Dear Readers,

This piece builds on my last few articles, but it itself is at an early, exploratory stage. There’s more thinking to be done here, more examples to draw, and more scholarship to engage with. Given that, I'll be interested in whatever reactions you may have; I'm at the stage where all comments are useful. Thanks in advance for your thoughts. --Anya

Selective Empiricism in Legal Interpretation
Anya Bernstein
anyabern@buffalo.edu

Introduction ........................................................................................................................................................ 1
I. Corpus Linguistics in Linguistics and in the Law ...................................................................................... 2
   A. Corpus Linguistics in Linguistics ........................................................................................................... 3
   B. Legal Corpus Linguistics ......................................................................................................................... 5
II. Which and Whose Ordinary Meaning? ..................................................................................................... 7
   A. The Language of Legislative Production .............................................................................................. 7
   B. The Language of the Governed ........................................................................................................... 12
   C. Signs of the Ordinary ............................................................................................................................. 14
III. The Status of Statutes ............................................................................................................................... 18
   A. How is This Language Different from All Other Language? .......................................................... 18
   B. Statute as Utterance versus Statute as Object .................................................................................... 21
IV. Meaning as Fact versus Meaning as Activity ......................................................................................... 22
V. Toward a Broader Empiricism in Statutory Interpretation .................................................................. 23

INTRODUCTION
Empirical inquiry has come to statutory interpretation—haltingly. This Article pushes it to go a little faster. I focus on one empirical approach that has recently gained prominence—legal corpus linguistics—to show that halting empiricism is almost as bad as none at all. Legal corpus linguistics, which bases conclusions about the ordinary meaning of legal texts on data sets of language use, introduces some aspects of reality to statutory interpretation. It provides an empirical basis for judicial claims about how everyday people speak and adds many new speakers to the traditional interpretive reliance on the intuitions of one or three judges. But it stops short of challenging the assumption that permeate legal interpretation or acknowledging the methodological choices it entails.

One reason corpus linguistics has quickly gained popularity is that it promises to provide neutral, correct answers about legal interpretation. There is a right answer to interpretive questions, proponents suggest, and we can find it. This notion promises particular benefits in a politically fraught field. As two prominent promoters put it, “the indeterminacy of the search for ordinary meaning” should not lead us to “find . . . a broad license for normative judgments about whatever
interpretation” improves “constitutional system.” Corpus linguistics—like empirical inquiry in general—seems to promise a path out of normative judgment, into simple and falsifiable facts.

Indeed, the approach has been promoted as a “scientific” solution to interpretive inquiry because it provides an empirical basis for claims about ordinary meaning.2

But this escape from normative judgment is illusory: as an enterprise that contributes to the formulation of law, legal interpretation unavoidably involves normative judgment. Legal corpus inquiry, for instance, involves selecting not only which texts and genres, but also which speakers and which practices, should count for giving meaning to law. Whether interpreters admit it or not, such choices take a position on what, and who, matters in our polity. That is, normative judgments pervade not only any search for the ordinary meaning of statutory terms, but even the decision that ordinary meaning is the thing to search for in the first place.

Limiting its realism to one particular arena, legal corpus linguistics ends up perversely supporting, and even relying on, some rather unrealistic tenets. I illuminate the limitations of legal corpus linguistics to argue for a broader, more expansive empiricism in legal interpretation. I argue that interpretive methodology should justify the speakers, audiences, and genres that interpreters use as reference points for interpretation. Legal interpretation should also incorporate substantive research on how statutes are produced, enacted, and implemented. Further, legal interpretation should take a cue from linguistics in treating the meaning of texts not as an inherent attribute, but as a social activity. In an overarching sense, legal interpreters should account for the indelibly normative quality of the decisions that go into interpretation; rather than trying to avoid normativity, an empirically grounded approach would acknowledge its place in legal inquiry. Relatedly, a truly empirical approach would be more interested in reality than in certainty: it would recognize the inherently messy nature of legal interpretation. Corpus linguistics itself suggests as much, if we care to listen.

I. CORPUS LINGUISTICS IN LINGUISTICS AND IN THE LAW

American legal texts frequently treat ordinary meaning as a touchstone for interpretation.3 Assertions about ordinary language routinely serve to justify legal interpretations and are central to interpretive theory as well.4 At the same time, there’s no agreement on just what ordinary meaning is.5 In judicial opinions and scholarly commentaries, it can indicate a meaning that is allowable; or “most frequent”; or “exclusive”; or “prototypical.”6 There’s so little consensus on what ordinary meaning is that judges sometimes invoke different understandings of ordinariness within the same opinion.7 Recently, a wave of judges and commentators have proposed using corpus linguistics to

---

2 Jesse Egbert and Clark D. Cunningham, Scientific Methods for Analyzing Original Meaning: Corpus Linguistics and the Emoluments Clauses, __; see Carissa Byrne Hessick, __.
5. Lee & Mouritsen, supra note 3, at 796-98.
6. Id. at 800-02.
7. See id. at 803-04 (noting that the Muscarello majority implies that ordinary meaning is determined by most frequent meaning at some points, but by common meaning at others; while the dissent uses the notion both as possible or acceptable and as most frequent meaning).
address the problem of ordinary meaning. This Part provides a brief introduction to corpus linguistics in the field of linguistics and in legal interpretation.

A. Corpus Linguistics in Linguistics

Corpus linguists analyze compilations of language use—informal conversations, broadcast shows, articles, diaries, tweets, or whatever—for pervasive, implicit patterns. This method grows out of functional linguistics, which aims to describe language patterns—and inconsistencies—from evidence of language use in various contexts. Functional linguistics tends to view language as part of larger cultural complexes through which meanings of all sorts are produced. Corpus linguists are interested in deriving language patterns not from overarching rules but from instances of actual language use. Their data sets are the corpora they put together: collections, sometimes but not always quite large, of various kinds of language use. These data sets can include any kind of language use the researcher finds relevant: newspapers and magazines, academic texts, radio or television programs, naturally occurring conversations, storytelling sessions, classroom interactions, or whatever else. Linguists often compile their own corpora to fit their specific research question, though they also use existing corpora compiled by others if the materials are relevant. For example, researchers might use academic journal databases to find patterns in academic language; or newspaper databases to find patterns in newspaper language; or recordings of everyday conversations or of narratives to find patterns in those. Compiling or choosing a corpus is thus a methodological decision that scholars must make and explain.

Materials entered into the corpus can be marked with various information: parts of speech; demographic characteristics; languages used; communicative settings; or whatever else. Corpus linguists then look for patterns in these data to answer questions about how people use language in social life. Researchers often use mathematically complex tools to find these patterns. In this sense, much corpus linguistics sits at the intersection of linguistics and computer science. Doing it well usually requires training in both.

---

8 Lee & Mouritsen, supra note 3; [collect cites].
9 [Greenberg, Silverstein, Benveniste]. Functional linguistics is usually distinguished from “formal” linguistics, which searches for underlying or innate human characteristics that give rise to universal language structures describable through fairly precise, static rules. [Chomsky] Very broadly, formal linguistics tends toward the deductive or model-driven form of social science inquiry, while functional linguistics tends toward the inductive or grounded-theory form.
10 And yet again very broadly, from the standpoint of the influential structural theory of Ferdinand de Saussure, formal linguistics is associated with an interest in langue, or the rule-bound, recognizably correct standard language, while functional linguistics is associated with an interest in parole, or the more variable and unpredictable instantiations of speech in social context. Saussure; Gal. This is not say that functional linguistics is a botanical exercise in collecting instances; functional linguists usually also aim to express generalizations and theorize their findings. So, for instance, formal linguistics is more apt to describe unexpected language uses as flukes or errors that don’t represent the larger grammatical scheme, while formal linguistics accepts language uses that defy expectations as representing some subset of uses and seeks to articulate descriptions of patterning that would encompass them. Some functional linguists do look for cross-linguistic universals. In contrast to formal linguistics, though, they tend to ascribe such patterns not to underlying or innate grammatical classification, but to shared functions: certain patterns may achieve certain effects, like managing information flow, maintaining conversational focus, and so on.
13 There are numerous ways to use corpora, not all of which require complex computational tools. But a firm grasp on the choices the method affords and the rationales that argue for one tool over another require some significant computational abilities.
Linguists can pursue a wide range of inquiries with this method. For example, corpora allow researchers to trace how people introduce, maintain, and revive topics in texts or social interactions; when people tend to use grammatically optional forms and what effect that has on the emerging discourse; which forms characterize different genres of text or interaction and how similar-seeming forms appear differently in different genres; and many other areas of potential interest.

Some more concrete examples may help specify that kinds of things corpus linguistics can illuminate. Here’s one: when speakers introduce new factors into conversation, they generally do so in dribs and drabs rather than in big blocks. By “new factor,” I just mean some object of discussion—a noun and its surrounding phrase—that has not yet been introduced into the conversation. Linguists have found that, across languages, speakers rarely introduce more than one new factor in a single clause. New factors, moreover, tend to get introduced as subjects of intransitive verbs or as objects of transitive verbs, not as subjects of transitive verbs. That is, in conversation, we typically find new characters introduced first, before they are described as doing something to something else. This “Preferred Argument Structure”—a hypothesis still being developed and refined decades after its introduction—suggests (though of course it does not prove) that people process new information best when it comes in discrete chunks. Although the idea developed through studies of speech, the patterned distribution of new information has implications for writers as well: it may be helpful to introduce a new idea before launching into assertions about how it affects others.

Another example from the corpus linguistics literature. Set phrases that occur in the same form over and over, “lexical bundles provide interpretive frames for the developing discourse” by setting the scene for further content. In conversations, speakers tend to use formulaic clauses (utterances that include both a noun and a verb) such as “I don’t know why” or “what do you think.” And they tend to drop formulaic clauses into conversations whole, as pre-formed units preceded or followed by speaker-specific information: “I don’t know why X,” “what do you think about Y?” American academic writing, in contrast, tends to use formulaic phrases, in particular noun rather, such as “the fact that” or “as a result of.” Moreover, academic writers tend to use word

---

15 The standard linguistic term is “noun phrase” or “NP.”
16 Linguists working in this area generally use the term “new information” to describe a noun-phrase that has not yet been introduced in a conversation. I use the less technical term “factor” simply to avoid the implication that the pattern applies particularly to new facts that a speaker wishes to convey to a listener. Rather, it applies to any conversational focus that has not yet had a part in the conversation—person, place, thing, experience, concept, etc. I emphasize this because philosophers of language occasionally claim that everyday conversation primarily involves the exchange of information in this less technical, more fact-disclosure, sense. Marmor. Decades of empirical study have demonstrated the opposite: everyday conversation is full of strategy, play, aesthetics, affect-maintenance, and a range of other features that subsume its informational content. Jakobson, Silverstein
17 See sources collected supra, note 14. There seems to be some cross-linguistic consistency for new factors rarely to come in as subjects of transitive verbs (what is called the “A” position in linguistics). Whether speakers prefer introducing new factors as subjects of intransitive verbs (the “S” position) or as objects of transitive verbs (the “O” position) may differ depending on the kind of language (ergative versus nominative) and the kind of discourse (more narrative, more discursive, etc.). Karkkanien 1996.
18 Biber 2009 at 285.
19 Biber 2009 at 284.
20 Biber 2009, at 299 (“Conversation prefers fixed continuous sequences of words, with a preceding or following variable slot.”).
21 Biber 2009 at 284.
bundles not as units but as frames into which they insert their own contents, such as “the … of the,” where the ellipses can be filled in with a range of options depending on the situation: “the case of the,” “the end of the,” “the fact of the,” and so on.22

A final example: in standard American English, speakers can use “that” as a connector between a clause like “I think” or “he claims” and the content that follows.23 Speakers do not have to use “that” as a “complementizer” in this way; this usage is grammatically optional. For instance, both “I think he likes ice cream” and “I think that he likes ice cream” are grammatical, idiomatic utterances. But it turns out that speakers use this option in patterned ways. Speakers are more likely to use the complementizing “that” when the material that follows it is syntactically complex, as well as when its content is surprising and unexpected.24 Moreover, when the two parts of the utterance it joins together are closely related and express the speaker’s own commitment to a particular state of affairs, a speaker is less likely to use the “that” complementizer, while a speaker distancing herself from the assertion that follows the phrase is more likely to insert a “that.” So, I am more likely to say “I think he likes ice cream” if I am fairly sure he really does like ice cream; more likely to say “he thinks that I like ice cream” if I do not, in fact, like ice cream. This optional use of “that” thus tends to depend on a range of factors, including surrounding linguistic structures and surrounding narrative content. And it often has a subtle “modal” function—that is, it indicates the speaker’s attitude toward what is being said.

It is in this kind of work that corpus linguistics shines. It illuminates hidden but pervasive linguistic patterns. These patterns structure our speech and writing in ways that we do not articulate or even recognize. And yet, overall, we conform to them, usually without realizing that we are doing so. It thus gives us a glimpse into how discursive structures, linguistic genres, and social contexts can impose constraints on participants that are all the more powerful for lying outside of conscious awareness—an function of culture that is tractable on the transcript page.25 At their most exciting, corpus linguistics’ findings are surprising, un-intuitive, yet relatable. That is because how we think we use language often does not quite reflect what we actually do with it.

B. Legal Corpus Linguistics

In the last few years, legal thinkers have become interested in harnessing the analytic power of corpus linguistics for the interpretation of laws. Several influential publications have pressed the method, amicus briefs using it to evaluate legal questions, several judicial opinions have used or proposed it in the resolution of litigation, and a number of scholarly conferences have now centered on the approach.26 It makes sense: access to the realities of language use promises to simplify and

---

22 Id.
23 In this use, “that” is called a complementizer. See Thompson, Sandra A. & Anthony Mulac. The discourse conditions for the use of the complementizer that in conversational English, 15 J. OF PRAGMATICS 237, 249-250 (1991); see also T. Florian Jaeger, Redundancy and Reduction: Speakers Manage Syntactic Information Density, 61 COGN PSYCHOL 23 (2010) (arguing that this pattern distributes information within discourse).
24 Stefanie Wulff, Stefan Th. Gries and Nicholas Lester, Optional that in complementation by German and Spanish Learners, in WHAT IS APPLIED COGNITIVE LINGUISTICS? ANSWERS FROM CURRENT SLA RESEARCH 100-01 (Andrea Tyler, Lihong Huang, and Hana Jan eds., 2018). Wulff et al. also discuss other circumstances that influence speakers’ use of complementizer “that,” a topic that has been “intensively studied . . . over the last 25 years.” Id. at 99-100.
25 Recognized, articulated discourse patterns, in contrast, can become available to be identified, and challenged, as forms of grammar, which “arise[s] from patterns in the way language is used by speakers.” Thompson and Mulac, supra note Error! Bookmark not defined. at 250.
26 Lee & Mouritsen; [collect cites]; amicus cites; Carpenter v. United States (Thomas, J., dissenting); BYU and U of Chicago conferences.
resolve interpretive questions. And it promises to give legal writers a testable empirical basis for assertions about those linguistic realities—a kind of assertion that already figures prominently in judicial opinions and scholarly work providing interpretations of legal texts.27

So far, legal corpus linguistic analysis has mostly looked for relatively simple frequency data. They have asked how many of a given term’s appearances in a corpus have co-occurred with some particular other term—a query known as a “collocation.” For instance, if we want to know whether a person who has a gun locked in his glove compartment would normally be described as “carry[ying] a firearm,” we might look to see whether the word “carry” usually appears with words having to do with cars. If, in contrast, “carry” usually appears with words having to do with individual humans, that would indicate that it is ordinarily used in the sense of “to carry [something] on one’s person” as opposed to “to carry [something] in a vehicle.”28

Legal corpus inquiries have also looked to the utterance-level context in which a given term tends to appear, aiming to discern how often it is used in some particular way as opposed to others (a “key word in context” or “KWIC” search). So, to figure out whether the term “vehicle” normally encompasses airplanes, we might search a corpus for all the sentences or sentence fragments in which “vehicle” appears. Then we would determine what proportion of those can reasonably be understood to refer to airplanes, or to encompass airplanes within their scope of meaning.29 Even this basic frequency analysis, note, requires taking interpretive positions: sentence fragments don’t necessarily announce the full scope of meaning that most users would attribute to “vehicle.”

Any kind of language use, in any kind of format, from any source, may form the contents of a corpus: what evidence to use for the inquiry depends on what the inquirer decides is relevant. In legal corpus inquiries, analysts have mostly used several large, publicly available corpora, in particular the Corpus of Contemporary American English (COCA), the News on the Web Corpus (NOW), the Corpus of Historical American English (COHA), and the Corpus of Founding Era American English (COFEA).30 Unlike corpus research in linguistics, most legal corpus analyses using these corpora has not explained in detail what exactly its corpus should consist of and why.

Corpus linguistics holds out the promise of giving interpretive theory an empirical basis it lacks by providing information about how real people use the language that appears in statutes and, by extension, how they understand it. Legal corpus linguistics’ suddenly popular realist impulse is striking in a field still dominated by normative theories.31 And it could have real benefits for legal interpretation. To a large extent, though, those benefits have been muted by the way this perfectly respectable social science methodology has been deployed, which perversely undermines the very empiricism it seems to further. When used this way, this selective empiricism conscripts a social science methodology into a larger project defined by legal fiction and unacknowledged assumption. In the following Parts, I consider those fictions and how they could be made more real.

27 Bernstein, Democratizing Interpretation, supra note __.
28 Mouritsen; Lee & Mouritsen, supra note __; Muscarello.
29 Lee & Mouritsen; see McBoyle v. United States, 283 U.S. 25, 27 (1931); [Eskridge]
30 The COCA is available at https://corpus.byu.edu/coca/; the NOW Corpus is available at https://corpus.byu.edu/now/; the COHA is available at __; the COFEA is available at __. All are compiled by researchers associated with the J. Reuben Clark Law School at Brigham Young University, which runs a project on corpus linguistics and law.
II. WHICH AND WHOSE ORDINARY MEANING?

Corpus linguistics work in linguistics usually uses corpora that represent the genre, and the speaker population, being analyzed. Scholars tracking how speakers introduce information in ordinary conversations use corpora of naturally occurring conversations or spontaneously produced narratives. Scholars tracking patterns in academic English use corpora of academic articles. Scholars studying the speech habits of particular sub-groups, dialects, or registers use corpora collecting examples of language use in those forms. That is, to understand how people use and interact with a particular kind of language, corpus linguistics analyzes situations where people use and interact with that particular kind of language. The language targeted for analysis determines the language collected for evidence.

If legal corpus linguistics followed corpus linguistics in linguistics, we would expect it to take into account the speakers who produce the kind of legal text being analyzed, as well as speech in that and related genres. But legal corpus linguistics has, for the most part, taken a different tack. The News on the Web (NOW) corpus, for instance, as its name suggests, collects news articles from around the English-language internet (effectively from the areas of the former British empire). The COCA collects language from American sources, including newspapers, journals, academic sources, and television and radio broadcasts. Neither collects utterances that go into the production of statutory text. And although one can imagine that snippets of laws occasionally make their way into newspapers, television shows, and so on, neither corpus is oriented around showing how speakers use or discuss statutory language. Thus, unlike corpus linguists doing linguistics, legal corpus linguistics proponents often use corpora that do not exemplify the kind of language they seek to interpret, and do not focus on the kinds of speakers who produce that language. The language targeted for analysis is thus distinct from the language collected for evidence.

The difference between the evidence legal corpus analysts compile and the object they aim to interpret is rooted in a conviction that statutory interpretation should focus on the “ordinary meaning” of statutory terms. On this theory, statutory text should generally be taken to mean what it would mean in ordinary, non-statutory contexts. And so, the idea goes, we should look not to the language of those who write the statutes but to that of those ordinary people that statutes address and whose conduct they authorize or constrain.

This approach to ordinary meaning is itself rooted in ideas about the democratic legitimacy of law. Statutes should provide notice to those they rule, and to do that they should be written in the language of the governed. Departing from standard linguistic research practices by using a general corpus to evaluate the specific language of statutes is thus a normatively motivated decision. Yet this fairly straightforward, even commendable, normative commitment turns out to raise a number of conundrums.

A. The Language of Legislative Production

So far, statutory corpus linguistics has not incorporated language use by those who produce legislative utterances into its analysis. Interestingly, legal corpus studies of Constitutional text have not been similarly averse to looking at the language use of the people who produced the text. In that area, legal corpus inquiries have included material representing speakers closely related to the production and early reception of Constitutional provisions specifically, such as records of the Constitutional Convention and of state ratification debates; early “federal and state statutes, executive department reports, and legal treatises”; and “official documents, diaries and personal
letters written by and to” “George Washington, Benjamin Franklin, John Adams, Thomas Jefferson, Andrew Hamilton, and James Madison.”

Recent work using these materials has attempted to discern how terms like “emoluments,” “officer,” and “bear arms” were used by people like those who wrote them into the Constitution. It is striking that no legal corpus study of a statutory provision has taken a similar interest in the language of people like those who wrote it into the statute.

Recent scholarship has begun to uncover the everyday practices—including the language practices—through which legislation is produced. One thing this research emphasizes is the diversity of people involved in statutory production. Although theories of statutory interpretation consistently place elected members of Congress in the role of statutory speaker, members of Congress themselves have a rather attenuated relationship to the writing of statutory text, and “do not do the actual drafting.” Instead, they provide overarching policy directions. The actual text of statutes is largely written by non-partisan employees in the House of Representatives’ and the Senate’s Offices of Legislative Counsel. Legislative history, meanwhile—documents describing in greater detail the policies that statutes are meant to promote and results they are meant to accomplish—is produced by “the staff most accountable to the members” of Congress.

Collections of language use associated with the production of statutory text are, of course, readily available. Statutes themselves are representative of that genre, and would be readily compiled into a corpus. One could add to that volumes of the Congressional Record, the official transcript of many legislative utterances. Transcribed recordings from C-SPAN, which films many live communicative interactions in Congress, would be relevant as well. A corpus of such Congressional utterances would quite literally “give voice to the will of . . . lawmakers” in legal interpretation, allowing us to interpret the language of statutes “through the lens of” the language of the people who produce them.

A statutory language corpus might include stylistically and pragmatically related texts like Congressional committee reports produced for the purpose of elucidating statutory meaning for the members of Congress. Committee reports have the added benefit of providing section-by-section statutory overviews. A collection of these section-by-section explanations, combined with the texts

---

32 Jesse Egbert and Clark D. Cunningham, *Scientific Methods for Analyzing Original Meaning: Corpus Linguistics and the Emoluments Clauses* at *6. This and related studies use the Corpus of Founding Era American English (COFEA), available at: https://lawnel.byu.edu/. COFEA also includes “books, pamphlets, and other written materials published in America between 1760 and 1799.”
33 Id.
35 Denis Baron, *Corpus Evidence Illuminates the Meaning of Bear Arms*, 46 HASTINGS CONSTITUTIONAL LAW QUARTERLY 509 (2019).
37 Id. at 908.
38 Id. at 908
39 Id. at 908.
41 Lee & Mouritsen, *supra note __* at 795.
42 Abbe R. Gluck, *Congress, Statutory Interpretation, and the Failure of Formalism: The CBO Canon and Other Ways that Courts Can Improve on What They Are Already Trying to Do*, 84 U. CHI. L. REV. 177, 209 (2017) (“[O]ne of the most important pieces of
of statutes themselves, may reveal interesting patterns in how Congressional staff translate terminology between the language of statutes and the less specialized language meant for increasing legislators’ comprehension of them.

Moreover, employees in the administrative agencies to be charged with implementing particular legislative commands play a key role in producing statutory text: “agencies have their own legislative counsel whose sole work is to review and draft legislation.” Administrative agencies, acting as institutions in their official capacity, propose legislation, draft legislation as the request of members of Congress, and review and comment on legislation already drafted by Congressional staff. Individual agency employees, speaking on their own or on behalf of the agency, also comment on draft legislation in less formal ways and without White House oversight. Indeed, agency bureaucrats likely have the best grasp of anyone on the terminological valences of the substantive areas their agency regulates; they guard those terms carefully to ensure that statutes reflect agency practices, abilities, and needs.

A corpus of legislative production language would thus reasonably include agency work product such as Federal Register notices regarding statutory provisions or terms; guidance documents explicating them; recordings of hearings or meetings with administrators; internal documents released through FOIA requests, and so on. As recent empirical inquiry has demonstrated, that language of statutory implementation is in practice also the language of statutory production.

Some might object that, even if we were to decide such a thing would be useful, a corpus of the language of legislative production should include the language only of elected members of Congress, not legislative or administrative agency staff. Both judicial opinions and legal theory sometimes draw this kind of distinction, with the implication that the writing of legislative staff is less legitimate because of their unelected status. Administrative agency employees, moreover, are generally left out of judicial and even theoretical discussions of statutory production altogether.

Yet if we are interested in how those who actually produce statutory text use language, it makes sense to take into account what we know about the realities of that process. Statutory text itself, just like committee reports, is not usually produced by elected members of Congress. It is primarily the product of other people who work for, with, and around members of Congress.

---

43 Jarrod Shobe, Agencies as Legislators: An Empirical Study of the Role of Agencies in the Legislative Process, 85 GEO. WASH. L. REV. 451, 455 (2017). See also Christopher J. Walker, Federal Agencies In The Legislative Process: Technical Assistance In Statutory Drafting I, REPORT TO THE ADMINISTRATIVE CONFERENCE OF THE UNITED STATES (2015) (“Federal agencies draft statutes. Indeed, they are often the chief architects of the statutes they administer.”)
44 Shobe at 908.
45 Shobe at 908.
46 [Shobe, supra note ___].
47 See Nicholas Parrillo, [Agency Guidance].
48 See, e.g., Matter of Sinclair, 870 F.2d 1340, 1345 (7th Cir. 1989) (emphasizing that a committee report was written by “a staffer, not a member of Congress”).
49 [Abbe Gluck and Lisa Bressman, Statutory Interpretation From the Inside, Parts I and II.; Shobe, supra note ___; Walker, supra note ___]
That’s important for a couple of reasons. First, it means that strict distinctions between materials produced by members of Congress and those produced by their employees do not reflect the empirical realities of statutory production. To the extent that statutory interpreters seek an empirical basis for their conclusions, it makes sense to take into account how the texts they interpret are produced. The Constitution sets up hierarchies of authorization for the enactment of statutes, but that does not tell us how those statutes are produced in practice. That’s partly because a command is not a description; and partly because enactment is just one step in statutory production.

Second, it means that there are even more people involved in the production of statutes than even the already multiple members of Congress. One of the great things about corpus linguistics is its ability to incorporate the practices of lots of language users to draw conclusions about language use. That is, corpus linguistics recognizes the inherently social and multi-party nature of language use and uses it to advantage. Yet legal corpus work has, for the most part, embraced this multiplicity only on the back end, in the corpora used for evidence about statutory language. Legal corpus analysts have not looked to the multiplicity of contributors to any particular statutory utterance. Instead, they have generally treated statutory text as unified, singular.

Ignoring the speakers who produce the utterances they interpret, legal corpus proponents tend to treat statutes as uncreated creators of social effects. Refusing to look to utterers to help give meaning to their utterances gives legal corpus proponents something in common with the poststructuralist literary critics who announced the “death of the author.” For both, “it is language that speaks, not the author.” Poststructuralists saw the author’s irrelevance as proof of the radical indeterminacy of the text: “a text is not . . . the ‘message’ of the Author-God[] but a multi-dimensional space in which a variety of writings, none of them original, blend and clash.” Legal corpus proponents analyzing statutes have similarly rejected authorial interference, but tend to see it as leading instead to the fixation of meaning: an opportunity to evaluate scientifically what a text really means by evacuating authorial intent and leaving only the language.

Note that corpus linguistics is not primarily oriented to revealing internal states like speaker intent (though it is sometimes used to make such inferences as well). Rather, the method is well suited to uncovering patterns and trends in language use. Including speakers who take part in the legislative process in a corpus for statutory interpretation thus does not necessarily violate a principle that judges should interpret statutory texts rather than search for legislative intentions.

Corpus-based statutory interpreters thus eschew inquiry into how legislators use language. A century of research, in contrast, indicates that communicative practice involves more than texts and audiences: corpus linguistics in linguistics, as well as related fields that study language use, treat the speaker-role as highly relevant to communication. Ferdinand de Saussure’s classic Course in General Linguistics imagined communication as happening between the foundation dyad of speaker and

---

50 Roland Barthes, *The Death of the Author*, in *IMAGE MUSIC TEXT* 142 (Stephen Heath trans. 1977) (“[W]riting is the destruction of every voice, of every paint of origin. Writing is that neutral, composite, oblique space where our subject slips away, the negative where all identity is lost, starting with the very identity of the body writing.)

51 *Id.* at 143. Though note that Barthes qualifies his attribution of authorial death as pertaining to narrative specifically: “As soon as a fact is narrated no longer with a view to acting directly on reality but intransitively, that is to say, finally outside of any function other than that of the very practice of the symbol itself, this disconnection occurs, the voice loses its origin, the author enters into his own death, writing begins.” *Id.* at 142.

52 [Barthes]
Erving Goffman’s understanding of participant framework argued that different roles describe different functions speakers may have in a particular communicative event. Later elaborations imagined the speaker position as even more complex. Over and over, empirical studies of communication have a place not only for the audiences of an utterance but for those who produce it.

According to linguists, then, understanding communicative practices requires considering the role of those who produce, articulate, and stand behind an utterance, too. This argument may apply with even more force to statutes because of their unusual, perhaps unique, pragmatic effects. Laws are, after all, unusually efficacious utterances, often creating or modifying obligations, rights, or states of affairs. The “felicity conditions” that enable the efficacy of these utterances, moreover, depend a lot on the social position of the “principal” role in the participant framework—the duly elected political branches that produce and enact the law through the legally mandated process. In other words, the speaker role in legal utterances is crucial both in the same way that speaker roles are crucial to any utterance, and also in a unique way that applies only to the unique activity of legislation.

What could account for legal corpus proponents’ rejection of this aspect of empirical inquiry? A normative commitment. Corpus proponents generally aim to give primacy to the language of the governed. In this, legal corpus proponents echo the tenets of textualism, for which ignoring the speaker of a legislative text is part of a normative democratic commitment to limiting legal power to the text enacted by the legislature: “our Constitution provides for the enactment and approval of texts, not of intents.” For corpus proponents, using ordinary meaning to interpret statutes “assures notice to the public, protects reliance interests, assures consistency of application, and respects the will of the legislative body.” And those same normative commitments seem to imply looking to the ordinary meanings of terms as used by the audiences of statutes, rather than by their producers.

In a rule-of-law democracy, those who make the laws are also subject to them. This is particularly true of administrative agencies: agencies participate in the production of statutes that they themselves will be charged with implementing. Indeed, while there’s much to be said for

---

53 Ferdinand de Saussure, Course in General Linguistics (1916). Saussure pictured the roles thus:

54 Erving Goffman, Forms of Talk 144-45 (1981) (disaggregating the speaker role into functions including principal, or entity committed to or bolstered by the utterance; author, or entity that chooses the form of expression; and animator, or entity that utters the words).

55 See, e.g., Judith T. Irvine, Shadow Conversations: the indeterminacy of participant roles, in Natural Histories of Discourse 134 (Michael Silverstein and Greg Urban eds. 1996) (this may be one reason why even self-reported textualist judges end up resorting to arguments based on statutory purpose; it’s hard to have communication with no speaker. Anita Krishnakumar, Backdoor Purposivism, __ Duke L. J. __ (2019).


57 Lee & Mouritsen, supra note __ at 793.

58 The normative valence of agency participation in legislation can vary across political views and even across democratic cultures. In separate research with administrators in the recently democratized country of Taiwan, for instance, I have
reading law through the language of the governed, that normative commitment raises its own methodological issues, as I discuss further in the next section.

B. The Language of the Governed

Legal interpreters wary of looking to statutory speakers for clues about what legislative language means tend to focus instead on a statute’s audience: the people addressed, commanded, and constrained by the statute. This fits well into a textualist ideal as well, which instructs judges to interpret statutory language as it would be understood by those subject to it, not those with the power to create it.60 Who are these people subject to statutory command? In legal corpus methodology, they are what we might call generics: producers of generically American text across a range genres that are published or broadcast nationally. The underlying image here is a vaguely Austinian notion of law as a command by a sovereign to his obedient populace, a populace of generics whose language use should determine the meaning of that command.

This does not, however, describe how most statutes, especially most federal statutes, work. For the most part, Congress does not enact commands to generics. It usually issues commands to administrative agencies. This is so even when private parties stand to benefit from the statute. So, for instance, when a federal statute gave oil producers a tax credit for oil produced “from . . . tar sands,” but left “tar sands” undefined, a court had to interpret the term.61 An oil company that stood to benefit from the tax credit argued that the court should accept its own understanding of the term, which may have reflected the understanding of others in the oil industry at the time. On this reading, tar sands included any material so viscous that it required something other than standard oil production methods. The government, in opposition, argued that tar sands included only specific rock types that required especially unusual, high-tech extraction methods, which would mean that the tax credit rewarded those who furthered technological innovation, not just those who produced oil in a more labor-intensive way than was standard.62

A recent article by James Heilpern argues that, since a generalist court naturally has trouble determining what constitutes tar sands for oil production purposes, this would be a good context in which to use legal corpus linguistics to determine how the oil industry really used the term.63 This comports with that classic Austinian image described above: a statute issuing a command to the general populace should be read through the understanding of that general populace. In this case, the image is modified a bit: that general populace would be rather narrower: a particular industry that uses particular terms of art. Heilpern argues that legal interpreters should conceive of industries as speaking their own specialized argot or dialect.64

found that administrators view agency participation in legislation as enhancing their legitimacy. Anya Bernstein, Porous Bureaucracy: Legitimating the Administrative State in Taiwan, ___ L. & SOC. INQ. ___ (forthcoming).


62 Id. at 214.


Yet, the statute at issue did not govern the conduct of the oil producers who would benefit from it. Rather, as part of the Internal Revenue Code, it issued its instructions to a government agency, the Internal Revenue Service, stating that “[t]here shall be allowed as a credit against the tax imposed by [other sections of] this chapter” a certain amount of money for certain kinds of fuel.\(^{65}\) Moreover, the Federal Energy Agency, implementing a related statute giving the Executive authority to set oil prices, had already produced a ruling defining “tar sands.”\(^{66}\) The specialized technical audience the statute addressed, in other words, was the government itself, and specifically, agencies with expertise in taxation and extraction. The petroleum industry, in contrast, though it stood to benefit from the statute, was not constrained or commanded by it. It surely makes sense for the court to consider the definition offered by the entity the statute addressed, rather than its potential beneficiary.

At other times, a statute might govern the conduct of another government employee. So, for instance, in *Taniguchi v. Kan Pacific Saipan, Ltd.*, the Supreme Court considered whether a statute allowing judges to award attorneys’ fees for “interpreters” referred only to those who worked with verbal expressions, or also to those who worked with written text.\(^{67}\) The majority looked, among other things, to the “relevant professional literature” of those who render expressions in other languages.\(^{68}\) In a majority opinion written for six justices, Justice Alito concluded that professional translators and interpreters would not include the person who renders written text in another language an “interpreter,” and so neither would the Court. Professional translators and interpreters do stand to benefit from the statute’s provision, but are they relevant communicative community? A fee-shifting statute, after all, authorizes and constrains the conduct of federal judges, who receive limited authorization to award fees. These legal professionals were not addressed by the technical literature the majority cited.\(^{69}\)

Legal corpus proponents generally propose hanging a statute’s meaning on the language of statutory audiences outside of government, such as generics or industries. And as *Shell Oil* suggests, the choice of audience can have a big effect on interpretative conclusion a court reaches. In practice, though, most federal statutes are not addressed to people outside the government. Instead, federal statutes predominantly mandate and constrain the conduct of administrative agencies, though, like the statute in *Taniguchi*, some statutes govern the conduct of judges instead. For many statutes, then, a corpus representing those subject to statutory commands would be centered on the language use of federal bureaucrats or other federal employees.\(^{70}\)

Such a corpus could utilize the range of materials discussed in the previous section: *Federal Register* notices, guidance documents, interviews with agency personnel, and other agency work product. Since we know that agencies maintain an almost proprietary attitude toward terms that are

---

\(^{65}\) *Id.* at 215-16 (quoting a provision then codified at 26 U.S.C. § 29; this provision was later amended and recodified at 26 U.S.C. § 45K).

\(^{66}\) *Id.* at 214 (quoting Department of Energy Ruling 1976–4, 10 C.F.R. ch. II Rulings at 371).

\(^{67}\) 566 U.S. 560 (2012).

\(^{68}\) *Id.* at 571 (2012).

\(^{69}\) *Id.* at 576-77 (Ginsburg, J., dissenting).

\(^{70}\) See Gluck & Bressman, supra note ___ (arguing that most legislation drafters take administrative agencies as their primary addressees).
central to their regulatory work, they could isolate portions of the corpus that relate to particular agencies and particular terms to assist with interpreting them.

Interpreters of law may want to exclude government employees from their corpora on grounds of democratic legitimacy. The theory goes that those who are subject to the power of law deserve notice of what law requires. And to determine whether they have proper notice, we should examine how they use language to see how they would understand a statutory text. Yet, again, a rule-of-law democracy inherently places some people in the position of both governor and governed. The main people most directly constrained by federal statutes are those who work in administrative agencies. If corpus analysts want to base their analysis in the empirical realities of statutory constraints—rather than in an unrealistic ideal of direct command from Congress to populace—then it makes little sense to leave bureaucratic language out of a legal corpus.

Building on the normative claims of textualism, legal corpus proponents tend to avoid treating governmental language practices as relevant to legal interpretation, preferring instead to base interpretation on the language use of the people whom statutes address and control. Yet, in reality, those people will very often be government employees themselves. So far, no proponent of legal corpus analysis has confronted the obvious implication: that bureaucratic language practices sit at the heart of statutory interpretation in the modern state.

As others have noted, one aspiration of a corpus analysis in linguistics is to “maximize[] the chances of achieving a representative corpus.” A representative corpus, in turn, is one whose “sample includes the full range of variability in a population.” Achieving representativeness of any population is difficult. But even more difficult in the legal context is settling on the population to sample. Which kinds of language, which speakers, and which audiences should populate the corpus on whose basis legal meaning is determined? The question is inherently normative and political. At its base, constructing a corpus for legal interpretation involves deciding not just what language, but also which people, should get to influence the meaning of law. Most legal corpus analysis proponents, however, approach this issue as unproblematic, as though the defining question were the number of words a corpus contains, rather than which words, and whose.

C. Signs of the Ordinary

Legal corpus linguistics has focused heavily on frequency. Writers seek to determine how frequently a particular term appears in a particular (purportedly representative) corpus; how frequently it co-occurs with other terms of interest; and how frequently it is used in a particular way rather than in other ways. This is the main way that corpus proponents have addressed an underlying problem: although legal writers often refer to ordinary language, there is no consensus on what constitutes “ordinary.”

---

71 Shobe, supra note __.
72 [Carissa Byrne Hessick, on notice.]
74 Id. at 1594 (quoting Douglas Biber, Representativeness in Corpus Design, 8 LITERARY & LINGUISTIC COMPUTING 243, 243 (1993)).
75 Id.
Proponents of corpus linguistics in the law have noted that judicial opinions, as well as scholarly commentaries, use the term “ordinary meaning” to refer to a range of things.\(^76\) “Ordinary” may refer to any meaning that is allowable—that is, any meaning that people would recognize and respond to appropriately.\(^77\) Or it may indicate an “exclusive” meaning: the one thing that a term means, in a way that out other possibilities.\(^78\) A usage that is “common” may be called ordinary in some cases, but sometimes judges aim to find not just a common but the “most frequent” usage.\(^79\) Alternatively, ordinariness may be a measure of the “prototypical” meanings that set the terms for normal and outlier versions of a particular expression.\(^80\) A prototypical meaning might be one that most speakers would agree a term has, even if they do not use it all the time themselves. Alternatively, “ordinary” may refer to what most speakers would say a term meant most of the time, even if they were, statistically speaking, wrong—something like popular opinion about prototypicality.

There is so little consensus on what ordinary meaning is that judges sometimes invoke different understandings of ordinariness within the same opinion.\(^81\) Finding frequencies seems to be one way to address the problem of ordinariness. It seems intuitive that a word ordinarily means what it is frequently used to mean. Yet, as others have noted, legal corpus users have tended to assume, rather than justify or test, the importance of frequency to their analysis.

There are pitfalls to this assumption. For instance, a term may appear frequently because the underlying phenomenon it describes occurs frequently.\(^82\) For instance, if a corpus shows that the word “discharge,” when used to describe someone shooting a firearm, usually indicates a single shot of an individual bullet, rather than multiple shots from the same gun.\(^83\) This high frequency may occur because the word “discharge” is the preferred way to indicate a single shot, as opposed to multiple shots from the same gun, which tend to be referred to by other linguistic means. But alternatively, the high frequency may occur because corpus participants generally talk about the firing of a single shot, as opposed to talking about multiple shots. It may even be that people who shoot guns more often shoot a single bullet than multiple ones—or at least that when people shoot guns in a way that corpus participants discuss, they do so with single bullets. An analysis that just asks whether the single-shot or multiple-bullet meaning shows up more frequently does not reveal which of these reasons underlies that finding. But if we are using the corpus to decide whether someone who illegally shot a bunch of bullets all at once from his gun is guilty of a bunch of crimes of “discharging” his gun or just one, then the reason that one version appears more frequently than another matters.\(^84\) Contrariwise, a word may be used frequently in a particular way because the

\(^{76}\) Lee & Mouritsen, supra note , at 796-98.
\(^{77}\) Id. at 800-01.
\(^{78}\) Id. at 800-01.
\(^{79}\) Id.
\(^{80}\) Id. at 801-02.
\(^{81}\) See id. at 803-04 (noting that the Muscarello majority implies that ordinary meaning is determined by most frequent meaning at some points, but by common meaning at others; while the dissent uses the notion both as possible or acceptable and as most frequent meaning).
\(^{83}\) Rasabout
\(^{84}\) Similarly, an infrequent appearance does not negate a word’s meaning or belonging to a particular category. In a now-classic example, the fact that dodos are extinct in the wild and are not referred to often in print does not detract from the dodo’s status as a bird; the ordinary usage of bird would include dodos even if the word “bird” is rarely used to refer to a dodo. Lee and Mouritsen at **.
underlying phenomenon it refers to is infrequent, and therefore particularly notable or newsworthy.\textsuperscript{85} If light switches tend to appear in a corpus as failing to turn lights on, that may be because light switches generally fail to turn lights on. Or, on the contrary, it may be because light switches generally do turn lights on, and speakers generally expect them to do so. Speakers may thus be more likely to comment on the infrequent light switches that fail to fulfill their expectation than on those commonplace ones that succeed.

Corpus linguistics research in linguistics also considers the frequency with which a term or linguistic phenomenon appears in the corpus, though the specific role of frequency remains a matter of debate in this developing field.\textsuperscript{86} Linguists have also noted that not all frequencies reveal the same things. The word “dog,” like any noun in English, will frequently appear with the article “the.” That co-occurrence does not tell us much about “dog”; it is simply a function of how English treats nouns. The “dog-the” co-occurrence is determined by the rules of English grammar, but grammatically optional co-occurrences also vary in how much they reveal about typical usages. Say we search a corpus of naturally occurring conversations for collocates of (that is, words that appear along with) “dog.” We may find “dog” frequently co-occurs with the word “his.”\textsuperscript{87} The frequent appearance of the phrase “his dog” certainly tells us that a dog is something that can be possessed by an individual. Many things can be possessed by individuals, though, so although a high proportion of the instances of “dog” may go along with the word “his,” a relatively low proportion of all instances of the word “his” will end up going along with the word “dog.” These words do not give much “Mutual Information”;\textsuperscript{88} they do not mutually predict one another’s presence. In contrast, the word “stray” will likely appear much less frequently in a conversational corpus than “his.” But a higher portion of the appearances of “stray” will go along with appearances of “dog.” We can thus be more confident in saying that “stray” tends to appear with “dog” than saying that “his” tends to appear with “dog,” even though “his” will appear with “dog” more frequently than “stray.”

Relatedly, “associations are not necessarily reciprocal in strength.”\textsuperscript{89} The presence of “stray” is much more likely to predict the presence of “dog” than vice versa. “Stray” is a less common word and it is strongly associated with dogs; that is, if something in a conversational corpus is stray, it’s likely to be a dog. In contrast, “dog” is a more common word and is not as strongly associated with stray-ness; that is, if something in a conversational corpus is a dog, it may be many things other than stray. This kind of “asymmetric” relationship requires “directional measures” that find not just

\begin{itemize}
  \item where terms go together but where one tends to predict the presence of the other.
\end{itemize}

For instance, the term “vehicle” may only rarely occur in situations implicating bicycles. But if “bicycle” usually occurs in situations implicating vehicles, that may indicate that ordinary language nonetheless categorizes a bicycle as a kind of vehicle. The bicycle-vehicle relationship, moreover, is a token-type relationship. If we want to know whether a bicycle counts as a vehicle in a particular corpus, it is not necessarily sufficient to look for co-occurrences of “bicycle” and “vehicle,” since mentions of

\begin{itemize}
\item \textsuperscript{85} Ethan J. Herenstein, \textit{The Faulty Frequency Hypothesis: Difficulties In Operationalizing Ordinary Meaning Through Corpus Linguistics, 70 STAN. L. REV. ONLINE 112, 114 (2017).}
\item \textsuperscript{86} But see Biber 2009, \textit{supra note **}, at 280 (“The role of frequency and quantitative analysis in corpus-driven research is . . . controversial.”).
\item \textsuperscript{87} The “dog” example in this paragraph is taken from the evidence in Biber 2009, \textit{supra note **}, at 287.
\item \textsuperscript{88} Biber 2009, \textit{supra note **}, at 287 (discussing Mutual Information (MI) scores).
\item \textsuperscript{89} Ellis & Ferreira-Junior (2009: 198); Gries 2013, 146 (“[B]idirectional/symmetric association measures conflate two probabilities that are in fact very different: \(p(\text{word}_1 | \text{word}_2)\) is not the same as \(p(\text{word}_2, \text{word}_1)\), just compare \(p\) (of \textit{in spite}) to \(p\) (\textit{in spite of}).”).
\item \textsuperscript{90} Gries 2013 at 146.
\end{itemize}
tokens do not always go along with mentions of their type: one does not typically have to specify, for instance, that a car is a vehicle.91

Things get even more complicated if we want to look beyond two-word collocations, for instance by asking about the way that multi-word phrases function.92 The ordinary usage of “carry a firearm,” for instance, may turn less on the typical usage of the individual word “carry” and more on the typical usage of the entire phrase. “Carry a firearm” may function as a “lexical bundle”—a group of words that gets deployed as a single unit rather than as individual words whose meanings are added together.93 Take the slang phrase “packing heat,” for instance, colloquially used to indicate having a firearm at the ready. Although it may be useful to ask how often “pack” and “heat” co-occur with words indicating firearms and carrying on one’s person, it would be important to recognize that this phrase has become an idiom. The presence of the words together changes their meaning from what they might usually be to the meaning of the idiom as a phrase.

Moreover, meaning does not just arise from words that are co-present; it also depends on words that are absent. In linguistics, the distinction is expressed in the terms syntagm and paradigm. Syntagm describes the co-occurring chain of words that make up an utterance. In an utterance like “I like ice cream,” the words I, like, and ice cream are in a syntagmatic relationship to one another: they follow one another and their ordering (as well as various other grammatical functions) give meaning to the sentence. Paradigm, in contrast, refers to the set of words that could have appeared in place of each of those words but didn’t. “I” could have been replaced with “he,” “she,” “they,” “Maryanne” and so on—though not, typically, by “you,” even though that’s also a pronoun. “Like” could have been “dislike,” “love,” maybe even “make” or “steal.” And so on. The unused options are in a sense spectrally present in the words chosen. My addressee understands that I enjoy ice cream, but I’m not necessarily crazy about it, in part because my addressee knows that “love” was an unused option in the paradigm set of “like.” Corpus linguists has ways of approaching paradigms, for instance by mapping out collocations of collocations,94 or finding and coding large numbers of similarly structured utterances with different meanings. But simple collocational frequency of the kind employed in legal corpus analysis does not capture paradigm set choices.

In short, “[f]requency of occurrence, in the sense of pure repetition frequency, explains only a modest proportion of lexical variability.”95 But for the most part, corpus inquiries for legal interpretation have not attempted to address these possibilities or utilize tools beyond the relatively simple frequency, collocation, and keyword in context functions that are available as pre-programmed options on publicly available corpus analysis tools. This approach does not solve the problem of choosing which notion of “ordinary” legal interpreters should use. Nor can it. Deciding what kind of ordinariness should count for legal interpretation is, after all, is a fundamentally normative question. It asks what kind of language use, and what qualities of language use, are most important for determining the meaning of the law.

91 Gries 2013 at 159.
92 Biber 2009, supra note **, at **.
93 Biber 2009, supra note **, at **; Anya Bernstein, Before Interpretation.
94 Vaclav Brezina, Tony McEnery and Stephen Wattam, Collocations in context: A new perspective on collocation networks, 20 INTERNAT’L J. CORPUS LINGUISTICS, 139, 142 (2015) (arguing that “collocates should not be considered in isolation but rather as part of larger collocation networks” and introducing software that can display such networks graphically).
95 R.H. Baayen, Demystifying the word frequency effect: A discriminative learning perspective, 5 MENTAL LEXICON 436, 456 (quoted in Gries 2013 at 161).
Frequency searches on pre-made corpuses and corpus analysis software, in contrast, allow users to plug in search terms and use default query options without consciously endorsing, or even recognizing, the decisions that go into constructing the corpus and default options. Those decisions are instead invisibly outsourced to the programmers who construct the query functions and, sometimes, the corpus itself. The most basic aspect of reality that legal corpus linguistics’ selective empiricism leaves out, then, is the reality of corpus linguistic research itself.

III. THE STATUS OF STATUTES

Legal corpus linguistics aims to provide an empirical solution to the problem of ordinary meaning. But an empirically based inquiry must also consider whether its methodology fits its goal. That is, empirically grounded approach to statutory interpretation must consider whether, and why, ordinary language use provides a good basis on which to evaluate statutory language use. There may normative reasons to link statutory utterances to the everyday speech of that portion of the population they govern which does not work in the government. But there are also good reasons to doubt the validity of this linkage.

A. How is This Language Different from All Other Language?

A corpus inquiry into how a statutory term is used and understood by ordinary speakers must, therefore, consider what kinds of linguistic practices best represent ordinary language for the purposes of legal interpretation. Both the COCA and the NOW corpora skew toward published English. But written English produced with editorial assistance is reliably different from even standard spoken English, not to mention the varieties of non-standard American English. We would need to decide which was more relevant.

If we are looking for the way ordinary speakers would use language in a statute, it is not at all obvious that published, edited texts would be the most illuminating choice. Something like the Santa Barbara Corpus of Spoken American English (SBCSA), which records naturally occurring interactions, may get us closer to how American English is used in its everyday settings. The SBCSA is a particularly valuable general corpus because it includes crucial information about the contexts in which the speech it records occurs—kitchen table discussions among intimates, classroom instruction, forest walks, and so on. Such contexts play a crucial role in linguistic “pragmatics,” that is, the effects of social situations on communication. Pragmatic contexts form a crucial part of linguistic meaning. Indeed, many functional linguistics maintain that semantic meaning—the context-independent aspect of meaning that a term carries with it from utterance to utterance—is merely a subset of linguistic communication, which is largely controlled by pragmatic function.

But this corpus is necessarily smaller than corpora created by web crawling software. The much larger COCA does have some spoken data, but it is drawn from television and radio shows. Although the COCA documentation itself claims that such speech is representative of ordinary

---

96 The COCA is available at https://corpus.byu.edu/coca/; the NOW Corpus is available at https://corpus.byu.edu/now/.
97 Biber 2009
98 The SBCSA is available at http://www.linguistics.ucsb.edu/research/santa-barbara-corpus.
99 Michael Silverstein, Cognitive Implications of a Referential Hierarchy, in SOCIAL AND FUNCTIONAL APPROACHES TO LANGUAGE AND THOUGHT 125, 129-30 (Maya Hickman ed., 1987) (arguing that the part of language use in which speakers refer to and make assertions about things in the world through semantically constant meanings “is a special case” of language use, located within a broader category of language as “a form of social action, a meaning-dependent and meaning-generating activity” whose significance rests on its pragmatic context of use).
language because it is “unscripted,” television and radio show appearances differ in obvious ways from everyday conversations. They record interactions between professional media hosts and their guests, invited by show staffs to discuss particular topics for particular purposes, like providing entertainment, information, or opinion. Participants are carefully chosen for particular broadcast-related reasons, and generally include at least one media professional. The interactions happen in an extremely limited time frame—often just a few minutes. They tend to focus on specific, limited, predetermined topics; media professionals are there to keep them from deviating from those topics. The participants generally have highly constrained frames of mutual reference, with little shared personal experience, and little incentive to draw on any shared experience they do have. Although these interactions present media professionals talking to guests, both perform for the broader audience of the listening public. In sum, the oral parts of the COCA may be unscripted like naturally occurring conversations are. But because the pragmatics of their language use differ so significantly from that of everyday conversations, we can expect them to constitute a different genre of language use as well.

It may be that such speech is indeed relevant to interpreting legal text. The COCA records interactions that are nationally broadcast, and therefore available to be watched or listened to by large portions of the population addressed by American statutes. Radio and television programs often combine a host personality, who tends to speak a fairly standard, formal English, with guests who may be of different classes, educational backgrounds, regional origins, and ethnic identities. They may present a fair representation of a range of kinds of speech produced for wide public consumption. So, although radio and television interactions form a different genre of spoken language than everyday conversations, they may nonetheless provide relevant empirical evidence. But to use this data responsibly, corpus analysts would need to articulate the reasons they find compelling. They would need to consider whether and why the material in the corpus they choose provides evidence of the genre of language use that is most relevant to interpreting statutes. Using a corpus just because it is available, or large, does not suffice to render it legitimate or rational.

Some analysts may prefer to avoid choosing one genre over another. The COCA provides a corpus with both written and spoken language uses (though with the limitations discussed above). One might think that such a mixed corpus would give the best cross-section of American English—a broad-based representation of that ordinary language that legal interpreters seek. Yet there are dangers to mixing genres. For instance, we know that academic and conversational American English each have lexical bundles; that each genre is characterized by different lexical bundles; and that those lexical bundles also tend to have different functions in each genre.

We may not care about this genre-based distinction; we may just be interested in how lexical bundles function in English generally. But there might not be a uniform pattern of lexical bundle function across all genres of English usage. If that is the case, our search for a consistent through-line across pragmatic contexts may yield something else: an average of uses across contexts that conflates different patterns or trends rather than representing any particular one. The frequency with which particular lexical bundles appear, and the functions they tend to serve, may tell us less about how lexical bundles work in English generally, and more about the composition of that particular corpus. That is, the frequency and function of particular lexical bundles may depend on whether the corpus has more academic text, or more conversational transcripts. Although having lots of data can

---

100 COCA webpage
101 See supra Part II.
be useful, the conflation of different kinds of data, evidencing different kinds of phenomena, can also muddy an analysis.

On the other hand, we may decide to eschew everyday conversational language as well as broadcast interactions and just focus on published, edited text. “Since we are interpreting a written text,” some argue, “evaluating that text through the lens of standard written American English . . . may be the right approach.”\(^{102}\) On this view, what constitutes “ordinary” language for legal interpreters should be language that is similar to statutes in style and pragmatics of production. And the formal, edited, published English of popularly available publications certainly comes closer to the language of statutes than everyday conversational English does.

How close does it come, though? Despite the persistent appeal to ordinary language as a model on which to interpret statutory language, statutory language differs from non-statutory in some fairly obvious ways. Start with the parts that anyone reading a statute notices first: it’s really difficult to follow. Statutory phraseology is convoluted, with its overstuffed sentences, comically long qualifiers, and cross-reference mazes. Statutes utterly fail to conform to Preferred Argument Structure, packing new information into every available clause as though Congress were running out of paper. They don’t indicate the weight, credence, or value audiences should grant particular assertions—the sort of thing that speakers might indicate through modal function words like the complementizer “that.”

It is also safe to say that way statutes are produced, as well as the effects they have, are unique. No ordinary utterance is created quite like the byzantine, multi-player exquisite corpse creations that form our law. And nothing but law imposes quite the same constraints, or embroils us in quite the same arcane system of regulation and litigation. Production context and real-world effects—what linguists call “pragmatics”—are, however, two key aspects of linguistic communication, structuring both the meaning of linguistic utterances and their broader significance.\(^{103}\)

The syntax and pragmatics of statutory language is, in short, weird. Legal corpus linguistics proponents seem unconcerned with the syntactic and pragmatic distance between statutes and the language data contained in corpora like COCA and NOW. Perhaps they assume that semantics—the stable part of word meaning—will save the day. This is not a failsafe assumption, however. For one thing, statutes often use words in very odd ways. A statute might use an existing word for a new object (such as “Exchange” for health insurance marketplace).\(^{104}\) Or it might use a seemingly ordinary word in a way that make sense only given other parts of the statute itself (as when a “credit” works to offset a tax).\(^{105}\) Or it might take a historically specialized word and define it to mean something broader and more ordinary (as when a statute defines an animal “taking” not just as common-law hunting or trapping but as anything that “harms” animals).\(^{106}\)

More importantly, though, if we are going to be realistic about the distribution of terms in a corpus, it seems odd to be so unrealistic by over-relying on the role of semantics in linguistic communication. Semantics is important, but it is only a subset of communicative function; indeed,

\(^{102}\) Lee and Mouritsen, supra note 3 at 834.
\(^{103}\) Silverstein.
\(^{104}\) King v. Burwell, Affordable Care Act.
\(^{105}\) Shell Petroleum, Inc. v. United States, 182 F.3d 212, 214 (3d Cir. 1999); Internal Revenue Code.
semantic meaning often depends on syntactic and pragmatic contexts. Take a sentence like “[t]here shall be allowed as a credit against the tax imposed by [other sections of] this chapter” a certain amount of money for certain kinds of fuel.107 In the context of a statute, it issues a command from the legislature to an administrative agency. But in most contexts, speakers do not have the option of issuing commands to others.108 In a less unusual communicative context than a statute, the sentence may instead be a prediction. A prediction is a quite different thing from a command. For example, it is falsifiable: it can be proved wrong by failing to come to pass. A command may be contradicted, or it may be invalid, but it cannot be proved wrong by facts; commands are not falsifiable. The words stay the same; but they mean something different in a different context.

Statutory language, then, differs in many important ways from the non-statutory language that the most popular contemporary corpora collect. Looking for terminological frequencies, or even expressive patterns, in one area may thus have little to teach us about the other. The empirical thrust of corpus linguistics suggests not only collecting instances of language use, but considering how well they serve as evidence for the particular inquiry in which they’re used. In this case, in contrast to a widespread normative insistence on ordinary language as the proper basis for statutory interpretation, there are reasons to doubt that everyday language usage is sufficiently similar to actual statutory usage to provide a viable evidentiary foundation. We may wish it did, and our theories of democratic communication may insist it does; but the social realities of both statutes and language use should cast doubt on those ideological commitments. Whether we respond by rejecting the ideological commitments, changing the evidence we seek, or trying to ameliorate the disjunction with other theories is an open question; but the empirical impulse of something like corpus linguistics should at least make us recognize that the disjunction is there.

B. Statute as Utterance versus Statute as Object

Despite the close ideological connection between ordinary language and statutory language, people predictably approach statutes differently than they approach everyday conversations, newspapers, or TV shows. This results not only from the pragmatics of statutory production discussed in the preceding section, but from the role statutes play in our lives. Law is more or less sui generis, so its social effects do not much resemble those of everyday conversations, newspapers, and so on—just see the “shall” in Shell Oil.

If we are interested in how statutes’ audiences understand them—if we want to know not just how an ordinary person might use a word, but what they might think it means when they encounter it—this poses a problem. How audiences understand statutory language may be quite different from how they themselves speak. I do not generally use the word “exchange” to mean a private insurance marketplace established under the Affordable Care Act. And yet I am capable of understanding that, for the purposes of the Affordable Care Act, a private insurance marketplace established under that Act is called an “exchange.”109

---

107 Id. at 215-16 (quoting a provision then codified at 26 U.S.C. § 29; this provision was later amended and recodified at 26 U.S.C. § 45K).
108 Even when a speaker is in a position to issue a command, the verb “shall” is not the idiomatic way to do it in most situations.
109 King v. Burwell, Affordable Care Act.
Put another way, audiences do not orient themselves to words in general, but to words in social context. How an ordinary person would understand a term in a statute is not necessarily the same as how she would use that term on her own. So, for instance, a recent survey study by James Macleod asked how a demographically representative sample of people in the United States understood causation requirements in homicide, hate crime, and discrimination statutes.\footnote{James A. Macleod, \textit{Ordinary Causation: A Study in Experimental Statutory Interpretation}, 94 IND. L. J. __ (2019).} Recognizing the difference between using language and understanding it, Macleod did not ask how survey respondents used words like “causation,” “but-for cause,” or “proximate cause” in their own speech. (Indeed, it’s quite likely that many people simply do not use at least some of these terms with any frequency, if at all.) Rather, Macleod asked what respondents understood different phrasings having to do with causation to mean in the context of particular situations.\footnote{Id.}

This kind of research, generally absent from corpus inquiry, recognizes that people involved in communicative situations orient themselves differently to different kinds of utterances. Mining repositories of conversations, newspapers, or media appearances does not necessarily reveal how people orient themselves to statutes, which are so different both in their linguistic characteristics and in their social roles from all those other kinds of speech. The relevance of an empirical analysis of a generalist corpus to statutory interpretation is, thus, based on some quite fictive assumptions: that statutory language resembles ordinary language, and that people approach the meaning of statutes the same way they approach the meaning of non-statutory language.

\section*{IV. Meaning as Fact versus Meaning as Activity}

Corpus linguistic inquiry can reveal aspects of how people use language. It deals in patterns and tendencies, not in enduring meanings or underlying intentions. It can reveal the ways that certain people using a certain genre in a certain time period tended used a term or a phrase. Such patterns and tendencies, however, will give only limited information about what a statute means in a particular litigating context. We may accept that statutory language should be interpreted according to what it would have meant at the time of enactment. But the nature of litigation in the American system, combined with the nature of time, make some sort of updating inevitable. Statutes end up applying to things that did not exist at the time of enactment; to real-world circumstances that have changed; and within legal contexts that have been altered through precedent and legislation. Legal interpreters have to decide whether and how the statute functions in this new setting. A responsible use of corpus evidence would recognize the inevitable changes that application in new legal and practical circumstances will wreak on legal strictures. But legal corpus proponents sometimes write as though the method could produce a “correct” answer about the “real” meaning of a statute.

In a recent \textit{Yale Law Journal} article advocating legal corpus linguistics, Thomas Lee and Stephen Mouritsen put their view thus: “Our thesis is that words have meaning,” they write, “and that meaning can be theorized and measured using” tools from linguistics. The claim that words have meaning sure seems incontrovertible, and it’s certainly striking.

But it’s not quite accurate. Words don’t “have” meaning the way, say, a substance has a chemical composition: meaning is not an essence that inheres in the word, traveling with it across situations and domains, irrespective of the activities of others around it. Rather, words gain meaning through the activities of those around them. They are one of the tools people use to produce meanings together. From the perspective of disciplines interested in the patterned deployment of
language in society, meaning is not really a thing that inheres in some particular object so much as the outcome of a process. It is not a fact but a productive social activity.

This is a subtle distinction, but an important one. The more legal corpus proponents imply that corpus analysis can give us correct answers to questions about the underlying, enduring meaning of statutory terms, the further they stray from the sort of thing that empirical research can actually deliver, or even aims to achieve. Legal corpus proponents may be bolstered by the notion that their inquiries are “objective” because they are based in neutral evidence external to the litigation debate. But, as Theodore Porter has argued, the idea of objectivity as neutral and external merges easy with a very different idea of objectivity: as unassailably correct.112

Neutral, after all, does not necessarily yield truths; one can examine neutral evidence and still not find a decisive answer to one’s question. Linguistics, and especially the functional linguistics out of which corpus analysis grows, studies the co-production of meaning by communicative participants; it cannot reveal the enduring, static meaning of a statutory provision that is inherently subject to new application and to debate. Once again, the empirical aspect of legal corpus analysis only becomes relevant once we accept the fictive assumption that statutory language has one enduring meaning that can be scientifically uncovered—an assumption that goes against the findings of the very disciplines on which legal corpus linguistics builds.

V. TOWARD A BROADER EMPIRICISM IN STATUTORY INTERPRETATION

Empirical inquiry has a lot to add to the statutory interpretation conversation. But, as I have suggested, it can make that contribution only by recognizing the relevance of empirical inquiry across the board—not by maintaining and bolstering interpretive fictions.

Corpus linguistics, for instance, could do real service by connecting areas of statutory production and implementation, linking different utterances across legislative drafters and their primary audiences.113 If corpus inquiry recognized the distinction between use and understanding, it could be used to specify the particular characteristics that distinguish statutory text as a genre. Such a specification may help legal interpreters determine what kinds of evidence are best suited to elucidating what kinds of statutory terminology. When, for instance, statutory syntax or semantics resembles that of other genres of language use in patterned ways, those similar genres may form appropriate corpora for investigating that statutory language. Where it diverges from that of other genres to give statutes their unique characteristics, it may be that corporate of unrelated and dissimilar language use will be of limited utility.

Additionally, investigating how people understand statutory language, and the specific ways in which it differs from language use in other genres, may help legal corpus proponents address an enduring, but under-appreciated, difficulty in statutory interpretation: it’s quite likely that many speakers would have difficulty giving any meaning to a lot of what is published in the United States Code. Statutory interpretation discussions, and especially those that depend on claims about ordinary language, tend to assume that people with no connection to government and no training in

112 See THEODORE PORTER, TRUST IN NUMBERS: THE PURSUIT OF OBJECTIVITY IN SCIENCE AND PUBLIC LIFE 3 (1995) (noting that objectivity can have different meanings—an “ability to know things as they really are;” a disinterested or neutral judgment; a rule-based, non-discretionary expertise—and that these meanings are often conflated, leading people to claim certainty when in fact all they have is neutral, but not necessarily conclusive, evidence).
113 See supra, Part III.
its language would have clear ideas about what a given statute meant. But empirical evidence points the other way. “Anecdotal evidence suggests that most seasoned statutory players start with the section-by-section [summaries provided in committee reports] to understand the point of each section before turning to what is often the dense and unintelligible . . . minutiae of the statutory text.”114 If even “seasoned statutory players” can’t figure out what a statute means, how are “ordinary speakers” supposed to?

Given the complexity of modern statutes, the normative conviction that statutes should provide notice of their contents to an unrelated public may itself operate as a legal fiction—a legal fiction that seems sometimes to take on the quality of an almost religious counterfactual, a claim repeated insistently in the face of all contrary experience and evidence. Perhaps it is time to address that contradiction, and the attendant democratic qualms it should cause, head on. Using corpus linguistics could, that is, encourage legal interpreters to take a novel interest in how the very people for whom they often express an abstract solicitude actually experience the law.

Findings from corpus linguistics may also help legal thinkers evaluate the canons of construction that play such an important part in contemporary statutory interpretation.115 As I have written elsewhere, there is no consensus on just what it is that these canons are supposed to accomplish, or how they are supposed to evaluated or validated.116 Nonetheless, linguistic canons of construction are supposed to guide judges as they interpret statutory language. Reading up on findings in corpus and other functional linguistics about English language patterns and structures may give judges and scholars a basis on which to evaluate, for instance, how well the canons represent actual English language usage in any genre or social arena. And that may give judges and scholars a basis for considering what canons should accomplish and how well they do so. For instance, the rule of the last antecedent instructs that, “in a list of disparate items, a clause modifies the item nearest to it.”117 But this may present a less accurate vision of how English usually works than linguists’ research on Preferred Argument Structure, which investigates in more depth how it is that speakers structure the flow of information, such as the characteristics of various topics of discussion. If we do want to use standard conversational English, or standard written English, as a model for interpreting statutory English, at least we can use the best available information about how it actually works, rather than searching for word frequencies.

A related benefit might be a compendium of best practices for legislative drafters. Since linguistics researchers have uncovered quite a lot about how American English works in the wild, legal scholars could draw on that research not just to differentiate statutory language as its own peculiar genre, but to clarify how statutory drafters could better approximate the kind of English that people outside the government might be able to understand. Rules of thumb like “introduce only one new piece of information per clause” could help harried Congressional staffers edit their work. Of course, Congressional staffers tend to be pretty fluent speakers of American English themselves, so their failure to write statutes in an easily comprehensible style probably arises from

115 See, e.g., Abbe R. Gluck and Richard A. Posner, Statutory Interpretation On The Bench: A Survey Of Forty-Two Judges On The Federal Courts Of Appeals, 131 HARV. L. REV. 1298, 1302 (2018) (finding, based on interviews with federal judges, that “[t]he younger judges [in the study], most of whom were educated under the modern legislation curriculum, were generally more focused on, and accepting of, the canons of construction” as compared with older judges).
116 Bernstein, Democratizing Interpretation.
something other than lack of competence. Still, scholars propounding clear and discrete guidelines for producing comprehensible statutes might nudge legislative drafters to treat that goal as more important than they have in the past.

Finally, introducing corpus linguistics into the legal conversation could help shift the paradigm of how judges and scholars conceptualize statutory interpretation, helping us develop a more empirically realistic notion of who it’s done by and how it works. Discussions of statutory interpretation, as is well known, tend to focus on the statutory interpretation done by courts. Moreover, the image of interpretation that these discussions tend to draw—the prototypical, backgrounded interpretive situation they assume—pictures at its center a judge as a reader. The reader’s goal is, in a sense, an internal state: a state of understanding the statutory text. The action in this image is the interaction between the text and the reader’s understanding. This is not quite the way statutes work. Statutes are speech acts: they accomplish things in the world. When we interpret them, we may strive to understand what they really say. But in the end, the effect of interpreting is to determine how statutes get implemented.

For another thing, the image of a judge reading in a quiet chambers belies the cacophony that attends statutory interpretation in practice. In reality, there is no one person—not even a three-judge panel—who gets interprets the true meaning of the statute. This is not only because judicial interpretations of statutes only happen within the adversarial context of litigation. It is also because most efficacious statutory interpretation is done not by courts but by agencies. Administrative agencies do more statutory interpretation than courts, and most administrative interpretations are never reviewed by a court. A number of scholars have considered how agency interpretation might differ from that of courts, and whether it should. Empirical investigation into how agencies actually do that work is in its infancy. But what we already know should suffice to disrupt the house-is-quiet image of a reader reaching a true understanding of a text.

For one thing, agency interpretation involves a lot of people. We know this from the basic requirements that structure agency rulemaking, which require agencies to take input from anyone who cares to comment on proposed rules and to explain why those comments did or did not make a

121 Jerry L. Mashaw, Norms, Practices, and the Paradox of Deference: A Preliminary Inquiry into Agency Statutory Interpretation, 57 Admin. L. Rev. 501, 501-02 (2005) (“virtually no one has even asked, much less answered, some simple questions about agency statutory interpretation: As a factual matter, how do agencies interpret statutes?”); see also Abbe R. Gluck and Lisa Schultz Bressman, Statutory Interpretation From The Inside—An Empirical Study Of Congressional Drafting, Delegation, And The Canons: Part I, 65 Stan. L. Rev. 901, 905 (2013) (noting a similar paucity of empirical evidence for legislative drafting practices and presenting groundbreaking interview-derived evidence on that topic). Walker, supra note __, and Shobe, supra note __, are perhaps the only scholars to address the everyday practices of agency statutory interpretation to date. I am currently conducting empirical research on this issue as well.
difference in the final rule. We also know that agencies often distribute policy-making and interpretive work to groups comprised of subject matter experts, legal professionals, economists, and others with relevant skill sets. This is a stark contrast with the calm world in which interpreting is imagined to the cacophony that attends most statutory interpretation in practice. In reality, interpretation usually involves many—perhaps hundreds—of people, working in different ways on different aspects of the issue, and iteratively confronting and resolving problems, disagreements, incoherences, and values choices.

It also suggests that, in the statutory realm, the question of interpretation may be inseparable from the question of implementation. Statutes are not just things that mean; they are thing that do. The idea of “understanding” a statute, then, may be incoherent if it is divorced from the idea of effectuating or implementing it. And that is what judges do, too.

Legal corpus linguistics may have something to offer here as well. It, too, turns away from the centrality of the individual judge and her internal states, looking instead to the multiplicity of voices contained in a corpus. Extending that impulse to voices that are more clearly relevant to the production and implementation of a statute, legal corpus linguistics could help legal thinkers come to terms with the realities that attend statutory interpretation in the administrative state. Such commentators could also incorporate work in linguistics and related disciplines that clarifies how language use serves not just to state truths or falsehoods but also to accomplish things, to cause effects, in the world.

In the end, corpus linguistics’ great gift to the law may not be in providing answers about what statutes mean (or even what people think they mean), but in providing questions: highlighting the assumptions and evasions that characterize a lot of thought on legal interpretation. Choosing a corpus pushes us to recognize the wide range of language users who can be relevant to legal interpretation, and to question why some are included while others are left out. Considering what constitutes relevantly “ordinary” language leads us to question whether ordinary language is really a good benchmark by which to evaluate the meaning of statutes—utterances so utterly extraordinary in syntax, semantics, pragmatics of production, and pragmatic effects.

Recognizing how different statutory language is from most other language might in turn help us face the uncomfortable likelihood that a competent speaker of English who lacks governmental experience or legal training might simply have no idea what a typical statute means. Ordinary language may be of little help in interpreting the byzantine, multi-player exquisite corpse creations that form our law. And at the same time, learning about patterns that structure our speech in ways we don’t recognize might nudge legal interpreters toward a bit more humility regarding their own interpretive prowess. After all, how we think we interpret statutes often does not quite reflect what we actually do.

125 See, e.g., Lee and Mouritsen, supra note 3 at 800 (criticizing the range of meanings judges give the notion of “ordinary meaning”).
How useful corpus linguistics, or other empirical methods, can be to doing legal interpretation depends on how explicit practitioners are about their specific analytic design choices; how well they justify and explain those choices; and how fully they engage in debate about the necessarily wide range of options interpreters face. In contrast, treating corpus linguistics as a route to certainty about legal meaning rejects that kind of methodological responsibility.

By extension, claims to certainty eschew engaging with the underlying issues that both motivate and complicate legal interpretation to begin with. After all, it is not the act of recognizing the indeterminacy of legal language that gives interpreters a “broad license for normative judgments.”126 Rather, normative judgments are an inevitable aspect of legal interpretation, whether interpreters acknowledge it or not. A truly empirical inquiry has to acknowledge both its methodological choices and the inevitable role of normative judgment in making them. In other words, the mantle of empiricism can’t deflect the burdens of normative justification.

126 Id. at 794 (internal quotation marks omitted).