Power in Text: Grammar and Language in Comparative Delegation Dynamics
[author redacted]
Abstract: How should scholars measure delegation of authority? Throughout political science, delegation represents a key concept of interest, motivating a large literature on institutional design in different political contexts. However, most analyses rely on either labor-intensive coding schemes or imprecise quantitative measures. Natural language processing offers an alternative approach. By using NLP tools, scholars can identify and measure key traits of particular texts, creating powerful and scalable measurement schemes. In this paper, I analyze and compare existing delegation metrics, and then use two novel NLP tools to generate a new measure for the level of power delegated by a given document. Finally, I test my measure on delegation data obtained from the Comparative Constitutions Project (CCP). Generally speaking, I find that the NLP-based measure performs substantially better than competing approaches, opening exciting new research questions regarding textual power structures and allocation of authority.
In political systems, most decisions are reached through delegation. Consciously or not, legislatures, constitutional conventions, and even voters constantly engage in the power-distribution business, deciding when, whether, and to whom to delegate decision-making authority. In the United States, federal statutes like the Clean Air Act and the Affordable Care Act delegate regulatory authority from Congress to the President, allowing executive agencies to set air quality standards and establish insurance exchanges, respectively. These laws also set restrictions on the use of that authority, subjecting executive decisions to legislative approval, judicial review, deadlines, and evidentiary requirements. Delegated powers thus play a critical role in institutional design and policy implementation, establishing institutional structures and relationships and guiding policy implementation.

As a practical matter, actors can delegate formally or informally, through written rules and procedures or unwritten norms, traditions, and power arrangements. However, in this paper, I focus on formal delegation of power: i.e., the privileges and restrictions contained in national constitutions and ordinary statutes. Formal power arrangements represent an important starting point for institutionally-oriented scholarship, structuring relationships and power dynamics across the various branches of government.

Empirically-oriented scholars interested in these formal delegation decisions have used one of two measurement strategies. The first, developed by Epstein and O’Halloran (1996, 385–387; 1999, 86–112), uses hand-coding procedures to count the powers granted by a given law. By contrast, Huber and Shipan (2002, 44–77; Huber, Shipan, and Pfahler 2001) use word counts as an alternative metric. Usually, they argue, longer statutes provide more specific policy prescriptions, and thus delegate less power.

---

1 Code, data, and raw text files used in validity tests may be obtained from the author by request. Acknowledgments suppressed for review.
Other authors have applied these methods extensively, and have used them to generate an important body of scholarship (e.g. Oosterwaal, Payne, and Torenvlied 2011; Franchino 2004; Franchino 2007; Randazzo, Waterman, and Fine 2006; Randazzo, Waterman, and Fix 2010). Unfortunately, both of these measures possess important limitations, especially when applied to large datasets. Epstein and O’Halloran, for their part, provide a strong conceptual scheme. Unfortunately, their measurement strategy is labor-intensive and relies on secondary legislative summaries, which limits generalizability. Huber and Shipan’s approach is more scalable, but much less precise. Legal documents can vary in length for many reasons; as a result, overall length may not be closely related to delegation of authority. Worse, Huber and Shipan’s method is not actor-specific; in other words, they cannot determine how much power was delegated to each actor mentioned in a given document. Within a given administration or government, different officials have different preferences, and different levels of autonomy from their politically-appointed masters. As such, the decision to delegate authority to a specific official or agency may be of substantive interest.

In this paper I develop and validate a new measure for formal delegation of power. First, I suggest that delegation can be viewed as a function of scope (number of powers delegated) and discretion (average restrictions per power), which combine to form the overall authority delegated by a given text. Throughout the paper, I focus on institutional restrictions on discretion. Next, I evaluate and critique both the Epstein and O’Halloran and Huber and Shipan approaches, and use natural language processing (NLP) tools to generate an alternative metric. Finally, I conduct validity tests on the Huber and Shipan metric and on my own system, comparing my results to data obtained from Elkins, Ginsburg, and Melton’s (2009) Comparative Constitutions Project. Elkins et al. hand-code an array of attributes in all national constitutions.
enacted since 1789, including elements of *scope* and *discretion*; as a result, their data provide a useful benchmark.

This study provides two primary contributions. Methodologically, this paper offers one of the first applications for a particular suite of NLP tools in political science research (specifically, grammatical parsers and text segmentation methods). Substantively, the NLP-based metric I develop is actor-specific, and shows greater efficiency than existing approaches. As a discipline, we know relatively little about the relationship between delegation and basic structural features of government (e.g. presidentialism vs. parliamentarism). NLP-based metrics can enable scholars to examine these kinds of broad, comparative questions, opening an array of exciting new research topics.

**Conceptualizing Delegation**

To borrow from Dahl (1957), whenever *A* delegates power to *B*, *A* faces at least two decisions. First, what powers should *A* allow *B* to exercise? And second, what restrictions should *A* place on *B*’s decision-making authority? These two choices – which I term *scope* and *discretion*, respectively – combine to determine the level of power offered by a given text. In this section, I describe both of these components separately, and explain how they combine to form *delegation*.

**Scope**

*Scope* refers to the number of powers allocated by a particular document to a particular institution or entity. All legal texts, whatever their purpose, address a certain set of topics and policy areas, empowering certain groups to use particular kinds of instruments and address particular issues (McCubbins 1985, 726). The US Constitution, for example, famously divides the powers of government between the legislative, executive, and judicial branches (and the
states). Different powers are allocated differently; for example, in Article 2, Section 2, the President exclusively controls Commander-in-Chief power, but shares treaty-making and appointment powers with the Senate (which must approve Presidential treaties and appointees). As with most constitutional provisions, these clauses are phrased in general terms, leaving their precise content open to interpretation. Nevertheless, the Constitution establishes the basic guidelines, outlining the general scope of the powers delegated to each institution of government.

Statutes, by contrast, are usually much narrower, and address a more limited set of issues. Even so, the same basic dynamics apply. Different statutes provide different numbers and kinds of tools to relevant agencies, which give those statutes different levels of scope. Under the Clean Air Act, for example, the EPA possesses an array of distinct regulatory devices, including national ambient air quality plans, state implementation requirements, limited emissions trading schemes, and special regulations on vehicle emissions, fuel composition, and ozone-depleting substances. The Clean Water Act, by contrast, offers fewer such instruments (and less thus scope), with a single standards-and-permitting system and a few specialized regulatory schemes for storm-water discharge and dredging/ocean dumping (Brownell 2011; McCall III).

Whatever the context, though, scope provides the first basic element of delegation. Documents that address more topics, and provide more regulatory tools, offer more scope than those that address and provide fewer. Importantly, scope is actor-specific; since each power is allocated to a particular institution (or divided between several institutions), a given document can provide different actors with different levels of scope. Precise definitions of these tools and topics depend on document type (e.g. constitutions, ordinary statutes, or administrative regulations). Constitutions, for example, provide broader instruments and address broader policy issues than do ordinary statutes. As a result, comparing scope across different document types is
probably not practical. However, across all document types, scope remains a defining element, maintaining the basic model I outline.

**Discretion**

*Discretion*, by contrast, refers to the restrictions placed on the exercise of a *given* power or regulatory instrument. Different scholars have used different conceptualizations to group these restrictions. McCubbins (1985, 724–729), for example, uses a multilevel scheme, separating structural restrictions (restrictions contained in the original authorizing statute) from management restrictions (new legislation/amendments, oversight hearings, and other post-hoc constraints). Within the structural category, he identifies a number of separate sub-types, including institutional constraints (i.e. number and identity of players authorized to act in a particular policy area), procedural requirements, and internal organization rules for relevant agencies. Epstein and O’Halloran (1999, 101) and their followers (e.g. Franchino 2004; 2007; Oosterwaal, Payne, and Torenvlied 2011; Ainsworth and Harward 2009) divide much further, identifying as many as 14 separate “constraint types” in the documents they examine. Finally, Huber and Shipan lump these categories together, measuring “policy specific” language rather than specific restriction types (Huber and Shipan 2002, 44–77; Huber, Shipan, and Pfahler 2001).³

---

² To expand on this point, constitutions operate at a different level than ordinary statutes, providing broad policymaking authority instead of targeted regulatory powers. As a result, comparing the powers offered by Usually, statutes and constitutions are drafted using similar linguistic practices and structures, making the basic models and metrics I propose applicable in both cases. However, these ideas should be applied separately across different kinds of documents, and not used to compare different types.

³ In one section of their book, Huber and Shipan actually do distinguish between different constraint types. As they note, at least some variation in statute length may result from procedural controls – i.e., “nonpolicy instructions about how the decision making by executives should proceed” (Huber and Shipan 2002, 56). To test for this possibility, they describe four types of procedural controls, and code a sample of statutes based on the percentage of language devoted to each type. However, after completing their validity tests, they abandon this conceptual scheme.
All of these conceptualizations are useful, and all help categorize discretionary restrictions. However, for the purposes of this paper, I focus on institutional constraints. McCubbins (1985, 725-6), in his “institutional” sub-category, provides a useful summary of these types of rules:

Regulations can be administered through civil or criminal suits in the courts, or through independent commissions or executives agencies, through discretion granted to the president or state and local entities […] The choice of institutional setting by [a legislature] involves a decision on how much independence [that legislature] wishes to grant the administrators (independence from [legislative] control) and the extent to which other decision-makers […] restrict or influence the choices of the administrators.

These institutional restrictions can be extremely powerful. By manipulating the background institutional structure, legislators can force implementing actors to cooperate with other (often hostile) players, altering downstream policy outcomes and restricting implementer discretion. Policy language and procedural restrictions, on the other hand, are not always so strong. As with all textual provisions, evidentiary requirements, deadlines, and substantive decision-making standards only constrain executive discretion to the extent that they are actually enforced. But, virtually all legal language is compatible with an array of interpretations and implementation styles. Institutional structures can help sharpen these constraints, assigning outside groups to monitor, amend, and approve agency policies.

As such, in this paper I operationalize an actor’s discretion on a given power by the number of other actors involved in the execution of that power. Obviously, this metric contains some significant limitations. Most importantly, involving additional players in a given policymaking process will not constrain administrative discretion by a constant amount. Instead,

---

4 In this context, I define “involved” quite broadly. For my purposes, any actor with some legally-prescribed role in a given policymaking process is “involved” in that process. So, actors with veto authority, legally-mandated consultation requirements, and policy proposal/amendment authority would all be “involved” in the execution of the relevant powers. For a given power, then, the universe of “involved” actors may be much larger than the set of veto players.
the size of the added restriction will depend (in part) on the *ideological distance* between existing players and new ones. If a legislature gives a court oversight authority over an existing policy area, that court will impose sharper restrictions on administrative discretion if it tends to disagree with agency judgments. This problem is somewhat analogous to Tsebelis’ (2002, 230) critique of the “numerical criterion.” Under the veto-players framework, counting the number of veto players forms “a questionable basis for comparative statics across different countries […] it ignores differences in the ideological positions of different veto players, which may be very important from one country to another” (Ibid). Worse, if legislators strategically introduce ideological allies into relevant policymaking processes, that choice will affect the relationship between institutional structure and administrative discretion.

Nevertheless, institutional restrictions provide an important starting point. By involving additional players in a given policymaking process, legislators will virtually always reduce existing actors’ discretionary authority. Even if their policy preferences are relatively similar, different actors will have different institutional interests; or, to paraphrase Madison, “ambition” in one institution can “counteract ambition” in another (Federalist 51, Hamilton, Madison, and Jay 2008), even if both actors desire similar policy outcomes. Moreover, though legislators almost certainly design institutional structures in a strategic fashion, they cannot predict the precise ideological relationships between different actors. Simply introducing more players into a particular policymaking process provides a “safe” alternative, allowing lawmakers to spread policymaking power across multiple actors and reduce overall discretionary authority. As such, more complex decision-making processes should virtually always indicate lower administrative discretion (though the specific relationship between institutional constraints and discretionary authority may not remain constant).
Scope, Discretion, and Delegation of Power

Thus conceptualized, scope and discretion combine multiplicatively to form the overall level of power delegated by a given document. As mentioned earlier, delegation is actor-specific; individual documents can delegate different levels of power to different actors. Documents that allow an actor to intervene in more policy areas delegate more power to that actor; conversely, documents that place more restrictions on a given actor’s actions delegate less. By imposing more discretionary restrictions, a legislature or constituent assembly essentially withholds some percentage of the “scope” it delegates, limiting the options available to implementing actors.\(^5\) As I describe later in this paper, “delegation” may thus be thought of as “weighted scope,” with discretion acting as a kind of a “weight” on the scope of powers delegated to a given actor.

Existing Metrics

Over the last fifteen years, political scientists have employed a number of different metrics to analyze the level of executive discretion contained within a particular text. However, two have gained particular traction: specifically, Huber and Shipan’s (2002) word-count approach, and Epstein and O’Halloran’s (1999) “discretion index.” In this section, I address these models in turn, discussing the positive and negative features of both approaches. To illustrate some of these critiques, I use biodiversity policy as a case study, comparing biodiversity law and implementation in the US, Australia, and Canada.

Epstein and O’Halloran

\(^5\) This conceptualization is very similar to that used in Epstein and O’Halloran (1996, 93-109)’s “discretion index,” though with an important difference. As described in the next section, Epstein and O’Halloran identify a set of 14 “constraint types,” which they treat as an exhaustive list of possible constraints on executive discretion. Here, I make no such assumptions, leaving the possible number of constraints open-ended.
As mentioned earlier, Epstein and O’Halloran’s (1996; 1999) conceptualization is basically compelling. Their primary measure – the “discretion index” – is composed of two basic parts: namely, the “delegation ratio” and “relative constraints” (Ibid, 90). Using *Congressional Quarterly*’s year-end legislative summaries, they count the number of “major provisions” contained in their statutes of interest. Next, they determine the proportion of those provisions which delegate some authority to the executive, which they term the “delegation ratio” \( (r) \) (Ibid). Then, they identify 14 types of “constraints” on executive action, and count the number of those constraint types that occur at least once in each statute they examine. These “constraints” include both procedural and institutional restrictions, including deadlines, consultation requirements, legislative vetoes, and appeals provisions (Ibid, 101).^6^ They then divide the number of constraint types present by the total number of constraint types they identify \( (f) \) and scale the result by the delegation ratio, producing the “relative constraints” imposed by a particular statute \( (c_i = f_i \ast r_i) \) (Ibid, 104). Finally, to produce their “discretion index,” they subtract the “relative constraints” value from the “delegation ratio” \( (d_i = r_i - c_i) \) (Ibid, 108).

This method possesses some clear advantages. By relying on *CQ*’s year-end summaries, Epstein and O’Halloran can easily measure the *scope* and *discretion* offered by the statutes they examine, granting their coding system a level of cross-contextual consistency. Their “constraint” (*discretion*) conceptualization does contain some odd elements; most notably, they count a constraint type as being present in a given document if that document contains at least one instance of that constraint, rather than counting the total number of constraints present in a given document. In addition, Epstein and O’Halloran equally weight all 14 of their constraint types.

---

^6^ The full list of constraint types they code includes: appointment power limits; time limits; spending limits; legislative approval requirements; executive approval requirements; legislative vetoes; reporting requirements; consultation requirements; public hearings; appeals procedures; rule-making requirements; exemptions; compensation; and, direct oversight (Epstein and O’Halloran 1999, 101).
However, some of the limitations they identify (e.g., time limits and reporting requirements) are likely less restrictive than others (e.g., appeals procedures, executive approval provisions). On the whole, though, their definitions of “delegation” and “constraints,” as well as their methods for combining these features, provide a sound delegation measure.

Unfortunately, Epstein and O’Halloran’s metric is difficult to generalize. Most obviously, in order to generate their list of “significant provisions,” the authors rely on the CQ’s year-end legislative summaries. However, even within their own work, these summaries do not cover all of the laws they examine; for example, in a study on financial legislation, the authors were forced to use the Federal Deposit Insurance Corporation’s website to code some statutes of interest (Epstein, O’Halloran, and McAllister 2010, 15). In projects examining other kinds of legal texts – e.g., statutes passed by subnational governments, national constitutions, or non-US laws – these kinds of resources are not likely to be available. Moreover, their method is quite labor intensive, requiring trained researchers to code complicated variables in large numbers of texts. In projects examining large, comparatively-oriented datasets, then, Epstein and O’Halloran’s procedure is not practical.

Huber and Shipan: Word Counts

Huber and Shipan’s delegation metric is fairly straightforward. At the outset of their discussion, they argue that, when faced with “two statutes that address the same issue, the longer one typically places greater limits on the actions of other actors” (Huber and Shipan 2002, 45). They operationalize “length” through the word counts of their statutes of interest, and apply their method to state-level Medicaid legislation passed from 1995-1996. Because of a series of exogenous shocks, a number of states were forced to reform their Medicaid programs during this period, giving the authors a temporally and topically homogenous dataset from which to work.
Huber and Shipan conduct an array of validity tests, with basically encouraging results. As the authors themselves note, wordier statutes may simply contain a higher proportion of “general” (non-operative) language, as legislators fill their statutes with vague platitudes and aspirational statements. Longer laws may also contain more procedural restrictions, imposing administrative requirements without constraining substantive policy choices. To test for these problems, the authors hand-code 67 statutes, calculating the percentage of each document devoted to “general” and “procedural” topics (Ibid, 52–64). On both counts, their results are fairly promising; they find that 46% of all statutes devote 0-20% of their text to “general” language, and 36% devote 0-20% of their text to “procedural” language. In addition, they find that statute length is negatively correlated with percentage of words devoted to “general” topics ($r = -.27$), which suggests some cap on amount of general language.\(^7\)

More troublingly, though, Huber and Shipan ignore scope. As noted earlier, the authors assume that the scope of each document they examine is roughly the same, forcing any variation in statute length to result from differences in discretion. If their assumption is accurate, higher word counts should be correlated negatively with delegation. In most contexts, though, this assumption is not realistic. Though some statutes and constitutions may be more similar than others, all legal documents possess some variation in scope. And, at least intuitively, scope should be positively related to word count (i.e., more words should indicate more powers delegated). As a result, the relationship between word count and delegation may be either positive or negative, depending on the effect sizes associated with scope and discretion.

Finally, Huber and Shipan cannot differentiate between the powers delegated to different actors. Even within the executive branch, different actors will possess different interests and

\(^7\) Their validity tests for non-US legislation are less extensive, but similar to their American results (Ibid, 65–72). In a sample of 30 statutes across all nations they examine, they find that 1.3% of all words are devoted to “general” topics, and 2.3% are devoted to “procedural” topics.
different levels of freedom from their politically-appointed masters. As a result, the decision to delegate power to a particular agency or actor may be of substantive interest. However, Huber and Shipan’s measurement strategy does not allow scholars to study statutes in this fashion, and can only measure broad, statute-level delegation.

*Statutes in Action: Biodiversity Law in the US, Australia, and Canada*

Endangered species statutes in the US, Australia, and Canada provide a good illustration of these issues. Like many economically advanced democracies, all three of these nations possess powerful, well-developed biodiversity laws. By raw word count, Australia’s Environment Protection and Biodiversity Conservation Act (EPBC Act) is the longest, at 66,871 words; by contrast, Canada’s Species at Risk Act (SARA) and the US Endangered Species Act (ESA) possess 22,441 and 22,185 words, respectively. Based on the Huber and Shipan methodology, then, we would expect the EPBC Act to be the most restrictive of the three, with SARA and the ESA essentially tied for second place.

However, empirical findings do not support this hypothesis. For starters, most of the additional wordage contained in Australia’s EPBC Act is attributable to the statute’s *scope*, rather than additional restrictions on administrative flexibility. The EPBC Act, in short, is simply broader than the other two laws, allowing agencies to take more types of actions and use more types of tools than either the ESA or SARA (Shaffer 2013). On those issues it does address, though, the ESA is almost certainly the most restrictive of the three. Compared with the Australian and Canadian laws, the ESA possesses somewhat stronger procedural and policy directives, particularly during the later parts of the biodiversity protection process (Ibid). More significantly still, the ESA also allows more agency decisions to be reviewed in court than either

---

8 The full EPBC Act is actually much longer, at 261,396 words. However, the EPBC Act addresses a number of topics besides biodiversity conservation; as a result, for the purposes of this exercise, I count only the text contained under subheadings that directly address endangered species protection.
SARA or the EPBC Act, creating a more restrictive institutional structure (Ibid; Opalka and Myszka 2009). As a result, American biodiversity protection agencies possess less discretion than their Canadian and Australian counterparts, who are much freer to act.

One response to this particular example might be that these statutes are too temporally and substantively separated to properly compare. After all, the ESA was passed in 1973, while the EPBC Act and SARA were passed in 1999 and 2000, respectively. In addition, the EPBC Act addresses an array of non-biodiversity topics, including wilderness preservation and cultural heritage. In the word count listed above, I only include those provisions of the EPBC Act that explicitly address biodiversity-related topics (see fn. 6). But, these sections may not be so easily separable, giving the EPBC Act an unwarranted wordiness boost.

But, if Huber and Shipan’s metric is truly that limited, serious questions about its generalizability become apparent. Though closely-linked political entities like the US states may sometimes pass “bursts” of legislation on the same topic within a short period of time, a collection of countries will not always follow the same pattern. Rather, as in the case of biodiversity protection, issues are likely to become salient at different times in different countries, creating temporal space between policy statements. And, substantively, the topics addressed by these three laws are about as homogenous as we are likely to find. Even if the EPBC Act is excluded from the comparison, SARA and the ESA address a basically identical set of topic areas, making them ideal candidates for a comparison of the sort that Huber and Shipan propose. However, despite their similar word counts, SARA and the ESA offer diametrically opposite levels of delegation; throughout the literature on biodiversity protection, the ESA is “widely regarded as the strongest [most restrictive] legislation ever devised for the protection of

---

9 Indeed, at least one set of authors has offered precisely this criticism, and advocated for an Epstein and O’Halloran-type measurement scheme (Ainsworth and Harward 2009, fn. 7).
non-human species” (Doremus 1991, 265). By contrast, Canadian officials routinely ignore SARA’s deadlines and procedural requirements, emphasizing the importance of institutional constraints (VanderZwaag, Engler-Palma, and Hutchings 2011; Shaffer 2013). Huber and Shipan’s methodology, then, may not be as generalizable as it first appears.

Overall, then, both measures leave room for improvement. Epstein and O’Halloran, for their part, provide sound conceptualizations and measurement schemes. However, their measure is highly labor-intensive, and difficult to generalize beyond the US context. Huber and Shipan avoid these problems; however, their measure likely contains a significant level of random error.

In addition, though their assumptions regarding delegation seem reasonable, the relationship between word count and overall delegation is not immediately obvious, and requires further validation. Finally, Huber and Shipan’s method cannot identify the level of power delegated to specific executive-branch actors. As a result, scholars interested in investigating delegation to specific officials and agencies in large datasets must employ a different measurement strategy.

“Relevant Sections”: Natural Language Processing and Formal Delegation

In response to these shortcomings, I use natural language processing (NLP) tools to develop and test an alternative delegation metric. This measure – which I term the “relevant sections” metric – uses text segmentation tools and grammatical relationships to identify the sections of a document that actually describes the powers given to the actor (or, the sections that are “relevant” to that actor). The “relevant sections” metric has three major advantages over existing approaches. First, compared with Epstein and O’Halloran’s approach, the “relevant sections” count is scalable and generalizable, allowing scholars to investigate large collections of texts. Second, compared Huber and Shipan’s approach, the “relevant section” count cuts out
most general and procedural language automatically, focusing on policy-relevant provisions. Finally, the “relevant sections” metric is actor-specific; as a result, it allows scholars to measure the level of power delegated to particular officials and agencies.

Segmentation and Subject-Object Relationships

To identify the set of “relevant sections,” I count the number of segments of a document that contain at least one sentence with the actor of interest as the subject.

This approach requires some explanation. As I describe earlier, raw word counts possess a number of problems as a measure for delegation of power. However, as Huber and Shipan demonstrate, word count does seem to be correlated with amount of “operative language” contained in a given statute (albeit weakly). In other words, variance in statutory length does seem to be partly driven by variance in policy-specific language, rather than variance in general or procedural statements. As a result, statutory length is likely to be correlated with delegation; the key is to identify which sections of a document are relevant to a particular actor.

To accomplish this task, I start by segmenting documents into topically distinct units of analysis. Generally speaking, legal documents cover a broad set of topics, ranging between different subject areas and sectors of government. Splitting documents according to their topical organization can help focus the measurement process, dividing texts into comparable units of analysis. Here, I segment documents using the TextTiling algorithm (Hearst 1994; Hearst 1997). TextTiling is a widely-used procedure in computational linguistics that employs lexical cohesion to segment texts at substantive topical boundaries (see Online Appendix for further details). In the political science context, authors have used TextTiling to segment party manifestos, which contain distinct statements on a variety of different issues (Stuckenschmidt and Zirn 2012). Statutes and constitutions structured similarly, with each text divided into a number of topically-
distinct units (e.g. paragraphs, articles, or sections). As a result, TextTiling seems to offer a promising method to segment texts into substantively meaningful units.

After segmenting the text, I analyze grammatical patterns in each unit. Within legal texts, sentences with an actor as the subject almost always confer some kind of power to that actor. Consider these three examples:

1. The Secretary [of the Interior] shall […] determine whether any species is an endangered species or a threatened species because of any of the following factors:
   - the present or threatened destruction, modification, or curtailment of its habitat or range;
   - overutilization for commercial, recreational, scientific, or educational purposes;
   - disease or predation. (Endangered Species Act)\textsuperscript{10}

2. The President of the Republic shall be the Commander in Chief of the Hungarian Defense Forces (Fundamental Law of Hungary 2011, 23).

3. The Justices of the Court of Accounts of the Union shall be chosen […] one-third by the President of the Republic (Constitution of the Federative Republic of Brazil 1988, 54).\textsuperscript{11}

These three provisions come from different countries and different types of legal documents (two national constitutions, and one ordinary statute). Clearly, all three provisions delegate a power to some actor (the Secretary of the Interior in the first example, and the President in the second and third). In each case, the text delegates the power in question to the actor in the \textit{subject} position in the sentence. The first example clearly demonstrates this feature, delegating the “endangered species determination” power to the Secretary of the Interior. In the second example, the text mentions an array of actors and institutions, including the President, the Commander in Chief, and the Hungarian Defense Forces. However, by placing the President in the subject position, the

\textsuperscript{10} 16 U.S.C. §1533(a)(1)(A-C)
\textsuperscript{11} Both the Fundamental Law of Hungary, 2011, and the Constitution of the Federative Republic of Brazil, 1988, are stored on the Comparative Constitutions Project’s website, as well as on file with the author.
clause specifically delegates the “Commander in Chief” power to the chief executive. The third example deviates from this pattern somewhat, placing the President in a prepositional phrase set off by the word “by.” When rephrased in active voice, though, the subject/object pattern remains consistent; in the active form, that sentence would read, “The President shall choose one-third of the Justices of the Court of Account,” restoring the President to the subject position in the sentence. Words contained in these kinds of provisions, then, should be more closely linked to the powers delegated to the relevant actor.

(Table 1 about here)

**Implementation**

Using these two observations, I wrote a Python script to calculate the number of “relevant sections” in a particular legal document (described more fully in Online Appendix). The script takes two inputs: (1) the plain text of a given document, and (2) the name of the actor of interest. The script uses the TextTiling algorithm to segment the document, and iterates through individual unit. Each individual section is then parsed using the Stanford CoreNLP grammatical parser (see Fig. 1 for sample parser outputs). The script counts the number of segments that contain the actor of interest as a subject, and returns the logged count (defined below):

\[ d = \ln \sum_{i=1}^{n} S_i \]

Where \( d \) represents the total power delegated to a given actor and \( S_i \) is a dummy variable, coded 1 if the \( i^{th} \) segment contains at least one sentence in which the Head of State occupies the subject position.

(Fig. 1 about here)

**Validity Testing: Delegation in National Constitutions**
As a validity test for the “relevant sections” metric, I attempt to replicate delegation data drawn from the Comparative Constitutions Project (CCP). In particular, I use CCP data calculate target delegation values for the Heads of State in all in-force constitutions (n=177), and compare those values with the total word count and the “relevant sections” metric.\(^\text{12}\) First, I take all questions from the CCP survey that address the “assignment” of a particular power (e.g., power to appoint high court justices, or power to propose budgets; see Online Appendix for full list).\(^\text{13}\) For each “power assignment” question, CCP lists the institutions involved in the exercise of that power.\(^\text{14}\) Using this information, I identify all powers that the Head of State is involved in (scope), and calculate the total number of players involved in each power.\(^\text{15}\) I then take the inverse of the average number of players per power, which forms my measure for average discretion (\(\frac{1}{\text{avg. players}}\)). Thus conceptualized, discretion ranges from 0 to 1, with 1 representing full discretion (no other players involved in any decision made by the executive) and 0 representing no discretion (an infinite number of other players involved in all decisions made by the executive). I then use the following equation to generate a total delegation score for the Head of State:

\[
d = s \times v
\]

Where \(d\) represents the total power delegated to the Head of State, \(s\) represents the total scope granted to the Head of State, and \(v\) represents the average discretion granted to the Head of State.

\(^\text{12}\) CCP possesses data and plain text files for all 192 in-force constitutions, as well as most historical constitutions passed since 1789. Unfortunately, though, the plain text for most constitutions is generated from uncorrected OCR scans, making textual analysis difficult. At the time of submission, CCP has cleaned 177 out of a total of 192 in-force constitutions, which I use in my sample.

\(^\text{13}\) For a fuller description of each question, see http://www.comparativeconstitutionsproject.org/files/surveyinstrument.pdf.

\(^\text{14}\) Again, CCP defines “involved” quite broadly; veto authority, advisory capabilities, and similar legally-defined roles are all sufficient to code an institution as “involved” in a particular power.

\(^\text{15}\) “Other” responses are excluded from this count.
(where \textit{discretion} is defined as \(\frac{1}{\text{avg. players}}\)). As mentioned earlier, \textit{delegation} may thus be thought of as \textit{weighted scope}; here, \textit{discretion} acts as a kind of a “weight” on the \textit{scope} delegated by a given text, withholding some decision-making authority from the implementing actor.

For methodological simplicity, this approach relies on two key assumptions, both of which may be theoretically problematic:

1. \textit{Preferences of involved players are randomly distributed.} Throughout my analysis, I operationalize \textit{discretion} as the number of other players involved in the exercise of a particular power. As noted earlier, though, this assumption is probably not realistic, particularly if legislators are capable of delegating strategically. By ignoring ideological distances between relevant players, the measures I propose thus ignore a potentially important source of variation. However, for reasons stated earlier, this metric is still likely to capture basic delegation dynamics.

2. \textit{All powers are equivalent.} As implied by the equation above, I weight each power identified in the CCP survey equally. Clearly, this assumption is not realistic; within any legal document, some powers are more important than others. However, determining the appropriate weights for these powers is not easy. To illustrate, consider the US constitution; of the powers granted to Congress, the power “to define and punish Piracies” is clearly less important than the power “to regulate [inter-state] Commerce” (Art. 1, Section 8). But, is the power “constitute Tribunals inferior to the supreme Court” more important than the power “to declare War?” The power “to borrow money?” Or even the power “to raise and
support Armies” (Ibid)? And, how should we deal with powers that change in importance over time? For example, when the US Constitution was first written, the Commerce Clause did not have nearly the same importance that it now possesses. So, how should we deal with that provision’s extraordinary rise in importance? None of those questions are easy (or, likely, possible) to answer. Weighting all powers equally avoids this problem, making large-scale analysis possible.

Analyses

After extracting the relevant data from CCP, I then use the “relevant sections” measure to analyze each constitution. To replicate Huber and Shipan’s approach, I used a Python script to count the total number of words in each constitution, and took the natural log of that count.\textsuperscript{16} I then use the CCP data to calculate the total power delegated to the Head of State, and compare my results to those generated by the Huber and Shipan and “relevant sections” metrics.

As a face validity measure, I first examined the CCP data. Based on CCP values, \textit{scope} and \textit{discretion} are negatively correlated ($r = -.42$). This result matches predictions from McCubbins (1985), who argues that these two variables ought to follow a negative relationship. From a more empirical standpoint, Epstein and O’Halloran (1996, 106-107) find a similarly inverse relationship between \textit{scope} and \textit{discretion} as well, though the correlation they observe is weaker ($r = .266$).\textsuperscript{17} On their face, then, the CCP metrics seem to behave similarly to measures used elsewhere in the literature.

\textsuperscript{16} In their study, Huber and Shipan use Microsoft Word to obtain their word counts; however, word counts from Python and Word were essentially identical.

\textsuperscript{17} Compared with my conceptualization, Epstein and O’Halloran switch the sign on the value for discretion/constraint; according to their definition, a larger constraint value means less overall delegation, rather
Next, I turn to the indicators. Fig. 2 presents a convergent validity test, correlating CCP scores with the “relevant sections” and “total words” metrics. At least in this test, the “relevant sections” metric dramatically outperforms the “total words” measure, improving from a .15 to a .49 correlation with the CCP data (see Table 2).

(Fig. 2 and Table 2 about here)

Interestingly, for both measures, the correlations are positive; more language, in other words, seem to indicate more power delegated, rather than less. This finding runs counter to the assumptions used in Huber and Shipan’s study, and merits further investigation. As shown in Fig. 3, both the “total word count” and “relevant sections” metrics are positively correlated with scope, and negatively correlated with discretion (see Table 3). In a certain sense, then, Huber and Shipan’s basic measurement strategy remains reasonable, since rising word counts are correlated with lower levels of discretion. But, in the constitutional context, once scope is included in the dependent variable, statute length becomes positively related with overall delegation (scope * discretion). As such, if the statutes in Huber and Shipan’s Medicare reform dataset actually contain no variation in scope, word counts are likely negatively related to delegation. As mentioned earlier, though, this assumption is probably not realistic. Thus, without further validation, scholars should be cautious when applying Huber and Shipan’s findings and measurement scheme.

(Fig. 3 and Table 3 about here)

Conclusions

Based on these results, four points seem worth mentioning. First, delegation is an inherently difficult concept to measure, especially in documents that cut across different

\[ \text{than more. As a result, though the correlation they observe is positive, and mine is negative, the substantive interpretation of the two values is identical.} \]
countries and policy areas. Constitutions offer a particularly difficult case, addressing a huge array of country-specific topics and issues. Worse, many of the constitutions examined in this paper date from very different time periods, ranging from centuries-old documents to texts that are barely a year old. Developing automated measurement techniques that work in all of these contexts is not easy. Nevertheless, the “relevant sections” framework seems to provide a promising start, offering a useful framework for future research.

Second, and related, scholars should use caution when interpreting blunt-force metrics like total word counts. Variation in word counts can result from a variety of sources, many of which have little to do with delegation. As a result, if scholars wish to use word counts as an indicator for an underlying concept, they should try to eliminate sections not directly related to their concept of interest. The natural language processing (NLP) tools used to generate the “relevant sections” metric accomplish this goal as a metric for delegation of power, offering markedly greater efficiency than the “total words” alternative. As such, the “relevant sections” metric seems to provide a more desirable measurement approach.

Third, to echo Grimmer and Stewart (2013), scholars need to pay special attention to validity testing when using language-based metrics. Linguistics tools offer promising methods for large-scale, automated analysis of text, making them attractive for large-scale data analysis. However, even in relatively simple cases, linguistic indicators may not be simply or obviously related to the underlying concepts of interest. As such, scholars need to test their metrics extensively before deploying them. Existing, well-validated datasets like the Comparative Constitutions Project provide ideal validation opportunities, and should be leveraged whenever possible.
Finally, this study highlights the power of NLP-based metrics. Text segmentation tools and grammatical parsers have not been used extensively in political science (Grimmer and Stewart 2013, 3). In documents that display consistent grammar and usage patterns, though, parsers can provide useful measurement contributions, helping segment texts and identify sections that are relevant to particular actors and ideas. By combining these tools with other text-analytic techniques, parsers can strengthen other research methods in appropriate substantive contexts.

Overall, then, NLP technology offers a promising measurement approach for complex concepts like delegation of authority. As numerous authors have noted, many political variables are text-based; as a result, text analysis tools can allow political scientists to measure latent concepts in a consistent and automated fashion, without requiring large numbers of research assistant hours. Clearly, language-based measures require extensive validation before application. However, these concerns aside, NLP-based approaches open exciting new research agendas, allowing scholars to compare statutory and constitutional development across different regime types, different policy areas, and other covariates of interest.
Works Cited


### Table 1: Sentence Parsing Explanations

<table>
<thead>
<tr>
<th>Example Sentence</th>
<th>Actor</th>
<th>Position in Sentence</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>The President appoints the Supreme Court Justices.</td>
<td>President</td>
<td>Subject – provides a power</td>
<td>Here, the President acts as a conventional subject.</td>
</tr>
<tr>
<td>The President’s nominees are approved by the Senate.</td>
<td>Senate</td>
<td>Passive verb complement – provides a power</td>
<td>As written; reorganized in the active voice, the Senate becomes the subject.</td>
</tr>
<tr>
<td>The House of Representatives shall be composed of members chosen every second year by the people.</td>
<td>House</td>
<td>Passive subject – does not provide a power</td>
<td>As written, “House” is the subject of this sentence; reorganized in the active voice, “people” becomes the subject.</td>
</tr>
</tbody>
</table>
Table 2: Correlations Between CCP Delegation Values and Language Metrics
(N = 177)

<table>
<thead>
<tr>
<th>Metric</th>
<th>Correlation with CCP Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relevant Sections (log)</td>
<td>.49 (.36, .59)</td>
</tr>
<tr>
<td>Total Words (log)</td>
<td>.15 (.00, .29)</td>
</tr>
</tbody>
</table>

*95% confidence interval (Fisher’s Z-transformed) given in parentheses*
<table>
<thead>
<tr>
<th></th>
<th>Scope</th>
<th>Discretion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relevant Sections (log)</td>
<td>.49 (.37,.59)</td>
<td>-.21 (-.35, -.06)</td>
</tr>
<tr>
<td>Total Words (log)</td>
<td>.15 (.00, .29)</td>
<td>-.07 (-.22, .08)</td>
</tr>
</tbody>
</table>

*95% confidence interval (Fisher’s Z-transformed) given in parentheses*
Fig 1: Sample Parser Outputs (Stanford Parser dependencies, collapsed)

The Congress shall have Power To lay and collect Taxes, Duties, Imposts and Excises.
(Article 1, Section 8, US Constitution)

Here, Congress is the nominal subject ("NSUBJ") of a sentence; so, this section would be counted towards the "relevant sections" count for Congress.

det(Congress-2, The-1)
nsubj(have-4, Congress-2)
aux(have-4, shall-3)
root(ROOT-0, have-4)
nsubj(lay-7, Power-5)
nsubj(collect-9, Power-5)
aux(lay-7, To-6)
ccomp(have-4, lay-7)
xcomp(have-4, lay-7)
ccomp(have-4, collect-9)
xcomp(have-4, collect-9)
conj_and(lay-7, collect-9)
dobj(lay-7, Taxes-10)
appos(Taxes-10, Duties-12)
appos(Taxes-10, Imposts-14)
conj_and(Duties-12, Imposts-14)
appos(Taxes-10, Excises-16)
conj_and(Duties-12, Excises-16)

New States may be admitted by the Congress into this Union.
(Article 4, Section 3, US Constitution)

Here, Congress is the passive-voice subject ("AGENT") of a sentence.
nn(States-2, New-1)
nsubjpass(admitted-5, States-2)
aux(admitted-5, may-3)
auxpass(admitted-5, be-4)
root(ROOT-0, admitted-5)
det(Congress-8, the-7)
agent(admitted-5, Congress-8)
det(Union-11, this-10)
prep_into(admitted-5, Union-11)
Fig. 2: Correlation Between CCP Data and Language Metrics
(N=177, 95% CI)

- Total Words
- Rel. Sections

Correlation with CCP Delegation
Fig. 3: CCP Subcomponents and Lang. Metrics
(N=177, 95% CI)