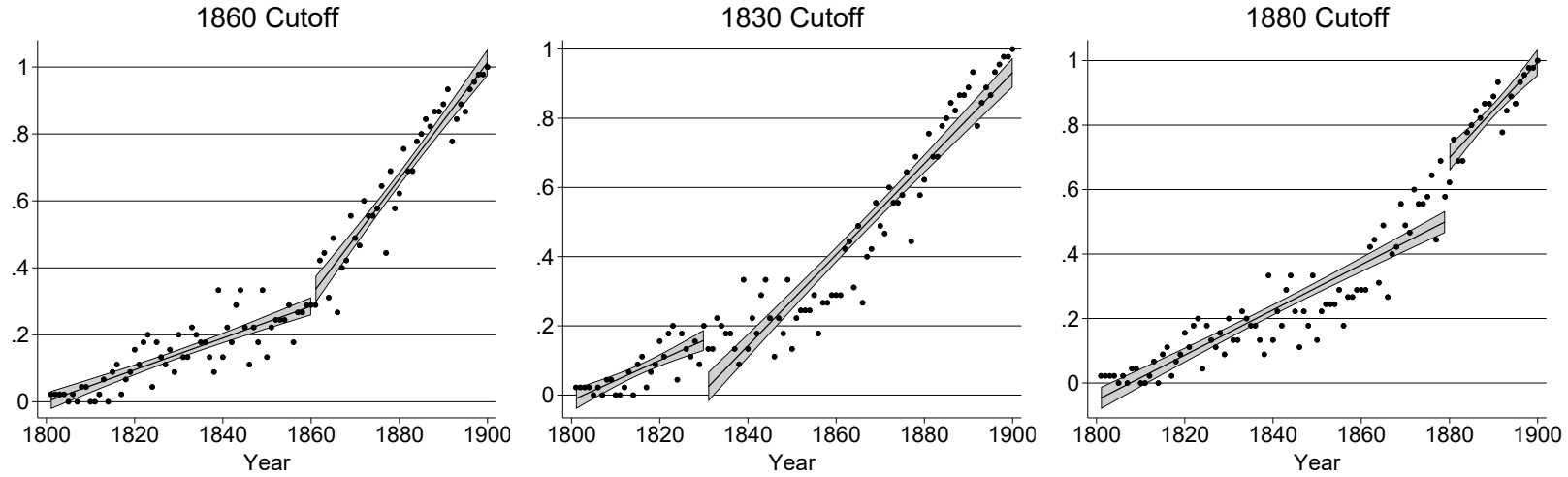
- For each year y from 1801 - 1860, an observation has a $\frac{y-1800}{200}$ chance of being coded as SINGULAR. For example, approximately 5% of all observations in year 1810 are coded SINGULAR = 1, while the same applies for approximately 30% of all observations in year 1860.

- For each year y from 1861 - 1900, an observation is coded singular with probability $0.3 + \frac{.7\times(y-1860)}{40}$.

Thus, by this data generating process, the slope in the period 1801 - 1860 is 0.5% per year, while the slope in the period 1861-1900 is 1.75% per year. These parameters roughly approximate the shape of our newspaper data.

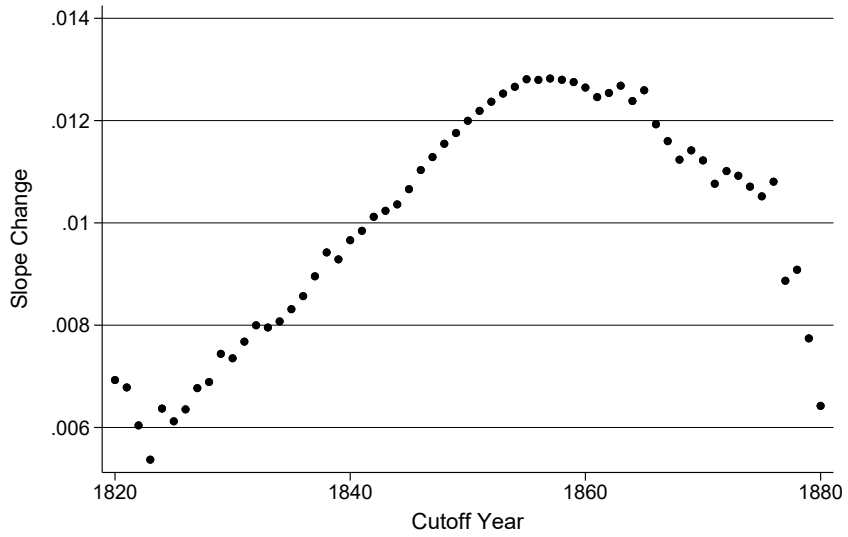
The simulated data, as well as the lines of best fit, are shown in the left panel of Figure S2. The center and right panels also show what happens when we intentionally choose the "wrong" cutpoints (1830 and 1880). We observe a positive slope change in all three cases, which reflects the underlying shape of the data.

Figure S2: Simulated Data with "True" Cutpoint at 1860



Note: The figure displays the lines of best fit, centering Y at 1860 and setting $P = 1$ if $Y > 0$.

Figure S3: Slope Change at all Cutpoints 1820-1880 in Simulated Data



Note: The figure displays the coefficients of slope change in our simulated data, centering Y iteratively at each year 1820–1880 and setting $P = 1$ if $Y > 0$.

Extending our example, we next plotted the coefficient of slope change, treating each year 1820-1880 as a cutpoint in turn. The results, shown in Figure S3, reveal an interesting pattern. In particular, while we observe statistically significant effects at all cutpoints, the substantive size of the slope change reaches a local maximum around the "true" cutpoint at 1860.

A similar pattern obtains when we repeat this exercise with our real newspaper data (see Figure 4). Overall, we take this further diagnostic evidence that the Civil War is indeed the turning point around which the trend in singular adoption accelerates in the North.

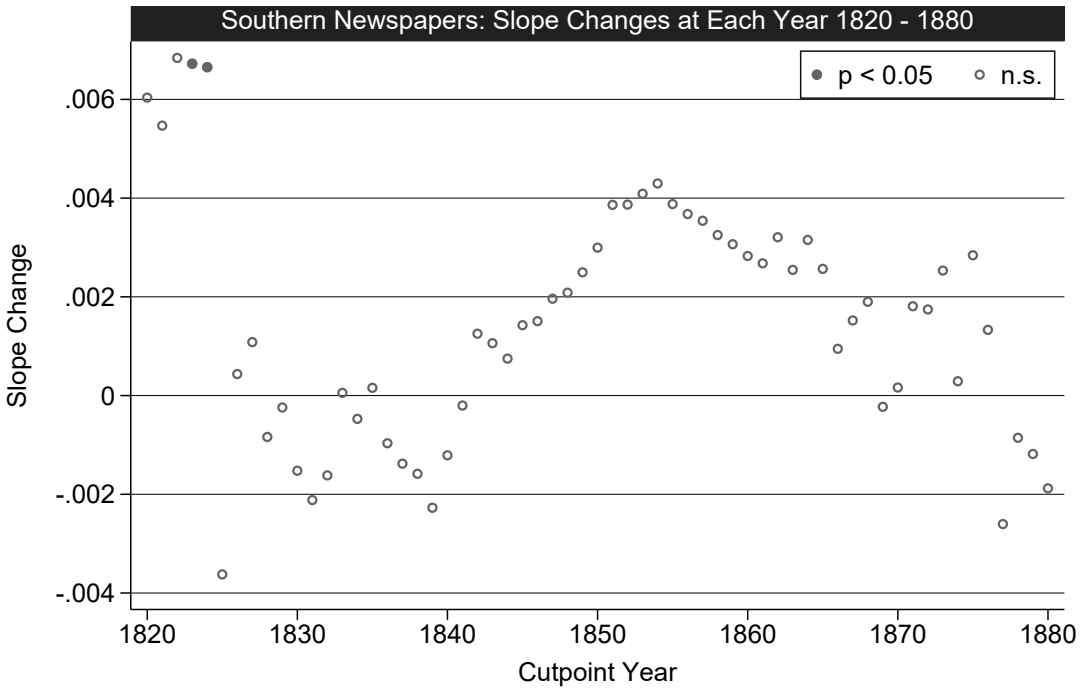
In Appendix section S4, we repeat this exercise with respect to time trends in (i) Southern newspapers, (ii) Lincoln counties, and (iii) McClellan counties. Consistent with the results reported in the main text, we find the greatest slope changes around the Civil War for Lincoln counties. In contrast, results for Southern newspapers and McClellan counties are substantively smaller and largely statistically insignificant.

S4 Additional Newspaper Results

This section reports some additional results that we removed from the main text due to space constraints.

Our main findings demonstrate little evidence that the Civil War coincided with changes in pre-war patterns of singular usage in Southern newspapers. In fact, as shown in Figure S4, the Southern time trend in singular usage appears roughly constant throughout the entire period 1820-1880.

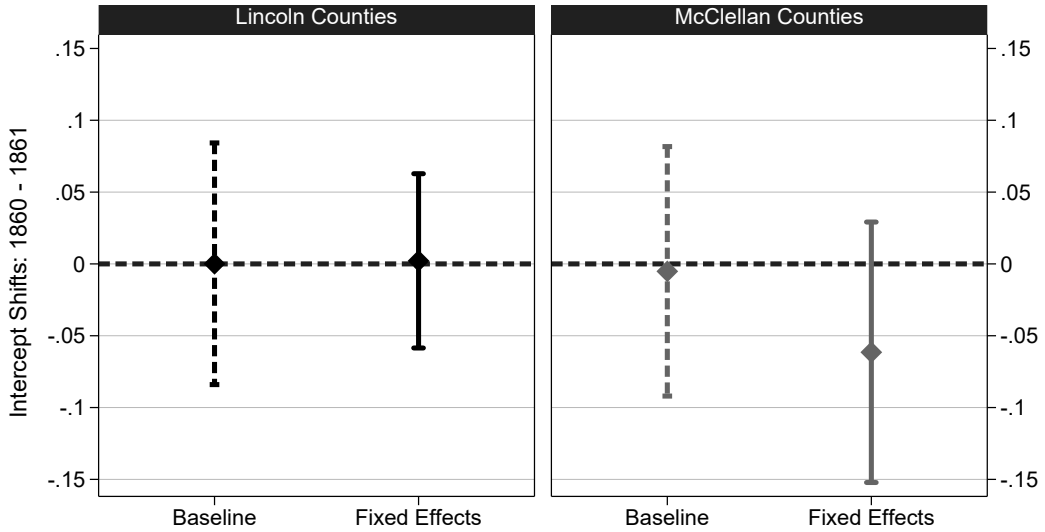
Figure S4: Slope Changes in Southern Newspapers at Each Year 1820–1880



Note: The Figure displays slope changes in the time trend of singular usage in Southern newspapers, centering Y iteratively at each year from 1820–1880 and setting $P = 1$ if $Y > 0$.

In our main sectional newspaper analysis, we found that the Civil War corresponded with a slope change in the time trend of singular usage for Northern newspapers. Given these results, our analysis of Lincoln vs. McClellan counties focused on slope changes. In Appendix Figure S5, we shown the results with respect to intercept changes. In parallel with our overall sectional results reported in Figure 3 of the main text, find no evidence of discontinuous or sudden jumps in singular usage in either Lincoln or McClellan counties.

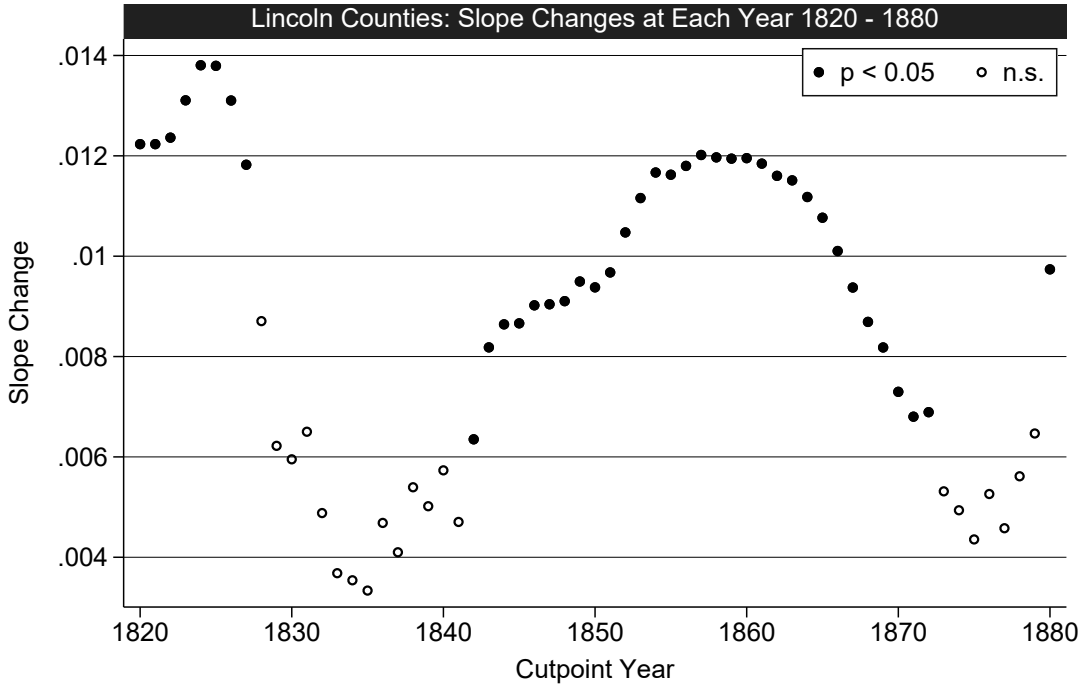
Figure S5: Singular Usage in Northern Newspapers, by 1864 Election Results



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 $Y > 0$. Full results are available in Appendix Table D6.

Finally, we consider the substantive size of the slope changes in Lincoln counties, setting each year 1820-1880 as a cutpoint. Coefficients are plotted in Appendix Figure S6. Consistent with the results reported in the main text, we find observe a local maximum around the time of the Civil War.

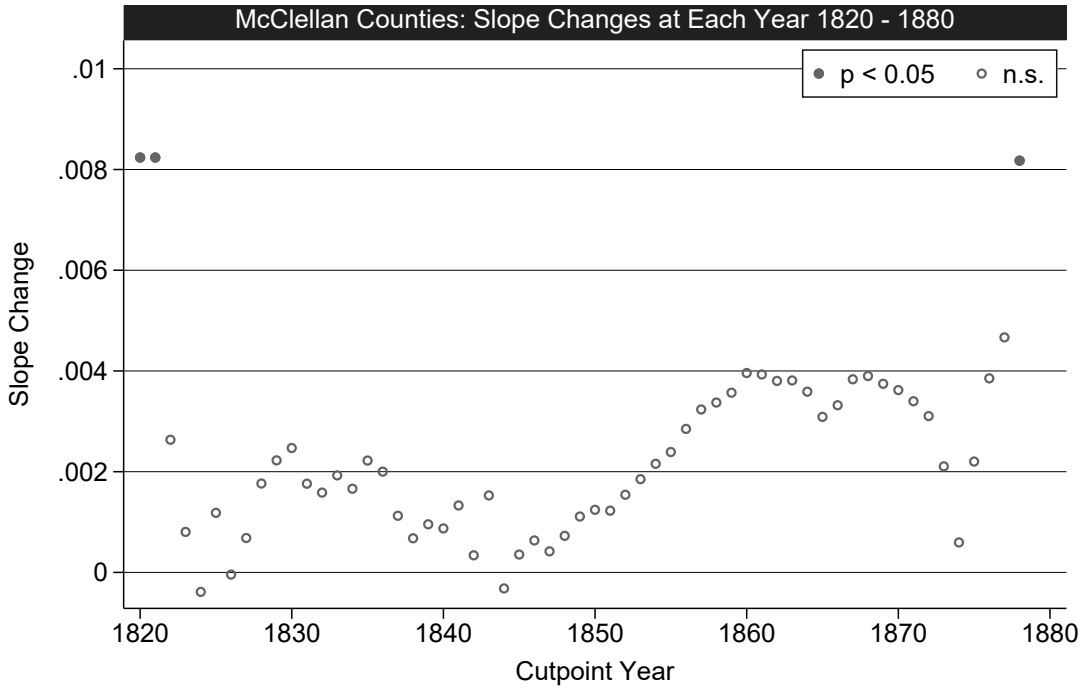
Figure S6: Slope Changes in Lincoln Counties at Each Year 1820–1880



Note: The Figure displays slope changes in the time trend of singular usage in Lincoln counties, centering Y iteratively at each year from 1820–1880 and setting $P = 1$ if $Y > 0$.

For completeness, we also plot analogous coefficients for McClellan counties in Appendix Figure S7. As expected, the estimates substantively smaller and largely statistically insignificant.

Figure S7: Slope Changes in McClellan Counties at Each Year 1820–1880

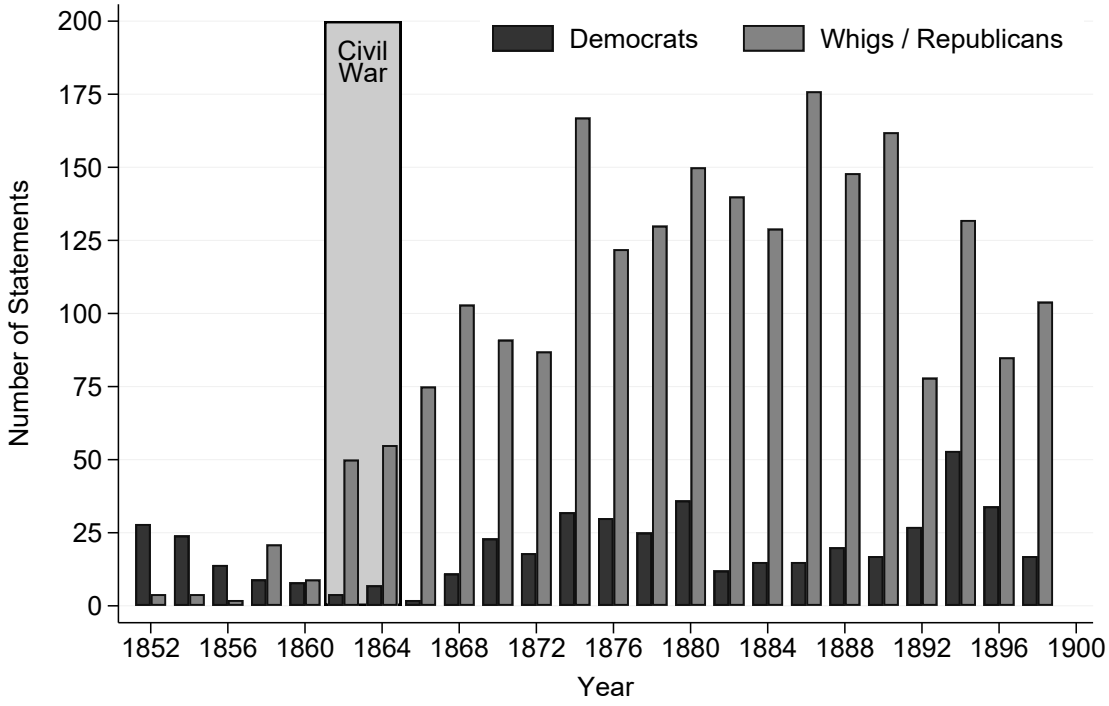


Note: The Figure displays slope changes in the time trend of singular usage in McClellan counties, centering Y iteratively at each year from 1820–1880 and setting $P = 1$ if $Y > 0$.

S5 Additional Congressional Corpus Results

This section reports additional analyses of our Congressional Speech corpus. We begin by examining Northern members of Congress. Appendix Figure S8 displays the number of statements made per Congress, by Northern Whigs/Republicans and Northern Democrats. We observe that, as reported in the main text, there is relatively little data available for the pre-war period.

Figure S8: Number of Statements by Party



Note: The figure displays the number of statements made by Northern Democrats and Northern Whigs / Republicans in the Congressional Corpus, divided between statements made before versus after 1860.

To more robustly examine partisan differences among Northern Congressmen, we focus on the period after 1860 only and estimate the following model:

$$S_{isc} = \alpha + \beta REP_{isc} + \delta W_{isc} + \gamma_c + \zeta_s + \epsilon_{isc} \tag{3}$$

where γ represents Congressional session (e.g. the 35th Congress) fixed effects and ζ denotes Congressmen random effects. This model allows us to look for partisan differences in the level of singular usage, without relying on the (noisy) pre-war data. Appendix Table S1 presents the results. Even restricting our attention to the post-1860 period, we observe that Northern Republicans use the grammatical singular at higher rates than Northern Democrats.

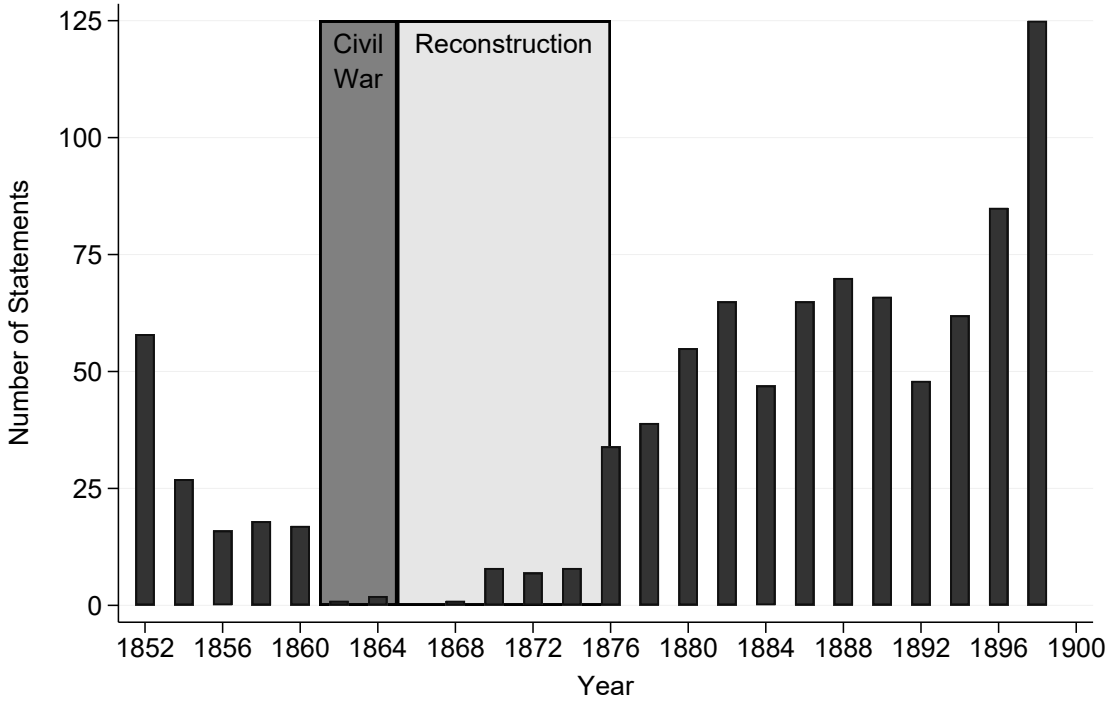
Table S1: Singular Usage among Northern Congressmen after 1860

	(1)	(2)
Republican	0.055+	0.061+
	(0.031)	(0.031)
County of Birth: % Urban (std)		-0.017
		(0.015)
County of Birth: Post Office Density (std)		-0.011
		(0.016)
County of Birth: Terrain Ruggedness (std)		-0.024+
		(0.013)
Born on River		0.033
		(0.030)
Born on Canal		0.029
		(0.029)
Year of Birth (std)		-0.004
		(0.015)
Attended College		-0.035
		(0.026)
Served in US Military		0.004
		(0.027)
Constant	0.392***	0.389***
	(0.066)	(0.070)
N	2582	2582
Congress Fixed Effects	Yes	Yes

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

Southern Congressmen. Next we turn to additional analysis examining Southern members of Congress. Here we provide some suggestive evidence that the partisan patterns we uncover among Northern Congressmen are also present among Southern Congressmen. Our analysis here is necessarily suggestive because of two features of the data. First, as Figure S9 shows, we have very few statements from Southern Congressmen during Civil War and Reconstruction. Second, as Figure S10 shows, the vast majority of our Southern sample comprises statements made by Southern Democrats. Further, in the post-war period, Southern Democrats also tended be former Confederate soldiers/officials.

Figure S9: Number of Statements by Southerners

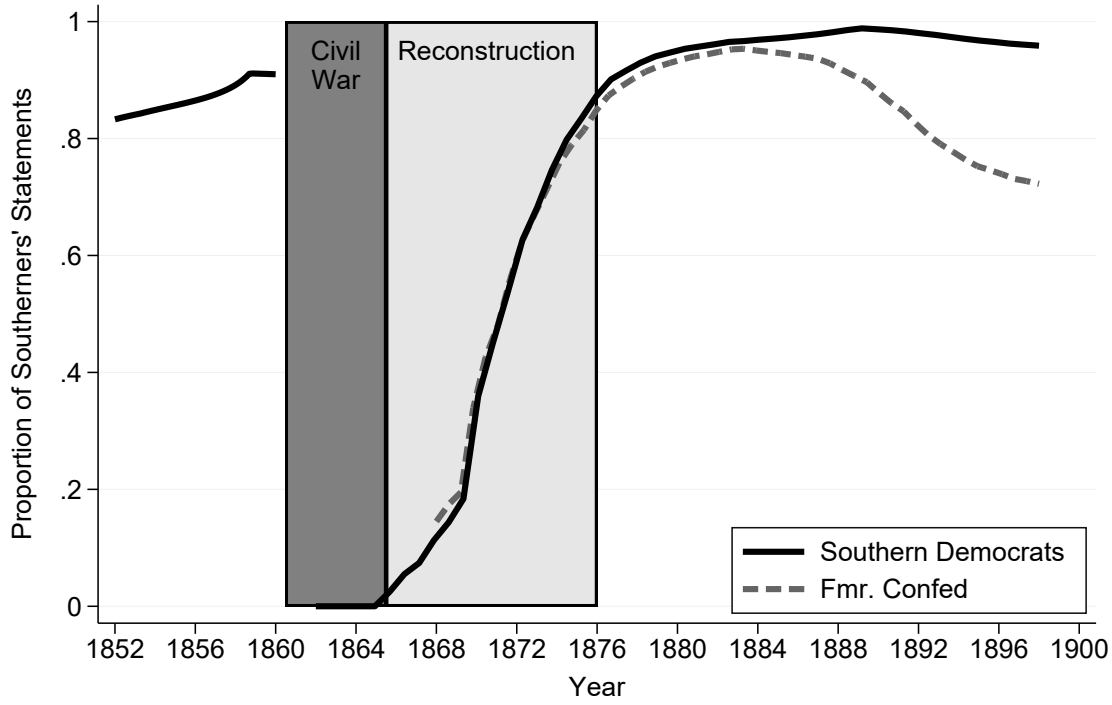


Note: The figure displays the number of statements made by Southern members of Congress during the Pre-War, Civil-War, Reconstruction and Post-Reconstruction Periods.

With these two caveats in mind, we examine singular usage among Southern Republicans and Southern Democrats. Briefly, our results below suggest that (i) Southern Republicans, particularly

those appearing during Reconstruction, were even prone to singular usage than other Congressmen, and (ii) the Civil War had little effect on the speech patterns of Southern Democrats.

Figure S10: Proportion of Southern Statements by Southern Democrats and Former Confederates



Note: The figure displays the number of statements made by Southern Democrats and former Confederate soldiers/officials as a proportion of the total number of statements made by Southern Congressmen in a particular session of Congress. Statements made by Southern Democrats predominate through our study period, except during the Civil War and early Reconstruction period. However, from the end of Reconstruction onward, Southern statements were made almost exclusively by individuals who were *both* former Confederates and members of the Democratic party, although the proportion of statements made by former Confederates begins to decline with the passing of the Civil War generation.

Southern Whigs and Reconstruction Republicans. Our entire sample contains only 49 statements made by Southern Whigs/Republicans. Of these, 13 statements were made before the War, 22 were made during Reconstruction, and 14 statements were made in the post-Reconstruction period. Not a single statement in the pre-War period used the grammatical singular. In contrast, 15 out of the 22 statements (68%) made during Reconstruction used the grammatical singular. This is almost the same rate of singular usage as for the post-Reconstruction statements (10 out of 14, of 71%), although the mean year of Reconstruction statements is 1872, compared to 1891 for post-Reconstruction statements. For comparison, the rate of singular usage among Northern Congressmen during the years 1866–1876 was only around 60%. In other words, it appears that Reconstruction Republicans were outliers in their tendency to prefer singular over plural.

Southern Democrats. Appendix Figure S11 shows patterns of singular usage among Southern Democrats. Visually, despite the gap in coverage, it appears that a single time trend explains the data, and that the Civil War had no effect on the speech patterns of Southern Democrats.

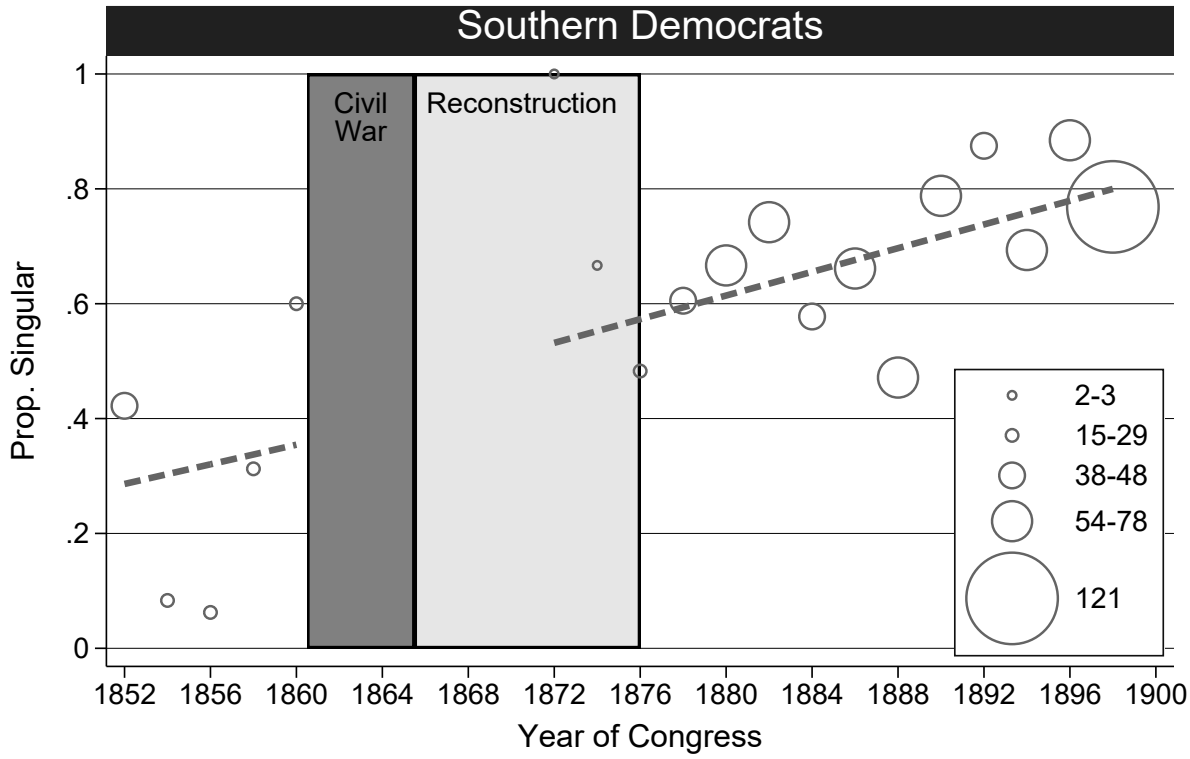
To assess this impression, we begin by estimating the following linear probability model, dropping statements made between the years 1860 – 1876:

$$\begin{aligned} S_{isc} = \alpha + \beta_1 Y_{isc} + \beta_2 P_{isc} + \beta_3 Y \cdot P_{isc} \\ + \delta W_{isc} + \zeta_{1s} + \zeta_{2c} + \epsilon_{isc} \end{aligned} \tag{4}$$

where Y indicates years, P indicates a *Post-Reconstruction* period dummy, W denotes a vector of covariates (described in the main text), and the ζs represent cross-nessted speaker and Congress random effects. Column 1 of Appendix Table S2 shows the results.

Notice that neither the coefficient on $Post - Reconstruction$ nor the interaction term $Year \times Post - Reconstruction$ are statistically significant in Column 1. We also re-estimate the model after dropping the interaction term. We thus ask whether, assuming a linear time trend for the entire

Figure S11: Singular Usage among Southern Democrats



Note: The trend lines show the rate of singular usage among Southern Democrats (n=863). Size of the bubbles indicates the number of statements in each session of Congress. The gray bars show the years of the Civil War and Reconstruction.

period, there is any evidence of an intercept shift? Results presented in Column 2 of Appendix Table S2 show that, again, the estimate for $Post - Reconstruction$ is not significant. Finally, we also drop the $Post-Reconstruction$ dummy and estimate a simple model with only a single $Year$ trend and W. Results in Column 3 show that the single time trend does indeed capture the pattern we observe in Appendix Figure S11: the rate of singular usage among Southern Congressmen appears to increase by about 1% per year.

Overall, these results suggest that, consistent with our visual impression from Appendix Figure S11, the Civil War had little effect on the speech patterns of Southern Republicans.

Table S2: Singular Usage among Southern Democrats

	(1)	(2)	(3)
Year of Speech	0.015	0.006	0.009**
	(0.023)	(0.005)	(0.003)
Pre-War	0.000	0.000	
	(.)	(.)	
Post-Reconstruction	0.077	0.104	
	(0.186)	(0.173)	
Year × Post-Reconstruction	-0.009		
	(0.023)		
County of Birth: % Urban (std)	0.001	0.001	0.001
	(0.018)	(0.018)	(0.018)
County of Birth: Post Office Density (std)	0.042*	0.042*	0.042*
	(0.021)	(0.021)	(0.021)
County of Birth: Terrain Ruggedness (std)	0.028	0.029	0.031
	(0.022)	(0.022)	(0.022)
Born on River	0.114*	0.113*	0.113*
	(0.053)	(0.053)	(0.053)
Born on Canal	-0.049	-0.050	-0.052
	(0.121)	(0.121)	(0.120)
Year of Birth (std)	0.053+	0.054+	0.056+
	(0.031)	(0.031)	(0.031)
Attended College	0.031	0.032	0.034
	(0.047)	(0.046)	(0.046)
Served in US Military	0.161+	0.162+	0.159+
	(0.096)	(0.097)	(0.096)
Constant	0.402**	0.362***	0.390***
	(0.134)	(0.089)	(0.075)
N	856	856	856

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

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.

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

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/.

Bensel, Richard Franklin. 1990. *Yankee Leviathan: The origins of central state authority in America, 1859-1877*. New York: Cambridge University Press.

Blevins, Cameron. 2021. *Paper Trails: The US Post and the Making of the American West*. New York: Oxford University Press.

Bockstette, Valerie, Areendam Chanda, and Louis Putterman. 2002. States and markets: The advantage of an early start. *Journal of Economic growth* 7 (4):347–369.

Cermeño, Alexandra L, Kerstin Enflo, and Johannes Lindvall. 2021. Railroads and Reform: How Trains Strengthened the Nation State. *British Journal of Political Science* 1–21.

Englebert, Pierre. 2009. *Africa: unity, sovereignty, and sorrow*. Boulder, CO: Lynne Rienner Publishers.

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

Gennaioli, Nicola, and Ilia Rainer. 2007. The modern impact of precolonial centralization in Africa. *Journal of Economic Growth* 12 (3):185–234.

Goldin, Claudia. 2006. Education. In *Historical Statistics of the United States, Earliest Times to the Present: Millennial Edition*, edited by Susan B. Carter, Scott Sigmund Gartner, Michael R. Haines, Alan L. Olmstead, Richard Sutch, and Gavin Wright, 2–387–2–397. Cambridge University Press.

Hechter, Michael. 2000. *Containing nationalism*. New York: Oxford University Press.

Herbst, Jeffrey. 2000. *States and power in Africa*. Princeton, NJ: Princeton University Press.

Hill, Mark J, and Simon Hengchen. 2019. Quantifying the impact of dirty OCR on historical text analysis: Eighteenth Century Collections Online as a case study. *Digital Scholarship in the Humanities* 34 (4):825–843.

Holt, Keri. 2019. *Reading These United States: Federal Literacy in the Early Republic, 1776-1830*. Athens: University of Georgia Press.

Jensen, Jeffrey L., and Adam J. Ramey. 2020. Going postal: State capacity and violent dispute resolution. *Journal of Comparative Economics* 48 (4):779–796.

John, Richard C. 1995. *Spreading the News: The American Postal System from Franklin to Morse*. Cambridge, MA: Harvard University Press.

Kernell, Samuel. 1977. Toward understanding 19th century congressional careers: Ambition, competition, and rotation. *American Journal of Political Science* 669–693.

Lee, Dong Wook, Melissa Z. Rogers, and Hillel David Soifer. 2022. The Modifiable Areal Unit Problem in Political Science. Working paper.

Loughran, Trish. 2007. *The Republic in Print: Print Culture in the Age of US Nation Building, 1770-1870*. New York: Columbia University Press.

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.

Myers, Minor. 2008. Supreme Court Usage and the Making of an 'Is'. *Green Bag* 11 (4):457–465.

Rogowski, Jon C., John Gerring, Matthew Maguire, and Lee Cojocaru. 2021. Public Infrastructure and Economic Development: Evidence from Postal Systems. *American Journal of Political Science* Early View.

Ruggles, Steven, Sarah Flood, Sophia Foster, Ronald Goeken, Jose Pacas, Megan Schouweiler, and Matthew Sobek. 2021. *IPUMS USA: Version 11.0 [dataset]*. Minneapolis, MN: IPUMS. https://doi.org/10.18128/D010.V11.0.

Santin, Bryan, Daniel Murphy, and Matthew Wilkens. 2016. Is or Are: The" United States" in Nineteenth-Century Print Culture. *American Quarterly* 68 (1):101–124.

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.

Swain, John W, Stephen A Borrelli, Brian C Reed, and Sean F Evans. 2000. A new look at turnover in the US House of Representatives, 1789-1998. *American Politics Quarterly* 28 (4):435–457.

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/.

van Strien, Daniel, Kaspar Beelen, Mariona Coll Ardanuy, Kasra Hosseini, Barbara McGillivray, and Giovanni Colavizza. 2020. Assessing the impact of OCR quality on downstream NLP tasks. In *Proceedings of the 12th International Conference on Agents and Artificial Intelligence*. https://doi.org/10.17863/CAM.52068.

Zhang, Nan, and Melissa M Lee. 2020. Literacy and State–Society Interactions in Nineteenth-Century France. *American Journal of Political Science* 64 (4):1001–1016.