Measuring the Content of Presidential Policy Making: Applying Text Analysis to Executive Branch Directives

AARON R. KAUFMAN

*The executive branch produces huge quantities of text data: policy-driving documents such as executive orders, national security directives, and agency regulations; procedural documents such as notices of proposed rule making, meeting transcripts, and presidential daily schedules; and public relations documents including press releases, veto statements, and social media posts. Political science is increasingly turning to methods of automated text analysis to rigorously interrogate such corpora. These tools have proved invaluable to scholars of the judiciary, of the legislature, and of political behavior; they hold great promise for the study of the presidency and bureaucracy. These methods' value lies in measuring complex, nuanced, yet theoretically critical quantities of interest like power, agenda setting, policy significance, ideology, and diplomatic resolve: by categorizing individual documents into a researcher-defined schema, automated text methods produce large-*N *data sets with fine temporal granularity. Recent work using text analysis to study unilateral actions illustrates these methods' promise for upturning conventional wisdom, settling long-standing debates, and illuminating new puzzles in executive politics in the United States.*

Keywords: text analysis, unilateral actions

Introduction: Text Analysis and U.S. Political Institutions

Among Congress, the Supreme Court, and the presidency, the latter is the most difficult to quantitatively study. Unlike the judicial and legislative branches, both of which produce a wide array of votes, the presidency produces very few instances of observable and quantifiable policy-relevant behavior, chief among them vetoes and unilateral actions. But many vetoes are foregone conclusions, especially during eras of divided government,

---

Aaron R. Kaufman is assistant professor of political science, Division of Social Sciences, New York University Abu Dhabi. **AUTHOR'S NOTE:** I thank Matt Blackwell, Jon Rogowski, Robert Ward, and Madeleine Wolf for helpful comments and feedback during preparation of this article. All remaining errors are my own.

DOI: 10.1111/psq.12629

and studying unilateral actions is fraught: no definitive counts of them exist, and while some are of great policy importance, most are ceremonial or inconsequential. This data limitation constrains rigorous quantitative study of the executive branch.

Scholars of the legislature and the judiciary, however, have turned in recent years to examine the observable outputs of their branches other than votes. Chief among these outputs are raw text: floor speeches, committee transcripts, bill drafts and amendments, press releases, oral argument hearings, amicus curiae briefs, and court decisions collectively compose the bulk of information produced by Congress and the courts. The Congressional Record alone accounts for more than 50 gigabytes of text data.

The executive branch, too, produces an enormous quantity of text. The presidency itself produces many types of documents driving policy: executive orders, public lands orders, policy directives, national security decision memoranda, presidential security directives, presidential proclamations, treaty proclamations, and statements of administration policy, among many others. Other types of documents from the White House may not set policy directly but still encode important information: pardons; press releases; signing or veto statements; agency or department circulars; and, since the presidency of Barack Obama onward, social media posts. Executive agencies produce even more data still, including regulations, notices of proposed rulemaking, State Department cables, military records, and Federal Reserve memos and actions.

While such archives represent an opportunity for evaluating long-standing theories and settling timely debates in political science, the prospect of carefully reading millions of pages of administrative text and recording relevant information about them is daunting. Only recently have the computational tools become available for quantifying and automatically analyzing these documents, and many political scientists, including political theorists, have taken note (Blaydes, Grimmer, and McQueen 2018; Grimmer and Stewart 2013; Hollibaugh 2018; Wilkerson and Casas 2017) of how these tools can enable scholars to identify new descriptive patterns on a scale not found in existing work, systematize measurement strategies for theoretical quantities of interest, and facilitate stronger and better tests of classic theories and reveal patterns that facilitate new theoretical advances.

The aims of this article are to illuminate the basic mechanics of text-as-data and promote its applications in the study of the executive branch. The next section defines the two basic paradigms of text-as-data in the social sciences, supervised and unsupervised learning, with citations to important papers in the study of U.S. political institutions as exemplars. The third section discusses how these tools have been applied primarily to Congress and the Supreme Court, followed by a section showing how those same designs might be applied to key debates in the study of the bureaucracy and the presidency, with an extended example of measuring the significance of presidential unilateral actions. The final section concludes with a discussion of the methodological future of the study of executive politics.

The Fundamentals of Text Analysis

The two components of text analysis are the *representation* and the *learner*. The representation is the transformation from raw text such as sentences, paragraphs, or documents into a familiar rectangular format; the learner is the model used to characterize or summarize those sentences, paragraphs, or documents. Representations and learners both range in complexity and in the assumptions they entail. The goal of text analysis is to summarize or characterize a document, placing it either in a category or at a point in a (usually unidimensional) space. We might, for example, want to identify the policy (a categorical label) or the partisan content (a continuous variable) that an item of legislation, a court decision, a federal regulation, or a presidential press release relates to or contains. We can choose one of two general approaches, each with its pros and cons (Table 1). Unsupervised learning looks for patterns in data, identifying clusters of observations or variables; supervised learning focuses on predicting an outcome or a measure using examples provided by the researcher.

Representations

The simplest representation of raw text as rectangular data set is a matrix in which each column is a word feature and each row is a document. This format is called a term-document matrix or word frequency matrix[1] (Brown et al. 1992). Sometimes called the "bag of words" model, this representation implicitly assumes that the ordering of words does not matter.[2] Representations may include a number of possible preprocessing steps with important substantive implications (Denny and Spirling 2018). Such steps include removing exceedingly rare or common words, combining related words together (e.g., write, writing, and written into "writ," called stemming), and removing punctuation, numerals, and capitalization. An important preprocessing tool is *n*-gram selection: rather than each column in the term-document matrix indicating a single word, each column is a set of sequential words. Including 2- and 3-grams reduces the amount of information lost in representing raw text as a matrix, but it comes at the cost of greatly increasing the number of variables.

Unsupervised Learning

Unsupervised learning has very low start-up costs and entails minimal assumptions in the modeling step, and for this reason it is more popular in political science than supervised learning. As a downside, the clusters or topics it produces as outputs are often difficult to interpret, requiring substantial researcher expertise (Chang et al. 2009). The most familiar unsupervised learning model is the topic model and its variants (Blei and Lafferty 2006; Blei, Ng, and Jordan 2003; Roberts et al. 2014), which view documents as a mixture of topics characterized by key words. A researcher provides a rectangular data set derived from a corpus of documents and indicates the number of topics to identify (k);

1. Documents are defined arbitrarily: they can be single sentences up to collections of books. Words are defined arbitrarily as well, and can be much more complex attributes of the texts like sequential sets of words, parts of speech, or functions of sets of words.

2. While there are more information-preserving ways to represent text (Devlin et al. 2018; Le and Mikolov 2014), they are rarely used in political science.

TABLE 1
A Comparison of Unsupervised and Supervised Learning Paradigms

	Unsupervised	*Supervised*
Term-document matrix	Yes	Sometimes
Requires input data	Yes	Yes
Requires input labels	No	Yes
Requires interpretation	Yes	No
Usage	Describing the data	Predicting an outcome or a measure
Examples	Topic modeling, k-means clustering, factor analysis, principle components analysis	Linear regression, neural networks, random forests, LASSO, support vector machines

the topic model produces a list of k topics with the words that define them and each document's unique topic mixture. A common use of topic models is as an exploratory analysis of a new corpus of documents, though document-level topic labels have substantive uses as well (Hollibaugh 2018; Mozer et al., forthcoming; Quinn et al. 2010; Rice 2019; Roberts, Stewart, and Nielsen, forthcoming). A topic model with $k = 5$ trained on presidential unilateral actions might identify topics related to social policy, foreign policy, economic policy, ceremonial actions, and pardons; a document about the Mexico City Rule might be 80% foreign policy, 15% social policy, and 5% ceremonial, while a resolution supporting International Human Rights Week might be 90% ceremonial and 10% foreign policy. However, this represents a best-case scenario, and more often than not, the topics identified by an unsupervised learner are ambiguous at best. Panel A of Table 2 presents a portion of the results from a topic model with $k = 10$ of presidential unilateral actions from 1953 to 2017 (Kaufman and Rogowski 2019); while four of the topics are substantive and coherent, one is administrative rather than substantial. Panel B shows topic memberships for the first six unilateral actions in the corpus.

Supervised Learning

Supervised learning requires more input from the researcher but produces less ambiguous results. Alongside the rectangular data set required by unsupervised learning (in this context, usually called *features*), supervised learning also requires a vector containing the dependent variable (usually called *labels*). If the labels are continuous, the supervised learner is akin to regression. If the labels are categorical, the supervised learner is called a *classifier*. Considering the presidential unilateral actions above, we might know ex ante that some of our documents relate to foreign policy, some to social policy, and others to economic policy; perhaps research assistants hand-coded some of the documents. If we provide a supervised learner with those ex ante labels alongside our (representation of the) text, our supervised learner discovers the relationship between word use and broad policy area, akin to estimating regression coefficients. Then, if we apply the trained learner to the remainder of our documents not yet hand-coded by a research assistant, the model will produce its best guess as to which policy area each document belongs. A diagram of this process is in Figure 1.

TABLE 2
Topic Model Output from 33,000 Presidential Unilateral Actions with 10 Topics

	Topic 1	*Topic 2*	*Topic 3*	*Topic 4*	*Topic 5*
Panel A					
Top Word 1	Land management	article	functions	National emergency	amplt
Top Word 2	bureau of	treaties	The director	tariff	ampgt
Top Word 3	Mining laws	contracting	As president	actu.s.c	pardon
Top Word 4	Information contact	convention	Hereby ordered	The trade	district court
Top Word 5	Land policy	dela	The functions	this proclamation	probation
Top Word 6	Public land	The convention	employees	emergency	court for
Top Word 7	Federal land	ratification	The administrator	determination	and unconditional

	Topic 1	*Topic 2*	*Topic 3*	*Topic 4*	*Topic 5*	*Topic 6*	*Topic 7*	*Topic 8*	*Topic 9*	*Topic 10*
Panel B										
Document 1	0.137	0.042	0.002	0.001	0.002	0.000	0.004	0.001	0.810	0.002
Document 2	0.173	0.010	0.005	0.000	0.001	0.001	0.001	0.010	0.797	0.001
Document 3	0.550	0.205	0.084	0.002	0.004	0.000	0.000	0.002	0.151	0.001
Document 4	0.319	0.073	0.059	0.009	0.002	0.001	0.001	0.004	0.531	0.002
Document 5	0.479	0.161	0.046	0.001	0.001	0.000	0.000	0.001	0.309	0.001
Document 6	0.118	0.005	0.003	0.000	0.001	0.001	0.001	0.006	0.865	0.001

Note: For Panel A: The seven most characteristic words for the first five of ten topics are shown. Topic 1 might relate to public lands, Topic 2 to international agreements, Topic 4 to national emergencies, and Topic 5 to the judiciary. Topic 3 is mostly administrative language; topics are not necessarily substantive.
For Panel B: The proportion of the first six documents belonging to each of the 10 topics. Topics 1 and 9 are the most prevalent.

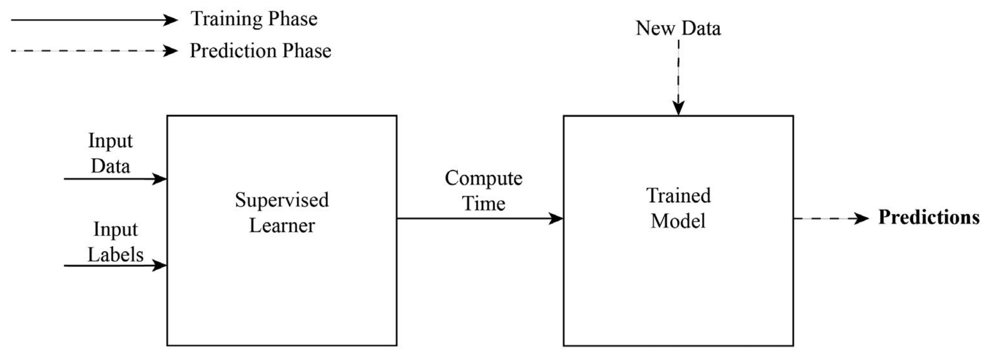

FIGURE 1. An Illustration of the Supervised Learning Process.
Adapted from Khodabandelou et al. (2013).

The final step of the supervised learning process is *validation*: verifying that the predictions or measures produced by the model are reasonable and appropriate. The most common form of validation is the *train-test split*. Before training the model, the researcher will reserve a portion (typically less than one-half) of the labeled observations and train the model using only the remaining data. The researcher then uses the model trained on the incomplete data set to produce predictions for the withheld set of data, which, importantly, already have labels. If the predicted labels generally conform to the reserved ex ante labels, that model is considered validated. There are many ways to measure such conformity, including correlation (for continuous outcomes), percent agreement (for categorical outcomes), and the area under the receiver-operator characteristic (for binary outcomes).

The key advantage of supervised over unsupervised learning is that because the researcher provides the labels ex ante, the learner's generated predictions are guaranteed to relate to the researcher's substantive interests. Researchers studying diplomatic resolve from White House statements as in Katagiri and Min (2019) require that their model characterize documents as showing high or low resolve; an unsupervised learning approach to State Department cables might instead identify topics related to geography or to ceremonial versus administrative purposes. For this reason, supervised learning is preferable in studying more complex phenomena that may be hard to explain but for which "you know it when you see it" (Kaufman, King, and Komisarchik, forthcoming).

Judicial and Congressional Text

Both supervised learning and unsupervised learning enjoy growing popularity in political science, but not all literatures have been equally quick to adopt these tools. In the study of political institutions, research on legislatures has enjoyed the greatest gains from text-as-data, though judiciary scholars have benefited as well. Legislative branch scholars' use of text analysis has focused on legislation itself and, recently, on news references to legislators or public statements such as press releases and tweets. Judicial scholars focus more on forecasting court outcomes and rely on either court decisions or transcripts.

Study of the executive branch, despite its great wealth of text data, has benefited the least thus far.

Purpura and Hillard (2006) use topic modeling, an unsupervised learning approach (see the Fundamentals of Text Analysis section), to map congressional legislation onto the Policy Agendas Project (Dowding, Hindmoor, and Martin 2016), while Quinn et al. (2010) use similar methods to measure political attention and agenda setting from congressional speeches. Many studies identify the words and phrases that distinguish Republican from Democratic speech, either in speeches (Diermeier et al. 2012; Laver and Benoit 2002; Monroe, Colaresi, and Quinn 2008) or in bills (Denny, O'Connor, and Wallach 2015), and apply these sets of words to examine ideological multidimensionality and polarization.[3] Grimmer (2013a; 2013b) and Radford and Sinclair (2016) address polarization by measuring "homestyle" (Fenno 1978) from press releases and Twitter, respectively, to study how district composition relates to position taking and ideology, respectively, while Hemphill, Culotta, and Heston (2013) use similar data to measure partisan differences in issue framing. Ban, Moskowitz, and Snyder (2016) and Ban et al. (2016) use newspaper references to legislators as a proxy for political power, among the most important concepts in political science (Dahl 2005; Riker 1964); Gentzkow et al. (2016) use the similarity between congressional text and newspaper text to study newspaper ideology and media bias.

A common use of text data in the study of the Supreme Court is decision forecasting with supervised learners, using either raw oral argument transcripts (Katz, Bommarito, and Blackman 2017; Kaufman, Kraft, and Sen 2019) or transcripts plus recorded audio (Dietrich, Enos, and Sen 2019).[4] This literature, though valuable, is less rooted in theoretical debates. Unsupervised learning approaches using topic models examine court decisions (Rice 2019)[5] showing the value in viewing Supreme Court opinions as a mix of topics. A related approach models amicus curiae briefs using random utility models (McFadden 1980), arguing these "friend of the court" documents are useful in predicting individual judges' vote choices (Sim , Routledge, and Smith 2015).

These papers address many of the most important debates in the studies of U.S. political institutions. Yet extant research on the executive branch using text analysis is rather more limited. Much of the work has studied executive agency rather than presidential text or examined debates in other subfields entirely (Katagiri and Min 2019); some work has examined presidencies outside of the United States or relayed methodological rather than theoretical findings. First, Workman (2015) uses supervised learning to identify the policy area of more than 226,710 executive agency regulations beginning in the 1980s and uses that data set to study the agenda-setting power of executive agencies themselves. Second, Anastasopoulos and Whitford (2018) study bureaucratic reputation using supervised learning to examine Twitter data from 13 executive agencies. Third, Katagiri

3. Peterson and Spirling (2018) similarly examine polarization from legislative speech taken from the UK House of Commons.

4. Aletras et al. (2016) and Sulea et al. (2017) also predict court decisions, focused on the European Court of Human Rights and the French Supreme Court, respectively.

5. Liebman et al. (2017) use topic models to study court decisions in China.

and Min (2019) address a key debate in the crisis-bargaining literature by studying both public and private U.S. State Department documents and training a supervised learner to measure their expression of Soviet resolve toward taking material action in Cold War–era Berlin. In a study of 73 Latin American countries, Arnold, Doyle, and Wiesehomeier (2017) find that presidents exhibit greater compromise in their State of the Union addresses as the legislature's composition changes. Finally, Benoit, Munger, and Spirling (2019) produce a new, generalizable measure of political sophistication derived from text and apply it to U.S. State of the Union addresses, finding that State of the Union addresses have become increasingly similar to the language complexity of a fifth grader.

The research methodologies and measurement strategies underlying recent text-as-data research on Congress and the courts can translate directly to research methodologies and measurement strategies in studying the most pertinent and timely debates in the presidency and bureaucracy literatures.

Parallel Studies in the Executive Branch

The executive branch has produced on the order of 10 terabytes of text data spread across published policy documents, public and private internal memoranda, agency rules and regulations, public relations documents, and meeting and hearing transcripts.[6] If scholars of the U.S. presidency were to leverage this executive branch text to conduct analyses parallel to those studies conducted by legislative and judicial scholars, what kinds of questions might we ask and what kinds of results might we find? Below, I consider three prominent debates in the study of the executive branch and propose automated text analysis methodologies to address them, noting which studies from the legislative or judicial literatures best parallel those research designs, followed by a detailed discussion of a completed application of supervised learning to the study of presidential unilateralism.

How Partisan Are Executive Agencies?

Much research has shown that bureaucrats are not purely procedural actors entirely reliant on protocol and hierarchy, and that they enjoy significant discretion to implement policy in accordance with their political preferences and partisan ideologies (Keiser 2001). Surveys of bureaucrats (Clinton and Lewis 2008) and other federal executives (Richardson, Clinton, and Lewis 2018) show convincingly that executive agencies are *political* institutions, and that the bureaucrats who comprise them are ideologically partisan through both their campaign donations (Chen and Johnson 2015) and their issue preferences (Bertelli and Grose 2011; Clinton et al. 2012). But none of these results address the *outputs* of bureaucracy, and it is conceivable that while bureaucrats themselves are partisan

6. This number excludes an even vaster quantity of internal data like Internal Revenue Service tax returns or routine military documents.

and that agencies have partisan reputations, they are sufficiently constrained as to produce nonpartisan rules and regulations.

Measuring the ideological composition of agency outputs is an ideal opportunity for automated text analysis. One or more domain experts with unlimited time might be able to read through every rule, regulation, and policy that a given agency develops and score each document according to its partisan bent, but it would take a prohibitive amount of time. Instead, we might develop a supervised learning algorithm (see the Fundamentals of Text Analysis section) to automatically characterize the ideology of every rule and regulation produced by an executive agency. One approach, parallel to Gentzkow and Shapiro (2010), uses the Congressional Record as a reference point: If an agency rule uses language more similar to Democrats than Republicans, it is scored as a liberal document. By aggregating the ideology of bureaucratic documents across time, across agencies, and across issue domains, researchers might address problems of agenda setting, polarization, and politicization (or depoliticization) of an important and understudied set of political institutions. Another approach might involve hand-coding a subset of these regulations, as in Workman (2015), rather than relying on the assumption that the partisan cues in the Congressional Record hold for regulations as well.

Control of the Bureaucracy

How do bureaucrats respond when the preferences of the executive branch and the preferences of the legislature diverge? Presidents can issue binding directives to those agencies; Congress, on the other hand, controls agency budgets and crafts legislation attempting to circumscribe their behavior. Both the president and the Senate play a role in appointing the senior leadership of executive agencies and may be the most important mechanisms for controlling the bureaucracy (Wood and Waterman 1991).

Little of the research on this question is empirical, almost by necessity: There are no behavioral (rather than survey) measures of whether bureaucrats feel more beholden to the White House or to Congress, and it is challenging to quantify with traditional tools whether an agency's outputs conform to the preferences of the president or the legislature. Lowande (2019) studied the length of time it takes for an agency to respond to a congressional inquiry or request, finding that while agencies are more responsive to the requests from majority-party legislators, more politicized agencies respond most to requests from copartisans regardless of congressional control. But while response times are useful proxies for responsiveness, they cannot capture whether agencies' policy outcomes coincide more closely with presidential or congressional preferences, just as agencies can respond to legislators' requests quickly in the negative.

One promising approach to addressing this question leverages topic models. Topic models are useful for characterizing the overall composition of a large set of documents: for example, the percentage of presidential directives related to domestic versus foreign policy. Purpura and Hillard (2006) and Quinn et al. (2010) rely on topic models to map congressional outputs to the Policy Agendas Project. Extending this topic model to include both presidential and executive agency outputs, one might compare the proportion of attention both Congress and the president devote to any given element of the Policy

Agendas Project and calculate the branch to which each relevant executive agency's attentions align more closely. The Environmental Protection Agency (EPA) has a broad mandate to regulate emissions, control pollutants like insecticides, manage radiation, and promote clean water. If the president's agenda prioritizes clean water while Congress passes more bills related to emissions, a topic model applied to the EPA's output rules and regulations can identify whether its activity relates more closely to emissions, clean water, or neither.

Delegation versus Discretion

When do presidents or legislators direct executive agencies with unilateral actions, appointments, or legislation, and when do they instead offer agencies discretion to implement policy according to their expertise? Executive agencies must implement laws as passed by the legislature and signed by the president, but these laws often give bureaucrats little direct instruction, instead offering them substantial leeway in governing according to their expertise. A large literature in formal theory argues that Congress will write stricter legislation on topics of particular importance or when congressional leaders disagree with agency heads (Gailmard 2002; Lowande 2018; Weingast and Moran 1983). Validating these theories with empirics is challenging, however, because it is difficult to measure how much discretion or delegation Congress and the president provide to agencies.

Huber, Shipan, and Pfahler (2001) illustrate a rudimentary text data approach by measuring statute length as a proxy for delegation from state legislatures to state agencies—longer bills proscribing bureaucracies more than shorter bills—and show that divided government, a unified legislature versus a contrapartisan president, and a professionalized legislature all lead to longer bills with less bureaucratic discretion. Yet this measure is imperfect. Longer bills may simply relate to more complex topics, and bureaucrats may take great discretion even with very long bills if they are written as such.

Automated text analysis provides an avenue for solving this measurement problem similarly to the approach in Anastasopoulos and Bertelli (2018). Researchers might interview bureaucrats or agency heads to identify bills exemplifying discretion or delegation and then collect the texts of those bills. By training a model to identify the lexical indicators of bureaucratic discretion across a range of policy areas, it is possible to measure the extent to which any new bill restricts or empowers bureaucrats to use their expertise. Using these measures, we might directly observe the relationship between discretion and theoretically motivated explanatory variables: divided government, policy salience, and legislative professionalism.

Application: Measuring the Significance of Unilateral Actions with Text Analysis

Do presidents issue more, and more significant, unilateral actions when faced with a hostile or gridlocked Congress? Do they issue more, and more significant, unilateral actions when the Supreme Court is ideologically aligned, or when public opinion is in their favor?

An influential literature suggests that unilateralism is the defining characteristic of the modern presidency (Moe and Howell 1999) and that unilateralism has eroded the system of checks and balances in favor of a more powerful executive (Posner and Vermeule 2011; Rudalevige 2008). Much empirical research focuses on the constraints presidents face in exercising their unilateral authority, either formal or informal, from agencies, legislatures, or judiciaries (Chiou and Rothenberg 2014; 2017; Howell 2003; Lowande 2018; Reeves and Rogowski 2015). Dozens of studies examine these relationships, but despite an intuitively appealing conventional wisdom expecting that presidents exert unilateralism more frequently during periods of divided government (Belco and Rottinghaus 2017; Deering and Maltzman 1999; Mayer 1999; Waterman 2009), the empirical research largely finds the reverse: presidents appear to issue fewer unilateral actions when Congress is not controlled by the president's party (see, recently, Bolton and Thrower 2016; Lowande 2018).

This existing empirical work, however, relies on an incomplete picture of presidential unilateral power. Despite widespread acknowledgment that unilateral actions may take many forms (Bailey and Rottinghaus 2013; Belco and Rottinghaus 2017; Lowande 2014), previous research examines primarily numbered executive orders (Belco and Rottinghaus 2017; Chiou and Rothenberg 2014; 2017; Howell 2003; Lowande 2018; Moe and Howell 1999; Warber, Ouyang, and Waterman 2018) to the exclusion of unnumbered executive orders, presidential proclamations, executive memoranda, land orders, treaty proclamations, agency memos, and other documents that provide evidence of direct presidential action. This gap in the literature raises important methodological and substantive concerns. As the executive order became a more high-profile weapon in the arsenal of the presidency, presidents may strategically substitute lesser-known tools to implement important policies; this is especially a threat to the validity of studies that span longer periods of presidential behavior.

Capturing the complexity of presidential unilateralism in all its many forms requires (1) an exhaustive (or nearly exhaustive) data set of presidential unilateral actions to avoid the challenge of strategic substitution and (2) a measure of the significance of each of these documents to avoid conflating policy-driving actions with purely ceremonial ones;[7] to study the constraints on presidential unilateralism, (3) measures relating the preferences of the president to the preferences of the legislature and the judiciary are also necessary.

Kaufman and Rogowski (2019) address each of these three components in turn, first by accumulating a complete set of presidential unilateral actions by combining the *CIS Index to Presidential Executive Orders & Proclamations* and the *ProQuest Executive Orders and Presidential Proclamations, 1789–2017* corpora. This data set includes 31,377 unilateral actions from 1953 to 2016 along with their issuance dates and full text.[8]

7. There are, to the author's best knowledge, no theoretical reasons in the literature to expect that the legislature or judiciary would constrain nonideological actions such as World Freedom Day (Document 2005-PR-7960) or Save Your Vision Week (Document 1965-PR-3641).

8. The full set of unilateral actions from 1789 to 2017 includes 98,117 documents.

To produce measures of these documents' policy significance, the authors turn to automated text analysis. Because significance scores for a number of these documents already exist (Chiou and Rothenberg 2014; Howell 2003), Kaufman and Rogowski (2019) leverage those estimates to train a supervised learning model. By providing such a model with the text of numbered executive orders labeled with significance estimates produced by Chiou and Rothenberg (2014), the model can learn the relationship between word usage and policy significance. This model finds that words like "policy," "direct," "necessary," and "establish" indicate policy-relevant actions, whereas words like "hope," "began," "northeast," and "count" indicate ceremonial or otherwise insignificant directives.

As in the Fundamentals of Text Analysis section, the required inputs to this supervised learning model are a vector containing the dependent variable and a matrix containing the covariates. In this case, the matrix of covariates is a term-document matrix consisting of 10,746 significance-labeled unilateral action documents and 8,863 unigrams and bigrams. The dependent variable vector contains unilateral action significance scores drawn from both Chiou and Rothenberg (2014; for numbered executive orders from 1953 to 2002) and from research assistant hand-coding (for a subset of the remaining unilateral actions). There are many possible supervised learning algorithms available for this task; Kaufman and Rogowski (2019) use a random forest model (Breiman 2001).

The authors validate this trained model first by calculating its area under the receiver-operator characteristic (AUC), and next by conducting a series of internal and external validity robustness checks. The AUC, a common method of evaluating the quality of binary predictions, ranges from 0.5 to 1; the authors' model achieves an AUC of 0.91. Their robustness checks assess a variety of threats related to heterogeneous prediction errors: If directives from the 1950s are measured less accurately than directives from the 2000s, or if memoranda or proclamations are measured less accurately than executive orders, making inferences using the full variety and time series of unilateral actions may be misleading. For full details, see Kaufman and Rogowski's (2019) Appendix B.

Finally, to estimate the significance of the remaining unlabeled unilateral actions, the authors supply the trained and validated random forest model with new data, a term-document matrix consisting of 20,631 unilateral action documents *without* scores either from Chiou and Rothenberg (2014) or from research assistants. In general, this study finds that, contrary to much of the existing literature, presidents tend to issue *more* significant unilateral actions as their ideology diverges from that of the legislature, while there is no relationship between judicial ideology and presidential unilateralism. In summary, neither Congress nor the Supreme Court constrain the president's use of substantial and proliferating unilateral power.

The Methodological Future of Executive Politics

Often the most influential studies are not those that answer big questions, but rather those that open new fields of scholarship altogether. This article has given an overview of the two modes of machine learning for political science and a literature review of their uses in studying U.S. political institutions. I propose, through three research designs and one application, that the study of the executive branch stands to benefit enormously by adopting these methods. The research designs discussed in this article may be just the very first efforts to introduce text-as-data methods to executive branch scholarship, but they will be far from the last: as governments and agencies discover the value these archives hold for generating insights and predicting outcomes, they may become more willing to release them to academics for study.

But text data are not the last innovation in the study of the presidency and bureaucracy. In addition to text, the executive branch produces troves of image, audio, video, and network data. Recent work measures emotional states from audio and video (Dietrich, Enos, and Sen 2019; Knox and Lucas 2018); applying these tools to recordings of presidents might yield insights into how their emotional states affect their policy making or their bargaining capacity. Static images may prove useful in measuring how presidents interact with and appeal to diverse constituencies through social media. The pictures accompanying presidential tweets or press releases, or the contents of the staged presences constructed by the White House, may contain information about how presidents conceive of "visual frames" (Torres 2018). Network data of internal White House or agency communications can offer glimpses into flows of information and influence (ben Aaron et al. 2017).

These computational methods, while promising, must be applied with careful thought, consideration, and substantive expertise. Text-as-data methods are only as useful as the inputs that are used to generate the resulting estimates, and researchers must take great care to ensure that the estimates are as substantively meaningful as the documents they characterize. Moreover, as with every other data source, the credibility of the inferences from text-as-data–based estimates depends on the quality of the research design. While text-based methods offer substantial promise for allowing applied researchers to systematically measure key parameters in theories of the presidency and the executive branch, careful iteration among data, substantive expertise, and theory provides the greater opportunities for unlocking its potential.

References

Aletras, N., D. Tsarapatsanis, D. Preoţiuc-Pietro, and V. Lampos. 2016. "Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective." *PeerJ Computer Science* 2: e93.

Anastasopoulos, J., and A. M. Bertelli. 2018. "Understanding Delegation through Machine Learning: A Method and Application to the European Union." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3207821.

Anastasopoulos, L. J., and A. B. Whitford. 2018. "Machine Learning for Public Administration Research, with Application to Organizational Reputation." *Journal of Public Administration Research and Theory* 29 (3): 491–510.

Arnold, C., D. Doyle, and N. Wiesehomeier. 2017. "Presidents, Policy Compromise, and Legislative Success." *The Journal of Politics* 79 (2): 380–95.

Bailey, J. D., and B. Rottinghaus. 2013. "The Historical Presidency: The Development of Unilateral Power and the Problem of the Power to Warn: Washington through McKinley." *Presidential Studies Quarterly* 43 (1): 186–204.

Ban, P., D. Moskowitz, and J. M. Snyder. 2016. "Leadership Power in Congress, 1890–2014: Evidence from PAC Contributions and Newspaper Coverage." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2765092.

Ban, P., A. Fouirnaies, A. B. Hall, and J. M. Snyder. 2019. "How Newspapers Reveal Political Power." *Political Science Research and Methods* 7 (4): 661–78.

Belco, M., and B. Rottinghaus. 2017. *The Dual Executive: Unilateral Orders in a Separated and Shared Power System*. Stanford, CA: Stanford University Press.

ben Aaron, J., M. Denny, B. Desmarais, and H. Wallach. 2017. "Transparency by Conformity: A Field Experiment Evaluating Openness in Local Governments." *Public Administration Review* 77 (1): 68–77.

Benoit, K., K. Munger, and A. Spirling. 2019. "Measuring and Explaining Political Sophistication through Textual Complexity." *American Journal of Political Science* 63 (2): 491–508.

Bertelli, A. M., and C. R. Grose. 2011. "The Lengthened Shadow of Another Institution? Ideal Point Estimates for the Executive Branch and Congress." *American Journal of Political Science* 55 (4): 767–81.

Blaydes, L., J. Grimmer, and A. McQueen. 2018. "Mirrors for Princes and Sultans: Advice on the Art of Governance in the Medieval Christian and Islamic Worlds." *The Journal of Politics* 80 (4): 1150–67.

Blei, D., and J. Lafferty. 2006. "Correlated Topic Models." *Advances in Neural Information Processing Systems* 18: 147–54.

Blei, D. M., A. Y. Ng, and M. I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3 (Jan): 993–1022.

Bolton, A., and S. Thrower. 2016. "Legislative Capacity and Executive Unilateralism." *American Journal of Political Science* 60 (3): 649–63.

Breiman, L. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32.

Brown, P. F., P. V. deSouza, R. L. Mercer, V. J. Della Pietra, and J. C. Lai. 1992. "Class-Based n-Gram Models of Natural Language." *Computational Linguistics* 18 (4): 467–79.

Chang, J., S. Gerrish, C. Wang, J. L. Boyd-Graber, and D. M. Blei. 2009. "Reading Tea Leaves: How Humans Interpret Topic Models." *Advances in Neural Information Processing Systems 21*: 288–96.

Chen, J., and T. Johnson. 2015. "Federal Employee Unionization and Presidential Control of the Bureaucracy: Estimating and Explaining Ideological Change in Executive Agencies." *Journal of Theoretical Politics* 27 (1): 151–74.

Chiou, F.-Y., and L. S. Rothenberg. 2014. "The Elusive Search for Presidential Power." *American Journal of Political Science* 58 (3): 653–68.

Chiou, F.-Y., and L. S. Rothenberg. 2017. *The Enigma of Presidential Power: Parties, Policies and Strategic Uses of Unilateral Action*. New York: Cambridge University Press.

Clinton, J. D., and D. E. Lewis. 2008. "Expert Opinion, Agency Characteristics, and Agency Preferences." *Political Analysis* 16 (1): 3–20.

Clinton, J. D., A. Bertelli, C. R. Grose, D. E. Lewis, and D. C. Nixon. 2012. "Separated Powers in the United States: The Ideology of Agencies, Presidents, and Congress." *American Journal of Political Science* 56 (2): 341–54.

Dahl, R. A. 2005. *Who Governs? Democracy and Power in an American City*. New Haven, CT: Yale University Press.

Deering, C. J., and F. Maltzman. 1999. "The Politics of Executive Orders: Legislative Constraints on Presidential Power." *Political Research Quarterly* 52 (4): 767–83.

Denny, M. J., and A. Spirling. 2018. "Text Preprocessing for Unsupervised Learning: Why It Matters, When It Misleads, and What to Do about It." *Political Analysis* 26 (2): 168–89.

Denny, M. J., B. O'Connor, and H. Wallach. 2015. "A Little Bit of NLP Goes a Long Way: Finding Meaning in Legislative Texts with Phrase Extraction." Presented at Midwest Political Science Association (MPSA) 73rd Annual Conference, Chicago.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova. 2018. "Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding." Working Paper.

Diermeier, D., J.-F. Godbout, B. Yu, and S. Kaufmann. 2012. "Language and Ideology in Congress." *British Journal of Political Science* 42 (1): 31–55.

Dietrich, B. J., R. D. Enos, and M. Sen. 2019. "Emotional Arousal Predicts Voting on the U.S. Supreme Court." *Political Analysis* 27 (2): 237–43.

Dowding, K., A. Hindmoor, and A. Martin. 2016. "The Comparative Policy Agendas Project: Theory, Measurement and Findings." *Journal of Public Policy* 36 (1): 3–25.

Fenno, R. F. 1978. *Home Style: House Members in Their Districts*. New York: HarperCollins.

Gailmard, S. 2002. "Expertise, Subversion, and Bureaucratic Discretion." *Journal of Law, Economics, and Organization* 18 (2): 536–55.

Gentzkow, M., and J. M. Shapiro. 2010. "What Drives Media Slant? Evidence from U.S. Daily Newspapers." *Econometrica* 78 (1): 35–71.

Gentzkow, M., J. Shapiro, M. Taddy, et al. 2016. "Measuring Polarization in High-Dimensional Data: Method and Application to Congressional Speech." Technical report.

Grimmer, J. 2013a. "Appropriators Not Position Takers: The Distorting Effects of Electoral Incentives on Congressional Representation." *American Journal of Political Science* 57 (3): 624–42.

Grimmer, J. 2013b. *Representational Style in Congress: What Legislators Say and Why It Matters*. Cambridge, UK: Cambridge University Press.

Grimmer, J., and B. M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21 (3): 267–97.

Hemphill, L., A. Culotta, and M. Heston. 2013. "Framing in Social Media: How the U.S. Congress Uses Twitter Hashtags to Frame Political Issues." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2317335.

Hollibaugh, G. E. 2018. "The Use of Text as Data Methods in Public Administration: A Review and an Application to Agency Priorities." *Journal of Public Administration Research and Theory* 29 (3): 474–90.

Howell, W. G. 2003. *Power without Persuasion: The Politics of Direct Presidential Action*. Princeton, NJ: Princeton University Press.

Huber, J. D., C. R. Shipan, and M. Pfahler. 2001. "Legislatures and Statutory Control of Bureaucracy." *American Journal of Political Science* 45 (2): 330–45.

Katagiri, A., and E. Min. 2019. "The Credibility of Public and Private Signals: A Document-Based Approach." *American Political Science Review* 113 (1): 156–72.

Katz, D. M., M. J. Bommarito II, and J. Blackman. 2017. "A General Approach for Predicting the Behavior of the Supreme Court of the United States." *PLoS ONE* 12 (4): e0174698.

Kaufman, A. R., and J. C. Rogowski. 2019. "*Interbranch Conflict and Executive Action*." Working Paper.

Kaufman, A., G. King, and M. Komisarchik. Forthcoming. "How to Measure Legislative District Compactness If You Only Know It When You See It." *American Journal of Political Science*.

Kaufman, A. R., P. Kraft, and M. Sen. 2019. "Improving Supreme Court Forecasting Using Boosted Decision Trees." *Political Analysis* 27 (3): 381–87.

Keiser, L. R. 2001. "Street-Level Bureaucrats, Administrative Power and the Manipulation of Federal Social Security Disability Programs." *State Politics & Policy Quarterly* 1 (2): 144–64.

Khodabandelou, G., C. Hug, R. Deneckere, and C. Salinesi. 2013. "Supervised Intentional Process Models Discovery Using Hidden Markov Models." In *IEEE 7th International Conference on Research Challenges in Information Science (RCIS)*, 1–11.

Knox, D., and C. Lucas. 2018. "A Dynamic Model of Speech for the Social Sciences." Technical Report. Working Paper.

Laver, M., and K. Benoit. 2002. "Locating TDs in Policy Spaces: The Computational Text Analysis of Dáil Speeches." *Irish Political Studies* 17 (1): 59–73.

Le, Q., and T. Mikolov. 2014. "Distributed Representations of Sentences and Documents." *Proceedings of the 31st International Conference on Machine Learning* 7: 1188–96.

Liebman, B. L., M. Roberts, R. E. Stern, and A. Wang. 2017. "Mass Digitization of Chinese Court Decisions: How to Use Text as Data in the Field of Chinese Law." 21st Century China Center Research Paper (2017–01).

Lowande, K. S. 2014. "The Contemporary Presidency after the Orders: Presidential Memoranda and Unilateral Action." *Presidential Studies Quarterly* 44 (4): 724–41.

Lowande, K. 2018. "Delegation or Unilateral Action?" *The Journal of Law, Economics, and Organization* 34 (1): 54–78.

Lowande, K. 2019. "Politicization and Responsiveness in Executive Agencies." *The Journal of Politics* 81 (1): 33–48.

Mayer, K. R. 1999. "Executive Orders and Presidential Power." *The Journal of Politics* 61 (2): 445–66.

McFadden, D. 1980. "Econometric Models for Probabilistic Choice among Products." *The Journal of Business* 53 (S3): S13–29.

Moe, T. M., and W. G. Howell. 1999. "The Presidential Power of Unilateral Action." *The Journal of Law, Economics, and Organization* 15 (1): 132–79.

Monroe, B. L., M. P. Colaresi, and K. M. Quinn. 2008. "Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict." *Political Analysis* 16 (4): 372–403.

Mozer, R., L. Miratrix, A. R. Kaufman, and L. J. Anastasopoulos. Forthcoming. "Matching with Text Data: An Experimental Evaluation of Methods for Matching Documents and of Measuring Match Quality." *Political Analysis*.

Peterson, A., and A. Spirling. 2018. "Classification Accuracy as a Substantive Quantity of Interest: Measuring Polarization in Westminster Systems." *Political Analysis* 26 (1): 120–28.

Posner, E. A., and A. Vermeule. 2011. *The Executive Unbound: After the Madisonian Republic*. Oxford: Oxford University Press.

Purpura, S., and D. Hillard. 2006. "Automated Classification of Congressional Legislation." In *Proceedings of the 2006 International Conference on Digital Government Research*, 219–25. Digital Government Society of North America.

Quinn, K. M., B. L. Monroe, M. Colaresi, M. H. Crespin, and D. R. Radev. 2010. "How to Analyze Political Attention with Minimal Assumptions and Costs." *American Journal of Political Science* 54 (1): 209–28.

Radford, J., and B. Sinclair. 2016. "*Electronic Homestyle: Tweeting Ideology*." Working Paper.

Reeves, A., and J. C. Rogowski. 2015. "Public Opinion toward Presidential Power." *Presidential Studies Quarterly* 45 (4): 742–59.

Rice, D. 2019. "Measuring the Issue Content of Supreme Court Opinions." *Journal of Law and Courts* 7 (1): 107–27.

Richardson, M. D., J. D. Clinton, and D. E. Lewis. 2018. "Elite Perceptions of Agency Ideology and Workforce Skill." *The Journal of Politics* 80 (1): 303–8.

Riker, W. H. 1964. "Some Ambiguities in the Notion of Power." *American Political Science Review* 58 (2): 341–49.

Roberts, M. E., B. M. Stewart, D. Tingley, C. Lucas, J. Leder-Luis, S. K. Gadarian, B. Albertson, and D. G. Rand. 2014. "Structural Topic Models for Open-Ended Survey Responses." *American Journal of Political Science* 58 (4): 1064–82.

Roberts, M. E., B. M. Stewart, and R. A. Nielsen. Forthcoming. "Adjusting for Confounding with Text Matching." *American Journal of Political Science*.

Rudalevige, A. 2008. *The New Imperial Presidency: Renewing Presidential Power after Watergate*. Ann Arbor: University of Michigan Press.

Sim, Y., B. R. Routledge, and N. A. Smith. 2015. "The Utility of Text: The Case of Amicus Briefs and the Supreme Court." In *Twenty-Ninth AAAI Conference on Artificial Intelligence*.

Sulea, O.-M., M. Zampieri, M. Vela, and J. Van Genabith. 2017. "Predicting the Law Area and Decisions of French Supreme Court Cases." In *Proceedings of the International Conference Recent Advances in Natural Language Processing*, 716–22.

Torres, M. 2018. "Give Me the Full Picture: Using Computer Vision to Understand Visual Frames and Political Communication." https://qssi.psu.edu/new-faces-papers-2018/torres-computer-vision-and-politicalcommunication.

Warber, A. L., Y. Ouyang, and R. W. Waterman. 2018. "Landmark Executive Orders: Presidential Leadership through Unilateral Action." *Presidential Studies Quarterly* 48 (1): 110–26.

Waterman, R. W. 2009. "Assessing the Unilateral Presidency." In *The Oxford Handbook of the American Presidency*, eds. G. C. Edwards and W. G. Howell. New York: Oxford University Press.

Weingast, B. R., and M. J. Moran. 1983. "Bureaucratic Discretion or Congressional Control? Regulatory Policymaking by the Federal Trade Commission." *Journal of Political Economy* 91 (5): 765–800.

Wilkerson, J., and A. Casas. 2017. "Large-Scale Computerized Text Analysis in Political Science: Opportunities and Challenges." *Annual Review of Political Science* 20 (1): 529–44.

Wood, B. D., and R. W. Waterman. 1991. "The Dynamics of Political Control of the Bureaucracy." *American Political Science Review* 85 (3): 801–28.

Workman, S. 2015. *The Dynamics of Bureaucracy in the U.S. Government: How Congress and Federal Agencies Process Information and Solve Problems*. New York: Cambridge University Press.