Technologies of Language Meet Ideologies of Law

Anya Bernstein
*University at Buffalo School of Law*

Follow this and additional works at: https://digitalcommons.law.buffalo.edu/journal_articles

Part of the Law and Society Commons, and the Linguistics Commons

Recommended Citation

© 2020 Michigan State University College of Law

This Article is brought to you for free and open access by the Faculty Scholarship at Digital Commons @ University at Buffalo School of Law. It has been accepted for inclusion in Journal Articles by an authorized administrator of Digital Commons @ University at Buffalo School of Law. For more information, please contact lawscholar@buffalo.edu.
A new technology of interpretation is taking the legal world by storm. Legal corpus linguistics, an approach generally unknown in the field until a few years ago, has suddenly become a focus for articles, conferences, legal briefs, and even judicial opinions. Taking advantage of evolving computational approaches and data collection
abilities, legal corpus linguistics searches big data sets of language use to help interpret legal texts.¹

Legal corpus linguistics is technological. It uses vast collections of language use examples, demonstrating a very contemporary interest in big data. It compiles those big data sets through various means, but

many involve either digitizing and collating historical texts, or automating the collection of more contemporary material, for instance by continually searching the Internet for new entries. It uses computer programming techniques to help annotate those texts, allowing each term to be labeled as a part of speech. Other computer programming techniques allow users to search the database, locating individual words and phrases within vast collections of language use examples. Being able to search databases to collect a wealth of instances of people using particular words in various contexts helps researchers understand how those words are usually used. Proponents argue that this knowledge allows legal corpus practitioners to give empirical grounding to claims about ordinary meanings that pervade legal interpretation. This technology, the argument goes, gives us access to what the words that appear in statutes really mean to people, making legal meaning-making less subjective, even “scientific.”

This Article challenges this new prevailing wisdom, puts legal corpus linguistics in the context of other meaning-making technologies, and suggests an approach for analyzing any technology of language in the law. One of my aims is to caution against technological exceptionalism—a view that computerized, automated, or big-data approaches are somehow special, perhaps more trustworthy, less subjective, and most likely to succeed. Rather, I argue that we should ask the same questions and make the same demands of any method of interpretation.

Any technology, after all, is only as useful as the way it is used: Technology depends on technique. So even when using or analyzing a new technology, we should ask some old questions. What are the means this technology employs, what are the ends it claims to reach, and does it convincingly relate the two? What are its premises, what assumptions do they rest on, and how do they structure interactions with the object of analysis? What other options are available and how do they compare?

Technology, in this sense, is not neutral nor passive. It is not just an effect yielded by external causes like scientific development. It is a cause in its own right. Because of this causal power, it is particularly important to examine the underlying assumptions that help construct, and are perpetuated through, a given technology. And this causality is

---

2. Brief for Professor Clark D. Cunningham & Professor Jesse Egbert as Amici Curiae Supporting Neither Party at 5, In re Trump, 958 F.3d 274 (4th Cir. 2020) (No. 18-2486) (calling corpus linguistics “a scientific discipline”) [hereinafter Brief for Cunningham & Egbert]; see also Lawrence M. Solan, Can Corpus Linguistics Help Make Originalism Scientific?, 126 YALE L.J.F. 57, 57 (2016).
true of old-fashioned techniques as much as of newfangled ones. The high-efficiency washing machine and the simple washboard are both technologies of clothes washing and should both be subject to the same interrogation about means–ends relations, built-in assumption, broader effects, and comparative advantages. To elucidate these points, I first draw on theorists who have influenced our understandings of the production of knowledge and technological development.3

I then introduce legal corpus linguistics, describing its origins in academic linguistics and the somewhat different way it has been practiced in legal interpretation.4 Having laid this groundwork, I ask how we should evaluate this emerging technology and its role in legal interpretation. I argue that legal corpus linguistics fails to coherently relate its methods, questions, aims, and claims. It promises more than it can achieve. That mistake, moreover, is not costless. The methodology inscribes a peculiar view of legal meaning: a narrow, asocial, and abstracted notion of things that are in fact broad, social, and practice-based. The illusion of simplicity that legal corpus linguistics propogates undermines our evolving understanding of the real complexities of law and leaves out participants and contexts that are crucial to the production of law as a social force.5

To probe its implications further, I then put legal corpus linguistics in the context of some other ways of giving laws meaning.6 I choose two that sit at the extremes of simplicity and complexity: dictionary definitions on the one hand and administrative rulemaking procedures on the other. These may seem unrelated or incommensurable at first glance. But in fact, they all are technologies of legal interpretation that should be considered in comparison to one another. Comparison also helps illuminate those aspects of legal corpus linguistics that fit it snugly into particular legal ideologies, but blind it to the realities of how law functions in society.

I. TECHNOLOGIES OF LEGAL MEANING

The term technology can indicate a broad range of meanings, most with a vaguely vector-like feel to them. It tends to convey something that rushes forward into the future, evoking a feeling of

3. See infra Part I.
4. See infra Part II.
5. See infra Part III.
6. See infra Part IV.
progress and often depending on some sort of machinery that exceeds the merely human. Even better, a particular kind of machine: computers, especially when they do things without constant human input. If the computer seems to be taking over human activities or replacing human judgment, that computer surely qualifies as technology. Discussions of technology often have a breathless quality, as though there were no time to waste: Who can bother with analytic niceties in the face of onrushing progress? Technological innovations are sometimes treated with a kind of apocalyptic urgency, as though everything were up for grabs, up in the air, falling on our heads as we speak. And that means we need to get on board or be left behind—or mangled under—this unstoppable juggernaut. When people talk about technology, they sometimes sound like they are standing in front of a speeding train that is about to run them over.

Here, I prefer to step off the tracks. While new tools can lead to real changes, any technology is always part of numerous larger systems of practice. It emerges from them, reflects them, and influences them even as it remains influenced by them in turn. So, the notion of technology should have a broader, more pervasive sense than that of machines, computers, or even innovation in general. Especially in the realm of law, it is important to remember that technologies are human creations that work with other human creations. If they help give meaning to law, they do so as part of a longer-term human enterprise of meaning-making.

I take technology not as a term for gadgetry but as a way of describing how things are done, how people and things are brought together and brought under control. This broad notion of technology replaces some ultimately unsatisfying definitional work—just how automated, mathematical, or gadget-y must something be to qualify as technology? Rather, my definition focuses instead on the ways legal meanings emerge from social processes.

Most importantly, keeping our understanding of technology broad allows us to develop the questions we would want to ask for any way of doing things—new or old, mechanized or not. And that, I posit, will help produce realistic evaluations of whatever it is we analyze. It helps analysts avoid being over-awed by the fact that something seems new, or uses unfamiliar software, or produces results at the click of a mouse. Instead, we can see how different approaches relate to and compare with one another along the parameters that matter to us—not the parameters with which technologies advertise themselves.
A. Presupposing, Entailing

My approach to thinking about technology draws on several key lines of thought. Michel Foucault talked about “technologies” as techniques for solidifying knowledge and power relations. Most centrally, Foucault understood technology “to refer not to tools, machines, or the application of science to industrial production, but rather to methods and procedures for governing human beings.” These modes and procedures might inhere in how people organize production or communication, how they shape individual conduct to a norm or subjugate individuals to structural power, or how they transform themselves for their own purposes. For Foucault, each of these is “a matrix of practical reason,” that is, a set of practices enacted for certain purposes on certain understandings. Law, of course, participates in each of the four social areas Foucault highlighted.

On this approach, technology is characterized by practices of control that render things visible and knowable, and therefore available to discipline, control, and further study. It is often embodied in micro-practices—small, hidden habits that invisibly structure our conduct and our relation to the world. In that sense, technology is never neutral. It also does not simply accomplish the end for which it is deployed and then stop. Rather, it has ongoing effects on the situation in which it emerges.

There are a number of ways in which technology has reverberating effects. For one thing, technological forms affect people in ways that exceed whatever localized purpose they are designed to address. A factory work schedule may start out as a way of ensuring that the necessary people are present to make the widgets and that production runs smoothly. But it has wider effects than that. The schedule structures those people’s lives and affects their family relationships (when they see a spouse, who cares for their kids); their physical state (whether they eat when they are hungry, whether they walk in the sunshine); their social lives (whom they can hang out with,

---

9. See id. at 18.
when they can hang out); and even their self-understandings (whether they see being at work early as a sign of virtue, whether they see those with no clock-in time as a different category of person).

A factory owner may or may not mean for the schedule to have those sorts of effects. But what some person means to accomplish with a technology is not really so important. We cannot, after all, control the effects of our actions in the world. The point is that the technology of scheduling has effects on bodies, souls, social worlds, values, and institutions. And those effects far exceed the widget production that the schedule itself directly controls.

There is another, related way that technology is nonneutral and nonpassive. The fact that this technology is deployed, rather than some other, has effects. Technology is part of “how it [is] that certain questions become important” while other questions that could have been pursued are instead devalued or ignored. The things that a technology focuses on, highlights, or makes visible are more easily available for further study, control, and valuation. One does not quite need to think that technological means suggest their own ends, but it is a little like that.

Technologies make certain things relevant to particular processes—for instance, counting, monitoring, evaluating—and those things then often come to seem important, in general, to form objects of interest quite aside from the end to which the technology was originally deployed. Pushing certain subjects from being merely relevant to some particular project to being important in some trans-substantive, overarching way is one way that technologies structure the worlds around them. If technological means do not exactly provide their own ends, perhaps one can instead say that technological means suggest other ends that then drive the development of further technological means.

11. See Hannah Arendt, The Human Condition 237 (2d ed. 1958) (describing action that has effects on the world as both irreversible and unstable: the meanings and implications of action exceed our control in ways we can only ameliorate, but never overcome, through “promising” and “forgiving”).
12. See Sawicki, supra note 10, at 162.
13. These ways of being nonneutral are related to the way that technologies incorporate values, assumptions, and knowledges that surround their creation, often in opaque ways. See generally, e.g., Virginia Eubanks, Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor (2018) (discussing how algorithmic decision-making builds in assumptions about value, propriety, and social stature); Oscar H. Gandy, Jr., The Panoptic Sort: A Political Economy of Personal Information 2–3 (1993) (noting that algorithmic sorting systems for economic decision-making are “based on theoretical models that reflect quite
A technology can thus clear a path for path dependence. To stick with the schedule example, the widespread use of schedules for work might push people to focus on how much time people spend working, and to use the length of time at work as a way to evaluate their performance as workers or even their worth as people. And it may make less visible other questions we could ask—what the quality of the work product is, or how the person contributes to workplace morale, or whether they further the cause of workplace ethics or provide a model of well-rounded living, or many other possible questions that might end up never being asked because the technology has no use for them.¹⁴

In sum, it is sometimes tempting to think of technology as something mechanically complex but philosophically simple: a tool someone uses to achieve a particular, limited end. But technology is neither neutral nor bounded in that way. Technologies affect both people’s conduct and their worldviews: how they act; what they find worthwhile; and what they do next. In this way, a technology’s effects reverberate beyond its immediate purposes or context of use and into the ongoing development of practices and ideologies.¹⁵

B. Maintained by Multiple Participants

Analyzing a technology thus involves more than looking at how it achieves the end to which it is put. It also requires asking how it

transitory fads or trends in social, economic, and political thought” but continue sorting on that basis even as those trends pass).

¹⁴. This echoes the way that identifying a particular area of life as a “system” can affect the values that are then applied to it. See Bernard E. Harcourt, The Systems Fallacy: A Genealogy and Critique of Public Policy and Cost-Benefit Analysis, 47 J. LEGAL STUD. 419, 421 (2018) (“The analysis of a metaphorical or figurative social system, in contrast to a real or tangible system such as an engine or a military weapon, is inevitably going to involve choices . . . that are . . . normative and political in nature.”).

¹⁵. I do not mean the term ideology to deride or devalue; rather, I use it to indicate broadly the ideas that give phenomena meaning. As Ian Hacking, following G. E. M. Anscombe, has noted, “some of the things that we ourselves do are intimately connected to our descriptions” of them, and such “actions under a description” are not understandable without taking into account how they are described—what ideas are applied to them. Ian Hacking, Making Up People, in THE SCIENCE STUDIES READER 166 (Mario Biagioli ed., 1999) (citing G. E. M. ANSCOMBE, INTENTION (1957)). In this sense, ideology “is characteristic of any sociocultural phenomenon.” Michael Silverstein, The Uses and Utility of Ideology: Some Reflections, 2 PRAGMATICS 311, 313 (1992). “[I]t must inhere in what makes any social entity . . . cohere as that social entity . . . [;]” in fact, “there is no such thing as a social fact without its ideological aspect . . . .” Id.
affects broader contexts outside of those ends, and even how it structures the construction of further goals and practices. Here I find it helpful to draw on Bruno Latour, whose work has brought attention to the multitude of participants that must be mobilized in order to deploy a technology, or produce a truth, or reach a legal conclusion. Participants, moreover, are not limited to the human beings who implement the technology, or even those who feel its effects. For Latour, many of the most important participants to any technological process are not human beings at all. They may be animals, or germs, or algorithms. All those things have effects in the world, and the power and endurance of any technology depends on such nonhuman actors.

Latour’s basic axiom is that it takes way more than a village to produce an effect, even if we often think of it as a “social” effect. It requires a kind of mass mobilization. For instance, in the area of disease, our understanding of germs forms part of their being in the world. But we did not get that understanding all by ourselves. If you just look at the decision-makers, or even just at the human beings, you will miss a lot. Think of the role that bacteria play in our lives now. They are very powerful: they make us wash our hands, heat our milk, take our anti- and pro-biotics. Reaching this level of efficacy took a lot of different participants in many processes. It took, of course, a genius like Louis Pasteur. But it also required bacteria themselves to participate in his experiments. Making that participation possible required the experimental apparatus that allowed those experiments to bear results. The laboratory had to be developed as a space particularly organized for doing and controlling experiments. Other people had to recognize that experimental results could have broader meanings beyond the confines of the laboratory. And so on. Each of these steps and participants was important—perhaps necessary—to producing our current bacterial regime.

Latour challenges us to not know the relative importances of entities ahead of time. Instead, he wants to follow the links connecting


17. See Bruno Latour, The Pasteurization of France 40 (Alan Sheridan & John Law trans., Harvard Univ. Press 1988) (1984) (“[D]istinctions among types of actors matter less than the fact that they are all renegotiating what the world is made up of, who is acting in it, who matters, and who wants what. They are all creating . . . new sources of power and new sources of legitimacy.”).
entities to one another. After all, it’s not like we just up and decided one day to start washing our hands and boiling our milk. The germs, through our ongoing interactions with them, restructured our world: bacteria helped make us the people we are. Although human beings came to understand bacteria through Pasteur’s work, understanding was hardly the only outcome of those interactions. Just as we constructed our idea of bacteria, the bacteria have reciprocally helped construct us. If we just look to Pasteur, we will miss most of what goes into making germ theory viable and making it stick.

In fact, Latour insists that making things stick—producing truths and practices that endure—is a terribly difficult process. It is not enough that something just happens to correspond to some reality or have some effect. Producing a truth or a technology that is recognized as such and treated as such, requires a lot of care and concern and effort and, especially, mobilization of participants. When Latour broadens the scope of analysis to take into account the important contributions of human and nonhuman actors who help make truths and technologies stick, he insists that there is no absolute way to rank the relative importance of these different participants. Some may look more important from one angle, less from another. Without a predetermined order of importance, the truths and technologies we are accustomed to can look less secure; there is less we can take for granted. Latour aims to make readers aware of the dense networks connecting the construction of ourselves to the construction of facts that matter to us—the theory of evolution, maybe, or the effects of pollution. These matters of concern require constant nurturing to maintain their status as things that have effects; otherwise, they might be deconstructed out of existence. And because of the multidirectional co-construction of entities, part of us would go with them.

Latour’s premise and exuberant research agenda have a number of fascinating implications for both scholarship and political action. For my limited purposes here, Latour’s approach helps highlight that any technology is a participant in numerous social (as well as nonhuman) processes, regardless of what its inventors or users intend. It reminds me to keep my horizons broad when I ask what gets mobilized in the deployment of any technology and what effects it has. It asks me to focus not just on decision-makers or even those affected, but also on the connections that put things in relation to one another.

It also reminds me that the intentions of the people involved are, not exactly irrelevant, but not special: an intention is just one more participant that can get involved in the deployment and effects of a technology. Latour emphasizes this with the funny term *actant*, which he uses instead of “actor” to underscore the way that intentionality, humanness, even animateness is not necessary for participants in networks to have effects on the world. An intention may be interesting, but whether it is important will depend on the particular situation I analyze and the angle I view it from. Taking a tip from Foucault and Latour, when we find technologies making protestations of innocence—I’m just a simple means to someone else’s end!—we should ask what participants help articulate that position and what larger effects it might have.

C. Representing, Intervening

Finally, the work of Ian Hacking can help structure how we think about technology. Hacking is primarily interested in how it is that we know things and how we think about knowing. He suggests that, historically, people have thought about science in two broadly contrasting ways, which is nicely captured in the title of his book on this topic: *Representing and Intervening.* The “representing” view sees knowledge as a description: We know that we know something when we can explain it or discuss it. The “intervening” view, in contrast, places the proof in the pudding: In this view, knowledge is not so much the ability to describe as to affect. We know that we know

---

19. See Latour, supra note 17, at 201.
21. See generally Ian Hacking, *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science* (1983). Latour and Hacking have different ideas about how we should treat nonhuman actors. Basically, Hacking thinks human beings are more special than Latour does. See generally Hacking, supra note 15 (“What happens to tuberculosis bacilli depends on whether or not we poison them with BCG vaccine, but it does not depend upon how we describe them. . . . Human action is more closely linked to human description than bacterial action is. . . . The microbes’ possibilities are delimited by nature, not by words. What is curious about human action is that by and large what I am deliberately doing depends on the possibilities of description. . . . Hence if new modes of description come into being, new possibilities for action come into being in consequence.”). This disagreement, while important, does not impinge on the way I draw on both scholars here.
something when we can make it happen, or change it, or stop it from happening.

As Hacking explains, scientific study has both a “describe it” and an “affect it” side. But so do many other things. Linguists and anthropologists, for instance, have shown how language also has this quality. Traditional views of language treat it as a transparent tool for identifying and describing things that already exist in the world. In fact, though, these are just the most easily noticed things that language does. Language also profoundly alters the world it sounds in. Linguistic acts create and reconfigure social situations in ways that may be less obvious to conscious notice, but no less important for social life.

Relatedly, popular images of law present it as primarily intervening in, rather than describing, the world. The classic formulation, after all, presents law as a command from a sovereign to a subject. Yet behind any constraint or authorization that law might give lies an implicit understanding of the world that law acts on. The law can really only intervene insofar as it represents, even if that representation remains implicit. Knowing and doing, Hacking’s typology reminds us, are not really separable—they are two sides of one coin. Whatever technology we have for understanding legal language, then, will include implicit representations of law, language, and the world they inhabit.

D. Conclusion: Evaluating Technologies of Legal Meaning

I do not mean to say that technology “really is” one thing and not another. Technology is a large concept that can encompass many


24. See Brian Bix, John Austin and Constructing Theories of Law, in The Legacy of John Austin’s Jurisprudence 1, 1 (Michael Freeman & Patricia Mindus eds., 2013) (describing Austin’s view of the law as “essentially the command of a sovereign to its subjects”).
things; there probably is no one thing that it “really is.” But I do think that there are ways we ought to approach whatever it is, ways we ought to view it and evaluate it. In other words, I am not making a definitional claim but a methodological one. My contention is that the things we generally call technology—machines, computers, and so on—have something fundamental, something important, in common with “ways of doing things,” ways of controlling action and having social effects, more generally. This commonality implies that when we interrogate machine-based or computer-based ways of doing things, we should not assume that those ways of doing things are somehow qualitatively different from other ways of doing things. We should develop questions that allow us to capture both high-tech and low-tech ways of doing things, and to compare them to one another.

In that sense, my claim is also a somewhat normative one: Rather than privileging mechanized or computerized ways of doing things, we should see them as belonging to the same category as other ways of doing things, and as subject to evaluation by the same standards. As a side note, I think this refusal to treat mechanistic technologies as special can also help us avoid some common pathologies, like the tendency to think that some result is better than no result, or that being able to get a result quickly through a mechanized or highly mediated process (like a mouse click) makes it valuable.

The orientations discussed in this Part yield central questions to ask of any technology. How does the technology represent the things it acts on—that is, what presuppositions about the world does the technology build in? What participants or factors contribute to making it work, and what claims make sense of it? What paths does it lay for future endeavors? And, of course, what alternatives does it leave out?

II. LEGAL CORPUS LINGUISTICS

In this Article, I want to apply this evaluative approach to the meaning-making technology of legal corpus linguistics. Legal corpus linguistics draws on approaches developed in the academic discipline of linguistics but tends to abandon some of their crucial aspects. I start at the source to describe corpus linguistics in the discipline of linguistics and then explain how that approach has changed in legal interpretation.

25. See Bernstein, supra note 1 (manuscript at 14–16) (discussing the relationship between academic corpus linguistics and its slightly different legal corpus cousin).
A. Academic Corpus Linguistics

Academic corpus linguistics uncovers widely shared speech patterns. Researchers compile data sets (corpora) of language use. These data sets are often specified by particular genres, registers, speech participants, or other parameters, and populated with examples of utterances produced in the natural course of things—whatever the natural course of things is for that genre, register, speech participant, and so on. So, a study of newspaper language would collect articles to use as examples, a study of academic writing would collect academic works, and research on ordinary conversation would draw on recordings of spontaneous interactions. This approach contrasts with some ways of getting linguistic data that some other methods of linguistics have used, like asking speakers to evaluate the grammaticality of a formulation, or asking them how they would say certain things, or otherwise eliciting speakers’ own thoughts about or reactions to language use, as opposed to collecting examples of language use itself.

Rather than eliciting utterances or opinions for the purposes of research, then, corpus linguistics generally goes out to find utterances

---

26. There is some give in this requirement. For instance, researchers have asked subjects to narrate the events of a short movie in order to collect numerous narrative texts oriented around the same subject. See generally WALLACE L. CHAFE ED., THE PEAR STORIES: COGNITIVE, CULTURAL, AND LINGUISTIC ASPECTS OF NARRATIVE PRODUCTION (1980). These narrations are produced for the corpus. But the researchers did not ask people to, for instance, evaluate the grammaticality of a particular phrasing or produce a grammatical formulation of some content, as non-corpus elicitation studies might. Rather, they provided a topic and then let speakers narrate it freely in order to compile a corpus of different speakers’ narrations of the same focal object, which could be used to find commonalities and differences across social groups, linguistic backgrounds, and even languages, in order to draw comparative conclusions about language use. Id.

27. See, e.g., CARSON T. SCHUTZE, THE EMPIRICAL BASE OF LINGUISTICS: GRAMMATICALITY JUDGMENTS AND LINGUISTIC METHODOLOGY xvi (2016) (“Throughout much of the history of linguistics, judgments of the grammaticality/acceptability of sentences (and other linguistic intuitions) have been the major source of evidence in constructing grammars.”). This kind of elicitation data can provide interesting insights, but it also has limitations that the corpus approach aims to overcome. For instance, people tend to be better at noticing some linguistic phenomena than others, so asking them for their ideas about language might reveal how people think about language, without revealing how they actually speak in practice. See generally Michael Silverstein, The Limits of Awareness, in LINGUISTIC ANTHROPOLOGY: A READER 382 (Alessandro Duranti ed., 2001) (discussing aspects of language that fall outside speakers’ awareness and are therefore not susceptible to speaker judgments or descriptions).
in their natural habitats. Moreover, corpus linguists usually try to collect a broad range of sources for analysis, with the aim of uncovering patterns and practices that are fairly general and widely distributed within their target genre, register, population, and so on. And corpus linguists often use computational tools of varying complexity to undertake their investigations. The idea is to reveal regularities or connections that are invisible to the naked eye and to help linguists recognize patterns and practices that characterize our speech but are not noticeable in the ordinary course of things.

A few examples can give readers the flavor of corpus linguistic findings. Linguists have found that in everyday conversations, people generally do not maximize the amount of information they convey at a time. On the contrary, across languages, speakers tend to introduce new characters, concepts, and other objects of focus no more than once per clause. And when speakers do name new things in conversation, they usually introduce them first, and only then describe those things affecting their surroundings. So for instance, in an everyday conversation about shopping in a grocery store, I might say, “this couple came in and they got some ice cream,” separating out the introduction of the couple from the introduction of the ice cream, rather than, “a couple got ice cream” or “a couple who came in got ice cream,” even though all three are grammatical and convey the same information. One might have thought that people would want to be efficient and convey as much information as possible as quickly as they can. In practice, though, that is not how we actually speak.


29. See generally, e.g., Stefan Th. Gries, 50-Something Years of Work on Collocations: What Is or Should Be Next…, 18 INT’L J. CORPUS LINGUISTICS 137 (2013) (discussing algorithms that may help map asymmetric collocation in corpora).

30. I discuss these examples in a bit more detail elsewhere. See generally Bernstein, supra note 1.


32. See Kärkkäinen, supra note 31, at 675. That is, linguists have found that people tend to introduce new objects of focus first as either subjects of intransitive verbs or as objects of transitive verbs, not as subjects of transitive verbs.

33. See, e.g., Du Bois, supra note 31, at 824.
Corpus linguistics can also reveal differences between genres that are not obvious to the casual observer. For instance, both everyday American English conversations and formal American English academic writing use “lexical bundles” or stock phrases to “provide interpretive frames for the developing discourse,” setting the scene for further content. But conversation participants use entire formulaic clauses (involving a noun and a verb), inserting their own individual content before or after a standard word bundle. Academic writers, in contrast, use formulaic noun phrases, and insert individual content right inside a standard phrasal frame. Genre thus influences how people use even formulaic expressions like word bundles.

Other people can influence us, too: In ordinary conversation, speakers often echo one another. We are so sensitive to the language around us that we often unwittingly structure our own speech to match that of our interlocutors. When one person uses a particular linguistic form, another conversation participant is likely to use that same form as well, even when other options are available—and even when the two are speaking different languages.

35. See id. For instance, a speaker might use the full clause what do you think, inserting her own content before or after it: “what do you think about this ice cream” or “I’m not sure about this ice cream, what do you think?” See id. at 299 (“Conversation prefers fixed continuous sequences of words, with a preceding or following variable slot.”).
36. For example, an academic article might use the formulaic noun phrase the ... of the, inserting the specific noun as needed: “the end of the,” or “the case of the,” or “the fact of the,” and so on. See id. at 284.
37. Studies of structural priming recognize that there is often more than one way to convey the same information, so speakers have choices about what syntactic arrangements to use. For instance, I could say “I gave her the ice cream” or “I gave the ice cream to her” to describe the same act. Research has found that speakers are more likely to choose a syntactic formulation they have recently heard in their own subsequent speech. So, if in a naturally occurring conversation you say to me, “I gave my ticket to my brother,” then when it comes to telling you about giving someone ice cream, I am more likely to match your syntactic formulation—“I gave the ice cream to her”—rather than choose the other option—“I gave her the ice cream.” See Stefan Th. Gries & Gerrit Jan Kootstra, Structural Priming Within and Across Languages: A Corpus-Based Perspective, 20 BILINGUALISM: LANGUAGE & COGNITION 235, 235–36 (2017) (describing demonstrations of structural priming, in which “speakers tend to re-use structures they have recently comprehended or produced themselves,” as “robust and widespread”). Structural priming works across languages, too: when conversation participants use two languages that offer some similar syntactic options, “hearing/producing a syntactic structure in one language will increase the probability of producing a related structure in another language.” See id. at 236; see also Melinda
Moreover, a speaker’s choice of one linguistic form can have social implications even when it does not explicitly affect the meaning of the utterance. In standard American English, for instance, a speaker sometimes has a choice about whether to use the word *that.*

“He claims he called you” and “he claims that he called you” are both correct and both mean the same thing. Or so it seems. But linguists have discovered that people tend to use this “complementizer that” in several kinds of situations: when the information that follows it is syntactically complex, when the information content is surprising, and when speakers want to distance themselves from it. Thus, when I say “he claims *that* he called you,” I subtly—probably unconsciously—imply that I lack total commitment to the proposition that he did, in fact, call you. Such implication through language pattern choice is called semantic prosody.

In short, corpus linguistics gives researchers a way to track patterns in various genres of language usage. It illuminates how language users distribute their focus, control information flow, deploy aesthetic choices, create rapport, convey attitudes, and much more. Speakers do these things, moreover, in ways that are neither formalized into rules, nor solidified as grammars, and not easily observable in individual instances of speech. Corpus linguistics,
mobilizing large data sets for qualitative and quantitative analysis, sheds light on these hidden patterns.

B. Legal Corpus Linguistics

In recent years, legal scholars and practitioners have drawn on corpus linguistics for legal interpretation. Corpus inquiries have appeared in articles, amicus briefs, and even judicial opinions. This legal corpus linguistics has diverged from academic corpus linguistics in some important ways. For one thing, legal corpus linguistics has generally taken a narrower focus than many academic corpus linguistic analyses. Rather than tracking information distribution within discourse or syntactic patterning across genres, legal corpus linguistics tends to focus on specific words, usually just one or a few words at a time. It usually asks how often a word appears in a corpus, or how often it appears alongside some other isolated word.

For an easily recognizable example, consider Muscarello v. United States. In these consolidated cases, one person had a gun locked in the glove compartment of his car while he engaged in a drug deal; another had a gun in his trunk. Did these people “carr[y] a firearm” during a drug transaction, thus triggering a statute’s enhanced punishment? To see whether ordinary people would describe a person who has a gun in the locked glove compartment or trunk of his car as “carr[ying] a firearm,” we could look to a large corpus of American English to see whether the word “carry,” or the phrase “carry a firearm,” often co-occurs with words having to do with cars. If, in contrast, “carry” or “carry a firearm” usually appears with words having to do with individual humans, then that might 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.”

44. See generally Bernstein, supra note 1.
46. Id. at 126–27.
47. See Mouritsen, supra note 1, at 1926. Muscarello is a favorite for legal corpus linguistic discussion because Justice Breyer, who wrote the majority opinion,
Legal corpus studies also often use a “key word in context” (KWIC) search, which examines the immediate context in which target terms appear in the corpus. Thus, for Muscarello, one could look at the paragraph level to see how phrases like “carry a firearm” are deployed to give a larger chunk of text meaning in the corpus. If, for instance, we find many instances where having a gun in a glove compartment or a trunk is contrasted with having a gun at the ready on one’s person, we can hypothesize that the glove compartment and trunk situations might not ordinarily constitute “carrying a firearm”—at least for the kinds of utterance situations captured in the corpus.

While many academic linguists construct their own corpus or use specific corpora particularly suitable for their inquiry, legal corpus analysts have generally used publicly available corpora compiled by others. Particular favorites have been corpora from the legal corpus linguistics project at Brigham Young University, whose law school has been a leading force in promoting the method: 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). The COCA collects American materials from “spoken, fiction, popular magazines, newspapers, academic texts,” and, since March 2020, also “TV and Movies subtitles, blogs, and other web pages.” The NOW corpus collects material “from web-based newspapers and magazines from 2010 to the present time,” continuously crawling the English-language Internet—essentially covering diverse descendants of the

---

engaged in something that looked a bit like a low-tech corpus inquiry. See Muscarello, 524 U.S. at 126, 129 (noting that the Court had “surveyed modern press usage . . . by searching computerized newspaper data bases” and selecting from “thousands of sentences in which ‘the words “carry,” “vehicle,” and “weapon” (or variations thereof) all appear”). As scholars have noted, this way of conducting the search prejudices the question of whether “carry” generally appears with “vehicle” by setting that co-appearance as a search parameter. See Mouritsen, supra note 1, at 1947 (calling the Muscarello majority’s inquiry “fatally flawed”); Anya Bernstein, Democratizing Interpretation, 60 WM. & MARY L. REV. 435, 450 (2018) (“This is a bit like asking whether most deaths are caused by terrorist attacks, then counting terrorist attacks rather than deaths. Since ‘carry’ appears in more utterances than ‘carry’ and ‘vehicle’ together do, searching only within the combination cannot tell us how ‘carry’ itself is ordinarily used.”).


49. See id.
British Empire—for new material.\textsuperscript{50} The COHA contains American texts from fiction and nonfiction books, magazines, and newspapers from the 1810s through the 2000s.\textsuperscript{51} And the COFEA has sources “starting with the reign of King George III, and ending with the death of George Washington (1760-1799),” including “documents from ordinary people of the day, the Founders, and legal sources, including letters, diaries, newspapers, non-fiction books, fiction, sermons, speeches, debates, legal cases, and other legal materials,” including “the U.S. Statutes-at-Large from the first five Congresses.”\textsuperscript{52} The documentation for each corpus boasts about the sheer number of words, often numbering in the billions, that they collect. But it does not always fully explore the methodological effects of their word-collection practices.

For instance, the COFEA contains what its documentation calls “documents from ordinary people of the day,” but does not detail who counts as ordinary people for its purposes, nor explain what its collection might reveal about ordinary usage. Given that colonial people were sharply differentiated in their opportunities to produce and preserve documents, those questions matter—especially if one is using these documents to evaluate how contemporaneous speakers used terms that appeared in statutes. To give a stark example, slave-owning White people and enslaved Black people were both “ordinary people of the day.” In fact, there were probably more slaves than slave owners. So if one is concerned with statistical frequency, as legal corpus writers tend to be, it is enslaved speech that should probably be considered the more ordinary. But a historical corpus cannot capture that ordinary speech because the resource and power disparities of our slave society meant that slave-owners were much more likely to produce and preserve documents than slaves were.

Treating these documents as representative of a generically ordinary people thus normalizes, and obscures, the power of the powerful. If the language of ordinary people is used as a guide to

\footnotesize
52. See Corpus of Founding Era American English (COFEA), BYU LAW, https://lawcorpus.byu.edu/ [https://perma.cc/P586-7C2E] (last visited Mar. 15, 2021) (describing each of several corpora developed or under development by the J. Reuben Clark Law School at Brigham Young University).
interpret what laws mean, moreover, this approach gives those with more text-producing and text-preserving ability a higher claim to determining the meanings of laws. Such disparities may be unavoidable—we cannot go back and record the speech of people long dead who lacked the means to preserve their utterances in writing. There may even be arguments for preferring the speech of slave-owners, if for instance we want a corpus that approximates the language of the lawmakers rather than the language of those subject to the laws but left out of their production. And in any event, such limitations certainly do not make the corpus useless. But they do make a phrase like “ordinary people of the day” inadequate to describe what it is the corpus offers those interested in evaluating what legal terms meant to a broad swath of American residents.

In the area of contemporary language, the COCA collects both written and spoken genres, with spoken texts drawn “Transcripts of unscripted conversation from more than 150 different TV and radio programs (examples: All Things Considered (NPR), Newshour (PBS), Good Morning America (ABC), Oprah).”53 The COCA’s documentation claims that these texts are representative of ordinary, naturally occurring conversations because, like ordinary conversation, they are “unscripted.”54 But there are obvious differences between naturally occurring conversations and those on radio or TV, even if they are not scripted. Radio and TV shows are hosted by hosts who invite guests on for particular reasons, to cover a very limited range of topics in a very limited time frame, and with little opportunity to draw on shared personal experience or create frames of mutual reference. Most ordinary conversations are not like that. In real life, it is quite common to sit around dinner tables or water coolers talking about this and that, not displaying any particular expertise, and referring to people we know in common or experiences we have shared. Even a relatively purpose-driven and time-limited interaction like a checkout counter conversation might include lots of nonpurposeful small talk. The pragmatics of these utterances—the circumstances that surround them—differ, and pragmatics are an important determinant of how people speak and what they say. Again, these limitations do not make the COCA’s spoken word sections useless. But it does mean that what they offer to someone interested in how ordinary people speak is

54. See id.
limited in particular ways that a researcher must recognize in order to use the material in a way that makes sense.

C. Using a Corpus

It may be useful to describe in a bit more detail how a corpus search might go. I take as my example here the COCA, a favorite among legal corpus users, and follow one path one might tread to consider the “carry[] a firearm” language in Muscarello. The example here is stylized in the sense that one would expect a person actually doing corpus-based research to pursue more leads and experiment with more alternatives. But it can serve to introduce readers to the basic idea, and the basic layout, in play.

The COCA user interface is pleasingly simple. The right-hand side of the screen contains a link to a discussion of new features, an offer to download the corpus, and a very brief introduction to the corpus with embedded links to further information. The top banner bears a few graphically depicted options: an introduction to the corpus, more details on its contents, a way to compare it to others, an option to download, an option to input a long text for analysis, and an opportunity to log in with an individual account to view corpora excerpts and searches associated with that account. Below these pictures are four simple options: search, frequency, context, download data. The frequency and context options depend on the search one does; download data produces the same screen as the download option on the top banner.

The star of the show is the search option, which displays on the left side of the screen when one accesses the COCA. It includes a search box into which one can type things for which to search the corpus. Above the search box sits an expandable list with several features: list, chart, word, browse, collocates, compare, KWIC. Each of these shows the target search terms in particular ways. “List” shows all occurrences in the corpus, while “chart” displays the frequency of occurrence by genre. The “word” option shows a kind of homepage for individual words, displaying pronunciation, dictionary definition, synonyms, common collocates, topical areas, common phrases in

56. See Corpus of Contemporary American English (COCA), supra note 48. A reader viewing this Article in a format that does not include figures can go to the COCA website and follow the path I lay out here. Note, however, that the COCA is regularly updated with more entries, such that the results returned will vary over time.
57. See id.; infra Figure 1.
which the term appears, and an initial list of the term in sentence context. “Browse” allows a search by sound, part of speech, and other non-lexical targets. “Collocates” returns to lexical targets, displaying the terms that appear alongside the target term, while “compare” allows one to see simultaneously the collocates of two separate terms (and where they overlap). Finally, “KWIC” yields “key words in context,” that is, the words that immediately surround the target term.

Figure 1.

If I wanted to investigate how the phrase “carry a firearm” appears in the corpus, I would type that phrase into the search field on the left side. To account for different grammatical forms of the base term “carry,” like “carrying, carried,” and so on, I could type an asterisk, like so: “carr* a firearm.”

58. See infra Figure 2.
The COCA returns four possibilities, listed in order of frequency of appearance: “carrying a firearm” (37 instances); “carry a firearm” (29); “carried a firearm” (11); and “carries a firearm” (1).\textsuperscript{59}

\textsuperscript{59} See infra Figure 3.
I can then choose one of these results for an expanded view of the instances of that form. For instance, if I click on “carry a firearm,” the screen shows a scrollable table with numbered instances. Columns include entry number; year of appearance (by default in reverse chronological order); the COCA genre category from which the occurrence was drawn (such as spoken, news, academic, fiction, and magazine); and the specific source from which it came (like the title of a TV show, newspaper, book, website, etc.). There are then three columns, labeled A, B, and C, to help researchers order and keep track of the occurrences. Finally, each row displays a long cell that shows the immediate context of the target word, which is highlighted. In this particular search, for instance, the first context cell, drawn from the CBS show *Face the Nation*, says: “crazy notion of arming teachers, teachers are going to quit before they carry a firearm in class. I—”; while the second, from the *Minneapolis Star Tribune*, says: “suspects before they are charged. # The man has a permit to carry a firearm and used his 9-millim.”

---

60. See infra Figure 4. The # sign indicates a paragraph break; text excerpts in this cell commonly cut off in the middle of a word, since the cell is a standard size.
I can further home in on any of these occurrences for a larger contextual frame. For instance, when I click on number nine in the list, I get a different screen. A table toward the top contains some basic information about where the occurrence of “carry a firearm” was found, again in a simple table form. Here I see the Date (2016), Title (Twisted Innocence), Author (Terri Blackstock), and Source (FIC: Twisted Innocence). Then, in a paragraph marked Expanded context, I see a slightly longer excerpt with the target phrase underlined and highlighted. For this particular entry, the excerpt reads as follows:

girlfriend, just to create trouble in paradise. She and her friends would giggle hysterically at the fight they imagined ensued, but the next day, as she nursed the punishment of a hangover, she would hate herself for it. # Not ready to give up on this fare just yet, Holly honked the horn again. The men up the street turned to look at her. Getting nervous, she reached into the pocket of her door, but of course her pistol wasn’t there. It was against the law for a cabbie to carry a firearm inside the car while they were on

---

61. If I had clicked on a newspaper article, the Title row would show the title of the article, while the Source row would show the name of the newspaper.
duty. It was locked safely in her trunk. # She thought of bucking the law and getting it out, but that would call more attention to her. This was stupid. She was a mother now, and the last place she should have been was in the slums, waiting for someone to blow her head off just for target practice. # But what if the woman who’d called wasn’t a kid at all, but someone stranded here who[.]62

Note that this rather tantalizing excerpt seems to give some support to both sides in Muscarello. In this passage, being “locked safely in [a] trunk” distinguishes Holly’s firearm from one that is “car[ried],” which suggests that “carry a firearm” would not describe a gun in the trunk. But the law referred to in this work of fiction specifies that cabbies may not carry firearms “inside the car,” which suggests that a firearm in a locked glove compartment might still be “carried” if the law failed to specify a location, as the statute at issue in Muscarello did.

This one example nicely illustrates the inherently qualitative assessment that goes into determining how a given term is used in a particular context and hypothesizing about what that might imply. In other words, even using corpus data, a lot of judgment and conjecture goes into any conclusion about ordinary usage. But my point here is not to determine what “carry a firearm” means when used by ordinary

62. See infra Figure 5.
people. Rather, I simply want to provide readers with an on-the-ground understanding of how one might actually search for terms or phrases in the COCA and its sibling corpora.

Having conducted this search, I would need to figure out what the results tell me. I might look through all of the results and come up with ways to classify them that seem to capture some relevant distinctions and similarities within my results, or I might apply a classification system I have already developed through other research, or I might borrow a classification system developed in the voluminous research literature that seeks to uncover how languages tend to categorize things, or perhaps I would simply use my own intuition about how our language cuts up the world as a basis for creating categories into which to group my results. I could also do more searches to see, for instance, how often forms of “carry” appear with words other than firearms, how often versions of “firearm” appear with words other than “carry,” how often one appears with words having to do with vehicles versus persons, and so on.63

Such questions about frequency and co-occurrence, which have typified legal corpus inquiry, could not, of course, yield the kind of insight into hidden language patterns that academic corpus linguistics has revealed. Legal corpus searches do not track how words appear or interact over longer texts or numerous conversational turns, and they generally have not recognized genre distinctions. So this kind of search would not yield the kinds of insights into discursive structure, semantic prosody, or genre-specific usage discussed in Section II.A.

A simple frequency and collocation search also provides a fairly limited amount of information since words can appear or coappear more or less frequently for many different reasons that are not themselves obvious from the search. Words may appear together frequently because the things they describe occur together often, but they might also appear together frequently because the things they describe do not occur together often since nonstandard or unexpected occurrences may be more likely to be remarked on, or marked.64 Words may appear together because they fit into a shared category (tokens of the same type); or because one constitutes a superordinate category for the other (a type with a token); or because they form a

63. See Mouritsen, supra note 1, at 1965–66 (discussing the results of a corpus analysis of the phrase at issue in Muscarello).
64. See Ethan J. Herenstein, The Faulty Frequency Hypothesis: Difficulties in Operationalizing Ordinary Meaning Through Corpus Linguistics, 70 STAN. L. REV. ONLINE 112, 114 (2017); Biber, supra note 34, at 280 (“The role of frequency and quantitative analysis in corpus-driven research is . . . controversial.”).
salient contrast of different categories (two different types, two tokens of different types, or a type with a token of a different type). Or they may appear together infrequently or not at all, yet still possess these kinds of relationships.  

Such sociological vagaries pervade language use. One might even say they constitute language use. But legal corpus studies generally cannot account for them. Indeed, the field has been notable for the narrowness with which its writers have approached their topic. Legal corpus writing rarely refers to the extensive research in linguistics that illuminates the social construction of language use, nor to works addressing how to construct and analyze corpus searches. Thus, legal corpus users often borrow a very small part of corpus linguistics’ method—the part with a search bar—without fitting it into the overarching methodology or the theoretical developments that make sense of it.

Legal corpus writers also tend to be narrow linguistically. They have been almost uniformly uninterested in the kinds of broad patterns that linguists study, such as how people manage information flow, create joint attention on discursive objects, or implicate attitudes in seemingly neutral statements. Instead, legal corpus writers have usually focused on individual words or, at the most, small lexical bundles of three or four words, divorced from bigger discursive contexts such as statutory provisions, sets of provisions in one statute, related provisions across several statutes, Congressional records indicating the range of meanings the words were expected to have by those who made them law, and so on. And within the realm of one-to-four word units, legal corpus writers have been narrowly focused on a very limited range of phenomena, primarily frequency and

65. See Lawrence M. Solan & Tammy Gales, Corpus Linguistics as a Tool in Legal Interpretation, 2017 BYU L. REV. 1311, 1313–16 (2017) (noting that the absence of some token from a corpus does not indicate that it is not in fact a token of some type—an issue the authors call the “blue pitta problem” after the blue pitta, a rare bird whose name does not appear in the COCA, but which is nonetheless a bird).


67. A recent article provides a striking example, laying out a roadmap for legal researchers using corpus linguistics methods. See generally James C. Phillips & Jesse Egbert, Advancing Law and Corpus Linguistics: Importing Principles and Practices from Survey and Content-Analysis Methodologies to Improve Corpus Design and Analysis, 2017 BYU L. REV. 1589 (2017). Phillips and Egbert give a thorough, readable introduction to basic professional responsibility in this area. By outlining the minimum standards that a reliable corpus analysis must fulfill they also show how dramatically most legal work deviates from such standards.
collocation. They have not attended to, for instance, the way words can appear or function differently across different genres or language use situations.

At the same time, legal corpus writing has also been notable for its insistence that the corpus provides an indisputably correct answer about which meaning is really ordinary. This insistence has been particularly amusing when writers have used the same corpus, and sometimes even the same queries, to reach conflicting conclusions about legal meanings. Of course, it is not unusual for researchers in empirical fields to propose different interpretations of similar data, especially when the object of analysis is human behavior, which can indeed have similar manifestations for different reasons. Yet legal corpus writers tend to minimize the possibility of rational disagreement, sometimes claiming that the corpus provides clear, single, correct answers even in the face of manifest disagreement about the implications and causes of corpus results.

In other words, legal corpus work has stood out for its unwillingness to acknowledge the validity of disagreements that most empirical study finds pretty routine. Instead, it has borrowed the conventions of zealous advocacy from legal practice. That insistence indicates that legal corpus work wears an empirical mantle that does not quite fit, refashioning it to conform to the practices and habits of mind that characterize legal representation more than linguistic analysis.

68. Compare, e.g., People v. Harris, 885 N.W.2d 832, 839 (Mich. 2016) (concluding on the basis of a corpus search that “[e]mpirical data . . . demonstrates . . . [that i]n common usage, [the word] ‘information’ is regularly used in conjunction with adjectives suggesting it may be both true and false”), with id. at 850 (Markman, J., dissenting) (concluding on the basis of the same corpus search that the word “information” is presumptively true unless marked otherwise). See generally Carissa Byrne Hessick, More on Corpus Linguistics and the Criminal Law, PRAWFSBLAWG (Sept. 11, 2017), https://prawfsblawg.blogs.com/prawfsblawg/2017/09/more-on-corpus-linguistics-and-the-criminal-law.html [https://perma.cc/G8VX-TM3E] (discussing this case).

69. See generally Anya Bernstein, The Songs of Other Birds, in INSIDERS, OUTSIDERS, INJURIES, AND LAW: REVISITING THE OWEN BIRD’S SONG 219 (Mary Nell Trautner ed. 2018) (discussing the way that similar sociological phenomena can have different meanings depending on their cultural surroundings).

70. See, e.g., Wilson v. Safelite Grp., 930 F.3d 429, 448 (6th Cir. 2019) (Stranch, J., concurring) (expressing “concern with the implicit suggestion that corpus linguistics is a simple, objective tool capable of providing answers to the puzzle of statutory interpretation” and concluding that “[t]he use of corpus linguistics brings us no closer to an objective method of statutory interpretation”).
III. LEGAL CORPUS LINGUISTICS AS A TECHNOLOGY OF LEGAL MEANING

Legal corpus linguistics nicely fits some commonplace understandings of technology: It produces results through the mediation of a computer, and it does so in a way that is invisible to the casual user. To use a corpus prepopulated with search functions like the COCA, one need not know anything about the process through which it yields results, or the theories and methods it rests on, or the alternatives to it that are available. One simply types a snippet into a search bar and presses return. Poof! That certainly feels like technology.

Legal corpus linguistics also constitutes a technology under the more expansive and practice-oriented notion of technology that I laid out in Part I. It situates legal language in larger linguistic contexts as a way to attribute meaning to them—to say what the language of the law means to ordinary people and what we should therefore interpret it to mean for legal purposes. Note that this process is a bit different from the way the technology works in the academic context. Building on academic work, for instance, users might track how the optional complementizer “that” works in legal texts and whether it shows a similar or a different patterning to that of informal conversations. Looking for aesthetic composition—which in linguistics is sometimes called “poetics”—they could look for patterns of alliteration or vowel alternation that surround target terms. Investigating genre distinctions, they could track differences in the syntax of statutes and of other formal writing like academic articles or mathematical proofs. The corpus approach provides a broad range of possibilities for inquiry.

Within that broad range, legal corpus linguistics focuses on a very particular kind of work, which differs from that of academic linguistics. It uses the counting and contextualizing functions of its preferred corpora not just to analyze patterns and distinctions in people’s use of language as a communicative medium, but to provide arguments about the meanings we ought to attribute to legal language. Legal corpus linguistics is a meaning-making machine.

71. See Pierre Bourdieu, The Force of Law: Toward a Sociology of the Juridical Field, 38 HASTINGS L.J. 814, 818 (1987) (“Unlike literary or philosophical hermeneutics, the practice of interpretation of legal texts is theoretically not an end in itself. It is instead directly aimed at a practical object and is designed to determine practical effects. . . . Reading is one way of appropriating the symbolic power which is potentially contained within the text. Thus, as with religious, philosophical, or literary texts, control of the legal text is the prize to be won in interpretive struggles.”).
On this understanding of technology, moreover, the world of law is replete with technologies of interpretation; there are many machines for producing meaning. We have multiple normative theories that tell people how they ought to interpret legal language, such as textualism, purposivism, originalism, intentionalism, and so on. We also have multiple institutional sites where legal interpretation happens: courts, agencies, prosecutors’ offices, police departments, legal journals, and so on. And we have a range of practices that convert legal text into something that has social force—from reading to talking to prosecuting to rulemaking to researching and beyond.

This multiplicity is, in a sense, implicit in the whole idea of legal language, which does many things at once. It describes a situation in the world to which it applies, thereby making explicit or implicit claims about realities outside the legal process. That is, it represents the world. It affects that situation by creating authority, constraint, enforceable conditions, or conclusions about the law itself. That is, it intervenes in the world it describes. In doing so, legal language changes the circumstances of the participants involved, putting them on a different footing than they were before the language was expressed in its legally cognizable form. Legal language also often exerts its power over long-ish periods of time, purporting to constrain not just today’s actions but tomorrow’s, and even the conduct of people not yet born in circumstances not yet imaginable. Just from those factors—the fact that legal language represents and intervenes in society in a temporally expansive way—a Latourian perspective would predict that legal meaning will require ongoing maintenance and concern. Because legal language has such an important role in representing and intervening in society, we should expect that consensus on its meaning will have to be constructed again and again, and again and again will fall apart.

See generally Bernstein, supra note 1 (arguing that while academic corpus linguistics is primarily an empirical exercise, legal corpus linguistics fundamentally rests on a normative claim).

72. See Erving Goffman, Forms of Talk 128 (1981) (“A change in footing implies a change in the alignment we take up to ourselves and the others present as expressed in the way we manage the production or reception of an utterance. A change in footing is another way of talking about a change in our frame for events.”).

73. See Latour, supra note 18, at 246 (“Archimedes spoke for a whole tradition when he exclaimed: ‘Give me one fixed point and I will move the Earth,’ but am I not speaking for another, much less prestigious but maybe as respectable tradition, if I exclaim in turn ‘Give me one matter of concern and I will show you the whole earth and heavens that have to be gathered to hold it firmly in place’? . . . The critic is not the one who debunks, but the one who assembles.”).
The American judicial system, with its common-law style of development, reflects this fact. Every time a judicial opinion decides, with finality, the meaning of some legal language, that judicial opinion itself becomes available for interpretation in some future case. Legal language thus experiences moments of seeming closure while remaining continually available for ongoing reinterpretation. This reinterpretation occurs episodically and sporadically as cases get litigated. With its occasional moments of declared finality always subject to reinterpretation in future moments that yield new and different finalities, the system creates a feeling of stability within a structure of ongoing change. By couching new interpretations in the language of exposition or explication, judicial opinions can continue to claim that previous interpretations were final while changing them over successive cases. And because in the common law system each

---

74. Recent developments in the interpretation of pleading standards the Federal Rules of Civil Procedure provide a nice example of how the language of explication can be used to introduce new interpretations. Rule 8 states that a viable complaint in civil litigation must contain a “short and plain statement of the claim showing that the pleader is entitled to relief.” Fed. R. Civ. Pro. 8(a)(2). The rule says nothing about how, if at all, a plaintiff must present the factual scenario that allegedly harmed her; she is merely instructed to make a claim. What does it take to show that she “is entitled to relief”? For decades, the governing interpretation of that requirement emphasized that the Rules did not require a plaintiff to “set out in detail the facts upon which he bases his claim.” See Conley v. Gibson, 355 U.S. 41, 47 (1957). Rather, a plaintiff need only “give the defendant fair notice of what the . . . claim is and the grounds upon which it rests,” and the court should accept that claim as viable unless “it appears beyond doubt that the plaintiff can prove no set of facts in support of his claim which would entitle him to relief.” See id. at 45–46, 47. In 2007, the Supreme Court reinterpreted Rule 8. See generally Bell Atl. Corp. v. Twombly, 550 U.S. 554 (2007). Twombly, however, did not purport to overrule Conley. Rather, it reinterpreted what Conley accomplished: “Conley . . . described the breadth of opportunity to prove what an adequate complaint claims, not the minimum standard of adequate pleading to govern a complaint’s survival.” Id. at 563. Holding for the first time that pleadings relying on “a formulaic recitation of the elements of a cause of action will not do,” Twombly introduced a new interpretation of Rule 8. Id. at 555. But in rejecting the traditional interpretation of Rule 8, Twombly presented itself not as making a new rule but as merely explaining what Conley was really about. Twombly asserted that, despite how everyone had treated it for fifty years, Conley did not set out a pleading standard at all; it merely gave “an incomplete, negative gloss on an accepted pleading standard” that the Twombly court simply presented. See id. at 563. Two years later, the Court cemented this new approach, using Twombly as the primary interpretation of Rule 8. See Ashcroft v. Iqbal, 556 U.S. 662, 678 (2009) (“As the Court held in Twombly, the pleading standard Rule 8 announces does not require ‘detailed factual allegations,’ but it demands more than an unadorned, the-defendant-unlawfully-harmed-me accusation.” (internal citation omitted)); Anne E. Ralph, Not the Same Old Story: Using Narrative Theory to Understand and Overcome the Plausibility Pleading Standard, 26 YALE J.L. & HUM. 1, 1 (“The Twombly and Iqbal
precedential judicial opinion itself becomes law—and because the part that becomes precedential is not just the result but the reasoning, with its analogies, offhand remarks, and thought experiments—the legal text to be interpreted grows over the history of litigation, providing each successive court with more material to work with and to mold into new legal meanings.\textsuperscript{75}

The common-law style thus provides a system for sporadically reinterpreting already interpreted language through litigation, with no end in sight. A litigant controversy involving a legal text spurs legal interpretation\textsubscript{1}, which presents itself as final, functioning as a kind of stop sign. But like a stop sign, this interpretation creates finality only for a time. A new litigant controversy can put the original legal text and its first legal interpretation—now inscribed into law—into motion again, yielding legal interpretation\textsubscript{2}, which again presents itself as final. And so on indefinitely.\textsuperscript{76}

\textsuperscript{75} See Anya Bernstein, Before Interpretation, 84 U. Chi. L. Rev. 567, 596 (2017) (“Situating statutory text against the background of the corpus juris . . . does not necessarily constrain interpretation or make it predictable. . . . Picking and choosing within the vast legal realm, opinions present their claims about relevance as facts, rather than as arguments.” (internal quotation and citation omitted)).

\textsuperscript{76} See infra Figure 6. This system instantiates the semiotic process in general, as theorized by the American pragmatist Charles Sanders Peirce. Peirce argued that meaning-making involves a sign that refers to a referent and is understood through an interpretant, which then itself becomes a sign for interpretation in subsequent moments of semiosis. See generally Richard J. Parmentier, Signs and Society: Further Studies in Semiotic Anthropology 3–4 (1994) (introducing Charles Sanders Peirce’s Semiotic Theory); Barton Beebe, The Semiotic Analysis of Trademark Law, 51 UCLA L. Rev. 621 (2004) (applying semiotic analysis to trademark law’s dilution doctrine).
At the same time, one of American law’s central claims to legitimacy is to provide certainty and predictability for those subject to it. This claim, of course, runs headlong into the actual common-law style of ongoing change through which that very predictability is supposed to be solidified. Technologies of legal interpretation become crucial to maintaining, or contesting, shared visions of how to give the law meaning and stability within a flexible, changing regime. It is thus also not surprising that they can become a focus of political contestation.

A. Language in Legal Corpus Linguistics

Like any technology, the legal corpus linguistic method for giving meaning to legal text rests on underlying assumptions about how legal texts come to mean things. This observation in itself is no criticism. Some built-in assumptions underlie any technology and indeed any action. A lever builds on understandings of gravity and mass, even if the person building the lever cannot explain or even name the concepts of gravity or mass. A technology’s underlying or

77. As I discuss further in Part IV, American law is also stabilized and changed through administrative, and not just through judicial, interpretation. See infra Part IV. Administrative decisions do not become quite the legal texts that judicial ones do, but administrative procedures, like judicial ones, provide a means to interpret legal texts in a locally stable but globally changing way.
built-in understandings help determine how the technology functions in important ways. Knowing what assumptions a technology rests on gives us a clearer idea of what it does and how it works. As a technology of legal language, legal corpus work builds on understandings of both language and law.

What image of language does legal corpus work present? Think back to the COCA interface, which resembles the interfaces of other BYU corpora favored by legal corpus linguistic studies. It presents a simple, rather uncluttered screen. The search area in particular is quite spare, with a lot of white space. From a Latourian perspective, the search screen itself a kind of actant, a participant with effects that is mobilized in the production and use of this technology. It makes possible the participation of other actants—we who search the corpus for clues about language.

This interface embeds a crucial implication about linguistic meaning: that finding it is generally pretty simple. You just need a word and an Internet connection. In this sense, it is very inviting, soothing, and even allows a broadly democratic participation in the investigation of meaning. The BYU corpora, in particular, are free to use. They require no computational skills since the key search functions are already programmed in, available at the click of an on-screen box. A user thus never has to—nor has the opportunity to—construct a bespoke means of searching. These corpora also require no linguistics training. They come pre-coded for parts of speech and other characteristics. A user need not—and cannot—define or modify the parameters available for search. The overall message is that you do not need a lot of methodological preparation or theoretical sophistication; anyone can plug in a search term and get instant results.

This implication of popular availability is true: Anyone can indeed plug in a search term and get results. But the impression that availability suffices to draw conclusions about legal language use belies the complexities underlying corpus use. The search box does not require or encourage users to recognize, much less make, the methodological decisions that go into delineating a corpus or determining whether it is relevant to a particular inquiry. It gives no

---

78. As with the rest of my critique, I address the technology employed by the legal corpus linguistics literature, which has departed significantly from work done in academic corpus linguistics.

79. Showing how seriously he takes the idea of actants, Latour once gave a speaking part to a train—and a train that was never even built, at that. See BRUNO LATOUR, ARAMIS OR THE LOVE OF TECHNOLOGY 55–56 (Catherine Porter trans., Harvard Univ. Press 1996).
hint of the need to determine what it is that a given search can reveal and where its limits lie. Indeed, to the untrained user, it may suggest that the key to determining linguistic meaning is running a search—a process that the program has automated—rather than interpreting search results, a qualitative process requiring judgment that is itself informed by understandings gleaned from a century of research and theory in linguistics and related disciplines.

The commitment to simplicity visually implied by the search screen resonates in the ways that legal corpus work has generally proceeded. For instance, this literature generally stays focused on individual words or short phrases themselves, rather than asking how these words and short phrases contribute to ongoing meaning-making practices, cohere texts, or create relationships among texts or speech participants. Further, it treats individual words and short phrases as defined and modified through their interactions with their most immediate linguistic surroundings. Recall the results I got when I queried the COCA about carrying a firearm: a series of short snippets, each a sentence or so, showing the phrase in its immediate semantic context. I then had the option of choosing any one of these short snippets to get a slightly longer snippet—a paragraph or so—from the source text. For a legal corpus inquiry focused on finding the ordinary meaning of a particular word, the implication is that this meaning is primarily created through a series of such snippets, unconnected from one another and excerpted from their source text. Thus, a legal corpus inquiry might follow the phrase “carry a firearm” to all the corners of the corpus, asking a binary question about whether carrying is done by a person or a vehicle.80

The technique here implies that words already contain meanings, which they carry with them into various contexts. The contexts used to specify and examine those preexisting meanings, moreover, are small-scale relationships with immediately surrounding texts—what in linguistics is called the syntagmatic relationship of one word to those that appear around it. So, for instance, each word in the utterance “I have read that book” stands in a syntagmatic relationship to the others. The different role each word plays in the sentence, as defined by a given language’s conventional sentence structures and grammatical markings, helps audiences give it meaning. Although the word “read” is pronounced the same as the word “red,” its syntagmatic placement—as part of a verb phrase with the verb “have” rather than a noun phrase with the noun “book”—would help a hearer

disambiguate the homonyms. Similarly, prosody and syntax help an audience member understand whether the word “that” acts as a deictic pointing to a particular instance of the direct object “book,” or if it instead introduces a further clause with “book” as the topic, as in “I have read that book learning is worthless.” Syntagms yield meaning by relating the sentence placement, morphology, and utterance role of words that appear in an utterance.

Meanwhile, paradigm sets help give words meaning by contrasting other words with similar utterance roles—words that could have appeared in the same place, but did not. So, for instance, the direct object “book” in my example utterance is in a paradigm set with words like “article,” “novel,” “music score,” and other things that can be read. Similarly, the verb “read” is to some extent defined by the things it is not—words that could have appeared in the same place but did not, such as “seen,” “loved,” “bought,” “burned,” and so on.

Although the BYU corpora do provide ways of getting at similar parts of speech, the search screen gives no hint that a word’s meaning is constructed through contrasts in both syntagmatic and paradigmatic directions, and legal corpus writers have generally not investigated that paradigmatic axis. Academic corpus linguistics, meanwhile, “work[s] on the assumption that meaning is created on both [syntagmatic and paradigmatic] axes . . . . There is no reason why one should have a priority in meaning potential over the other.”

81. JOHN SINCLAIR, TRUST THE TEXT: LANGUAGE, CORPUS AND DISCOURSE 170 (John Sinclair & Ronald Carter eds., 2004); see also infra Figure 7.
The legal corpus technique thus focuses on the syntagms that appear on the screen, not on the other possibilities and relationships that contextualize them and give them meaning. Treating language as being primarily about words in the utterance leads it to ignore the key role of paradigmatic relations in meaning, as well as the key role of genre conventions in linguistic usage.  

Instead, legal corpus writers have generally treated corpora as internally undifferentiated. The COCA, for instance, is used to represent generically “ordinary” language use. Here, the technology itself does allow users to attend to genre: The COCA gives some genre specification, noting the broad category—newspaper, academic, spoken, and so on—a given result comes from. But legal corpus technique has not incorporated genre into its analysis. Instead, it treats ordinary usage as one-size-fits-all. In contrast, academic linguists have been attentive to genre distinctions within and between corpora. Indeed, as academic linguistics has demonstrated, different genres use even standard English differently—for instance by drawing on different kinds of lexical bundles and deploying them in different ways. 

Similarly, as I discuss in Part I, academic linguists often use corpora to trace relationships—within texts, among texts, among people, and between speakers and texts. On this approach, focal texts are not so much bearers of unchanging meaning as means of forming relationships among texts and people: They distinguish genres, distribute information, inject individuality, and so on. Even when linguists focus on a single word used in a particular grammatical formation, like complementizer that, they are likely to be interested in the way the word works to distinguish some utterances from others, to structure the conveyance of meaning, and to affect the relationships of speech participants. They speak in terms of functions or effects. For instance, linguists might ask what complementizer that does in a sentence, how it works, and how it influences the ongoing flow of meaning-making that surrounds it, rather than to try to assign some ultimate meaning to the word or adjudicate which of many usages represented in a corpus should count as “ordinary.” Legal corpus writers, in contrast, tend to fixate on the question of individual word


83. See Biber, supra note 34, at 283.
meaning, and to treat meaning as something that inheres in the word itself, rather than in its functions and relationships.\textsuperscript{84}

This blithe assumption that words simply “have meaning” ignores the way that both meaning and usage depend integrally on \textit{social}, and not just semantic or syntactic, context.\textsuperscript{85} When my host puts a plate down on the table between us and says, “That chocolate is delicious,” I am likely to understand that she is offering me the chocolate and encouraging me to try it as a part of her hospitality. But when, after I have taken a nibble, I respond with that same sentence, “That chocolate is delicious,” my host will likely understand that I am thanking her and showing appreciation, not offering her the chocolate as a demonstration of \textit{my} hospitality. That is because the host–guest relationship functions as a kind of social-interactional genre that carries with it patterns of practice, somewhat presupposable interactional roles, stock phrases and presumptions, and so on. In short, it forms a “frame” within which people make and interpret utterances.\textsuperscript{86} So when my kid, after spending an hour in the supermarket helping me shop, points to the shelf by the checkout counter and says coyly, “That chocolate is delicious,” he is counting on me to work within the frame we jointly occupy and to understand that he means neither “please eat some” nor “thank you for your hospitality.” Just from our everyday lives, it is easy to see how pragmatics—the social situation of communication—affects meaning

\begin{itemize}
\item \textsuperscript{84} See, e.g., Thomas R. Lee & Stephen C. Mouritsen, \textit{Judging Ordinary Meaning}, 127 \textit{Yale L.J.} 788, 795 (2018) (“Our thesis is that words have meaning, and that meaning can be theorized and measured using principles and methods devised in the field of linguistics.”).
\item \textsuperscript{85} See \textit{id.}
\item \textsuperscript{86} Charles J. Fillmore and Beryl T. Atkins, \textit{Toward a Frame-Based Lexicon: The Semantics of RISK and its Neighbors, in FRAMES, FIELDS AND CONTRASTS: NEW ESSAYS IN SEMANTIC AND LEXICAL ORGANIZATION} 76–77 (Adrienne Lehrer, Eva Feder Kittay & Richard Lehrer eds., 2012) (“[I]n semantic theories founded on the notion of \textit{cognitive frames} or \textit{knowledge schemata} . . . a word’s meaning can be understood only with reference to a structured background of experience, beliefs, or practices, constituting a kind of conceptual prerequisite for understanding the meaning. Speakers can be said to know the meaning of the word only by first understanding the background frames that motivate the concept that the word encodes. Within such an approach, words or word senses are . . . related . . . only by way of their links to common background frames.”); see also Charles J. Fillmore, \textit{Frame Semantics, in COGNITIVE LINGUISTICS: BASIC READINGS} 373 (Dirk Geeraerts ed., 2006) (“By the term ‘frame’ I have in mind any system of concepts related in such a way that to understand any one of them you have to understand the whole structure in which it fits; when one of the things in such a structure is introduced into a text, or into a conversation, all of the others are automatically made available.”).
\end{itemize}
making and structures the semantic meanings and syntactic forms that interest legal corpus analysts. Yet pragmatics does not generally enter into those studies at all.

The pragmatic contexts of laws are, of course, much weirder and more unusual than those in my chocolate-based examples. They occur only under very particular, very peculiar, conditions, structured by myriad rules, legal constraints, and traditions, within very specific social institutions peopled by very unordinary participants, and with effects on the world matched by no other speech act. That weirdness and rareness does not make pragmatic relations in law any less important. On the contrary, it should alert us to the difficulty of finding language use examples that would help evaluate the language of legal texts. Making claims about what legal terms mean in nonlegal contexts, in other words, may require more than simply investigating nonlegal contexts. We need reasons to believe that those nonlegal contexts give information about the legal texts they are used to examine.

Work in legal corpus linguistics, for the most part, does not bother to justify or theorize the relationship between, say, a general corpus like the COCA and the highly unusual pragmatic circumstances of legal texts. It merely recites the familiar adage that legal language should be interpreted based on its ordinary usage. That view makes sense if words are meaning-containers that can be moved about like vases, retaining their integrity and their contents over different situations. But if paradigmatic, genre, and pragmatic contexts all change word usage and meaning—if a word is less like a meaning-containing vase and more like a chameleon that retains structural integrity while changing color with its situation—then directly imputing the ordinary usage of one frame into another becomes more complicated. Legal corpus linguistics generally progresses on the word-as-vase assumption and does not discuss how changes to contexts and frames could affect the analysis of legal language. Yet linguistics itself has demonstrated that linguistic meaning involves multiple contexts—contexts that embed not only individual words but also one another. Legal corpus linguistics focuses most on the
syntagm, which is the most localized context and also the one most embedded in other contexts that affect meaning.\(^87\)

![Figure 8.](image)

A legal corpus proponent might counter that it is courts, and not legal corpus scholars, who insist that we should interpret legal language according to its ordinary meaning. And indeed, the notion of ordinary meaning as a basis for legal interpretation has a long history in American law. For most of that history, however, courts have understood ordinary meaning in a negative sense as a way of helping rule out outlandish or implausible interpretations: “if a meaning can be the intended one in everyday usage, then that meaning passes the ordinary meaning test.”\(^88\) This “casual understanding of what is ordinary” differs from most legal corpus proponents’ attempts to use a corpus to glean the single real, correct, or truly ordinary meaning of

---

87. See infra Figure 8. This image is not meant to exhaustively show all the various contexts that go into creating linguistic meaning, but simply to illustrate some major meaning-making components.

88. Lawrence M. Solan, *Corpus Linguistics as a Method of Legal Interpretation: Some Progress, Some Questions*, 33 INT’L J. SEMIOTICS L. 283, 286 (2020); see id. at 287 (arguing that, in nineteenth century, the Court used “ordinary meaning” to express the notion that in everyday speech, the meaning that the Court attributes to the statute would be acceptable to the normal speaker/hearer of English”).
a legal term, as opposed to simply indicating the wide range of ways a word can reasonably be used.89

Legal corpus linguistic technology and technique thus build in a particular idea—one might even say ideology—of what legal language is and how it works.90 In this image, meaning is relatively simple to find and stable across populations, genres, and situations. Further, meaning inheres in words—and specifically words found in works of fiction, popular publication, academics, and national broadcast shows. And of course, the implication of using this technology is that it will help us answer the questions we pose.

B. The Law in Legal Corpus Linguistics

The questions we pose, in turn, are legal ones. And so it is worth thinking about what visions of the law legal corpus techniques embed. Again, any technology embeds presuppositions. These visions are not necessarily problematic or wrong—a successful lever will build in assumptions about gravity and mass that are very useful to moving things. But illuminating a technology’s presuppositions gives us a chance to determine whether we agree with them and what views we implicitly lend support to when we utilize the technology ourselves.

The ideologies of language described in the previous section yield clues to the notion of law embedded in legal corpus work. As Don Kulick has written, “language ideologies seem never to be solely about language” so much as they are about other things that surround language use.91 Often those other things involve categories or groups to which a speaker is seen to belong: Language ideologies might predict how people of particular genders, races, or class backgrounds will speak, and reciprocally attribute particular speech patterns to the gender, race, or class background of particular speakers.92 In the case

89. See id. at 286.
90. As Michael Silverstein has noted, ideology—as a set of understandings that explain a phenomenon and allow people to orient their conduct and worldviews toward it—“is characteristic of any sociocultural phenomenon,” and “must inhere in what makes any social entity . . . cohere as that social entity”: “there is no such thing as a social fact without its ideological aspect.” Silverstein, supra note 15, at 313.
of legal corpus research, ideologies of language merge with and enable ideologies of law. One might say that legal language ideology inscribes assumptions about how law itself speaks.

In the preceding section, I explained how legal corpus linguistics presents language as primarily determined by the syntagmatic relations of individual words to the words around them, relying on frequency counts to declare which kind of syntagm wins. This vision of language, I argued, ignores some of the primary constituents of communication like paradigm sets, genre conventions, and pragmatic circumstances. Legal corpus linguistics does something similar with the concept of law.

Legal corpus linguistics imagines law as primarily a written text, and specifically a written text that comes in clumps of one to four words. Those are the things whose meanings legal corpus writers seek by contextualizing them within a series of syntagms drawn from areas other than the law. The meaning of the law, meanwhile, is set from the start: It inheres in the individual words or word clumps, and it does not change over time or circumstance. It does not depend on its larger textual surroundings, so we need not look to the rest of the statute and ask how this provision makes sense in light of the statute’s overall project. And we need not ask what circumstances pushed Congress to make the enormous effort required to produce and enact a statute.

What members of Congress understood the statute’s effects on the world to be is similarly irrelevant to this inquiry. With the important exception of Constitutional interpretation, legal corpus inquiry usually looks specifically to sources outside the legal realm, produced by people other than those who produce laws.93 So the way one law fits into or relates to other legal schema, as well as the way legal terms change through the ongoing process of legal interpretation itself, is erased: the words mean what the nonlegal corpus shows them

93. In contrast, legal corpus work in constitutional law routinely looks to contemporaneous legal materials and writing by those who were involved with producing, debating, or ratifying the Constitution. See, e.g., Jennifer L. Mascott, Who Are Officers of the United States, 70 STAN. L. REV. 443, 443 (2018) (looking to contemporaneous political and legal actors and practices to evaluate original meaning of the constitutional phrase “officers of the United States”). See generally Brief for Cunningham & Egbert, supra note 2 (using contemporaneous legal and political sources to evaluate the constitutional term “emoluments”). Legal corpus writers have not explained this differential treatment for constitutional and statutory law, but it does map on to the differential treatments given by statutory textualists and constitutional originalists.
meaning, not what they might have come to mean through the ongoing common-law process or through historical change to the social situation the law regulates.

In this vision, legal words come prepackaged with meaning; the interpreter’s job is to unearth the true meaning that inheres in the words. In other words, for legal corpus linguistics, legal meaning is frozen at the time of enactment into words isolated from social, and even legal, surroundings. And we can know what that eternal frozen meaning is by looking to how those individual words are used by people who have nothing to do with producing or implementing the statute.

These convictions stand in some tension with the common law system of American adjudication, which leaves legal texts open to recurring reinterpretation over time and creates pathways for meanings that are distinctly legal rather than ordinary. For an example, consider something as basic as the federal question jurisdiction of the federal courts. Both the Constitution and federal statute give federal courts jurisdiction over issues “arising under” federal law. But, as law students learn in their first-year civil procedure course, those words have come to mean different things in different contexts. The reach of Supreme Court jurisdiction under the Constitution “may extend to all cases in which a federal question is ‘an ingredient’ of the action.” But a lower federal court may only entertain a civil suit if federal law is integral to the plaintiff’s ability to lodge her complaint and central to the court’s ability to decide it.

94. See supra Figure 6.
95. Thanks to Mark Rosen for suggesting the arising-under example.
96. See U.S. CONST. art. III, § 2 (“The judicial power shall extend to all cases, in law and equity, arising under this Constitution, the laws of the United States, and treaties made, or which shall be made, under their authority.”); 28 U.S.C. § 1331 (“The district courts shall have original jurisdiction of all civil actions arising under the Constitution, laws, or treaties of the United States.”).
This odd situation routinely confounds civil procedure students. Yet the basic idea that the same words might mean one thing in the Constitution and another in a statute is not that hard to grasp. These phrases, though identical in their word composition, have lived separate jurisprudential lives, becoming interpreted and redefined over the course of iterative litigation—resolving moments. And they were introduced into their respective legal texts under different political and social circumstances, at different points in history, for different reasons. While arising-under jurisdiction was part of the Constitution in 1789, Congress did not write it into statute until 1875, after the experience of the Civil War had helped normalize an increased role for national government.

From the legal corpus linguistics standpoint, this example poses a bit of a conundrum. If individual words or small word bundles contain immutable legal meanings that can be discovered by looking to nonlegal syntagmatic arrangements in which these little texts appear, the different landing points of this same phrase become difficult to explain. Did the phrase “arising under” mean something significantly different among ordinary speakers—whoever they might be—in 1875 than it did in 1789? Did the ordinary speakers in 1789 habitually use “arising under” to mean “implicated in” or “an ingredient of,” while the ordinary speakers of 1875 used it to mean a contrasting notion of “clearly required for” or “integral to,” the way Constitutional arising-under jurisdiction contrasts with statutory jurisdiction continues to pose puzzles for lawyers and scholars alike. See Lumen N. Mulligan, A Unified Theory of 28 U.S.C. § 1331 Jurisdiction, 61 VAND. L. REV. 1667, 1668–69 (2007) (“Given the weighty doctrinal and pragmatic consequences that flow from determining whether a claim falls within the scope of § 1331, it is surprising to learn that we lack a coherent view of what statutory federal question jurisdiction entails.”). See Anthony J. Bellia, Jr., The Origins of Article III “Arising Under” Jurisdiction, 57 DUKE L. J. 263, 265 n.2 (2007).

In 1875, Congress gave federal circuit courts jurisdiction of ‘all suits of a civil nature, at common law or in equity . . . arising under the Constitution or laws of the United States, or treaties made, or which shall be made, under their authority,’ subject to an amount-in-controversy requirement of five hundred dollars. Act of Mar. 3, 1875, ch. 137, § 1, 18 Stat. 470. Following reenactments and reformulations of the 1875 act, federal district courts ‘have original jurisdiction of all civil actions arising under the Constitution, law, or treaties of the United States.’ 28 U.S.C. § 1331 (2000). See id. See generally LAURA F. EDWARDS, A LEGAL HISTORY OF THE CIVIL WAR AND RECONSTRUCTION: A NATION OF RIGHTS (2015) (analyzing changes to legal understandings of national government authority in both north and south during the Civil War and its aftermath).
jurisdiction? It seems unlikely, and indeed a brief review of the relevant BYU corpora suggest no such thing.\textsuperscript{101}

If this inquiry seems like a very roundabout way to go about figuring out what “arising under” means, or should mean, in either the Constitution or the statute, that is because the legal corpus approach erases most of the stuff that gives these words meaning and differentiates the phrases from one another—history, social context, legal genealogy. Legal corpus linguistics, in other words, may indeed be quite useful for identifying the broad range of meanings a given word or word phrase could acceptably have had at some particular point in time—a “casual understanding of what is ordinary.”\textsuperscript{102} It can help determine if ordinary speakers at some particular time would have found the attribution of some particular meaning to a given word or phrase completely outlandish. But it cannot reveal what a given word or phrase \textit{should} mean in a specific legal context because a corpus of instantiations does not reveal which of a multitude of acceptable uses is the one that should apply in some particular instance.

For legal interpretation, though, we need to decide more than the broad range of uses that speakers would accept. We need to give the term meaning in the law at issue. Journals and newspapers cannot tell us what that ought to be. Legal corpus users, in contrast, routinely claim that corpus study reveals not just the range of potential uses of a word in general, but the best or most likely meaning of the word in its particular legal context. Such contentions, often presented as empirical findings about what a term actually means, in reality articulate normative positions about how we ought to interpret the term.

Moreover, as soon as a legal text is written, the world around it changes. For one thing, a legal text—if it works—changes the world itself. Beyond that, though, people invent new machines, develop new

\textsuperscript{101} The Corpus of Founding Era American English (COFEA) lists 233 instances of the phrase “arising under,” most of which have something to do with legal claims and the laws that support them. See \textit{Corpus of Founding Era American English (COFEA)}, supra note 52. Due to the nature of the COFEA, these examples primarily come from texts where people propose, debate, or discuss Constitutional or legal texts; this will not tell us how people not connected to the legal establishment used the phrase “arising under” under other circumstances. The Corpus of Historical American English (COHA) returns no results for “arising under” in the 1870s, when the jurisdiction statute was enacted. Most of the results for the 1860s and 1880s (ten each) are quotations or paraphrases of the constitutional or statutory language. See \textit{Corpus of Historical American English (COHA)}, supra note 51.

\textsuperscript{102} See Solan, supra note 88, at 286.
practices, and set new norms. They even develop new laws that form
new contexts for existing statutes.\textsuperscript{103} The social, material, and legal
world in which any legal text lives is constantly shifting under its feet.
These normative convictions thus also stand in tension with the way
laws continue to apply to the world around them even when that world
changes, often in unpredictable ways, giving even old meanings a new
significance because of changed circumstances—much the way “that
chocolate is delicious” has a different social effect when spoken by a
host than when spoken by a guest.\textsuperscript{104} Legal corpus technology and
technique suggest that practitioners should ignore the essentially
social nature of law, treating it instead as a purely linguistic
endeavor—but a linguistic endeavor divorced from most of ways
language usually gains meaning.

These ideas about law should sound familiar. They are mainstays
in textualism and originalism, prominent normative theories of legal
interpretation. Textualism holds that the meaning of a legal text should
be seen to inhere in its words only, and it particularly focuses on
individual words or small word clumps. Originalism holds that the
meaning of a legal text should be seen to always mean what its words
would have meant to contemporaries when it was enacted. For both
normative theories, the meaning of legal words should not be seen to
change over time. Adherents to these normative theories have also
emphasized that we should use ordinary language to interpret legal
language. The technology and technique of legal corpus linguistics
thus build in, as structuring presuppositions, the normative
convictions of textualism and originalism. Where users do not
acknowledge and account for that relationship, using this technology
enacts those normative convictions.

Legal corpus analysts might respond that they make no
normative claims. Legal corpus analysis, on this telling, simply
provides judges and other interpreters with information that they may
use however they want, or not at all. In fact, legal corpus users might
say they have no stake in the outcome of their research. They are just
doing what judges—always on the lookout for ordinary meanings—

\textsuperscript{103} See, e.g., FDA v. Brown & Williamson Tobacco Corp., 529 U.S. 120,
161, 164 (2000) (holding that despite giving the Food and Drug Administration
jurisdiction over “articles . . . intended to affect the structure [and] any function of the
body,” the Food, Drug, and Cosmetics Act could not give the FDA jurisdiction over
tobacco products—which are articles intended to affect the structure and function of
the body—because subsequent statutes indicate that Congress did not wish to do so).

\textsuperscript{104} See generally WILLIAM N. ESKRIDGE, JR., DYNAMIC STATUTORY
INTERPRETATION (1994).
want them to do. They are a neutral party using a neutral technology to do neutral research that can then be used in various ways.

As we have already seen, though, this claim is a bit unrealistic: Technologies of meaning are not neutral or passive. People doing legal corpus analysis may indeed be unbiased, in the sense that they do not necessarily favor one litigant or one litigation outcome over another. But that hardly means the technology they use is somehow immune from the characteristics of technologies in general: building in worldviews, affecting conduct, distributing salience, and influencing the development of further ends. To evaluate legal corpus linguistics as a technology of interpretation, then, we should not get too hung up on the intentions of its users. Rather, we should consider what understandings of law and language the technology embodies, how it molds conduct, and what further questions it suggests and obscures.

In the case of legal corpus linguistics, the techniques it employs and the answers it yields present legal meaning as something that inheres in words that sit in syntagms and can be discovered scientifically—rather than something that arises from a social world in which legal strictures are mobilized and have effects. This technology supports some existing normative claims in legal theory—that legal meaning is a simple affair; that meaning inheres in one or a couple words and stays stable across social and historical contexts; and that figuring out what a statutory provision means is basically a matter of figuring out what each of its words or phrases mean, separately.

Even if one does not consciously share those views, this technology of meaning-making can be very attractive. The interface is so clean. The results are so fast. It makes the project of legal interpretation, which can seem hopelessly muddled, seem finally doable, and maybe even easy. The way the inquiry is configured makes it look like the terrain around legal language does not change.

And of course, in that classic way in which technologies create paths for path dependence, a person using this technology can find it easy to narrow her focus away from the social production and effects of law and toward the individual words or phrases she plugs into the search bar. We have this wealth of objective, neutral information at our fingertips; it can feel natural to put it to use. The frictionless feel of this method pulls users toward its implicit claims about words, law, society, and change. In this way, decisions about how to interrogate legal texts can be mistaken for characteristics of the text itself.

**IV. PUTTING LEGAL CORPUS LINGUISTICS IN AN INTERPRETIVE ECOSYSTEM**

Having thought about some of the presuppositions and entailments of this technology, we can also place it in a larger ecology of meaning-production technologies. Here, I choose just two comparators to sketch how that might work. These comparators themselves present a stark contrast: dictionaries on the one hand, administrative decision-making on the other.

Each of these might appear at first glance to be something other than a technology. Compared to computer programs, dictionaries seem so stodgy, so unidimensional, so *physical*. On the other hand, administrative procedures seem more like a party: social events that involve lots of people over lots of time, not at all like the solitary act of plugging in a search term and seeing what the corpus yields. Yet each of these is a meaning-making machine just like legal corpus research. Each gives people a way to attribute meaning to legal language by recourse to something other than the legal language itself. There is, in other words, no good reason to limit the comparison to options that strongly resemble legal corpus methods. On the contrary, comparing legal corpus work to things that look a lot like legal corpus work would not give us much information about what sets legal corpus work apart and what we gain or lose by choosing it as opposed to other ways of making meaning.\(^\text{106}\)

A. Dictionaries

Dictionaries have figured prominently in discussions of statutory interpretation.\(^{107}\) And several legal corpus proponents have, in private communications with me, defended their work as at least better than dictionaries, which judges are free to use anyway. So it seems worthwhile to compare the two technologies explicitly.

Legal corpus work undeniably provides more information about how a given word or word bundle is used in some range of communicative situations than a dictionary does. Yet these technologies of meaning-making bear some important resemblances as well. For instance, using a dictionary to determine the meaning of a statutory term encourages users to focus on individual words or very small word bundles, rather than the overarching contexts in which those words gain meaning and have social effect. This limited focus becomes apparent in the way judges use dictionary definitions, which differs from how an everyday reader of, say a novel, normally would.

When I come across an unfamiliar term, I might go to a dictionary to help me understand it. Personally, I like to use a tattered American Heritage I have had since I was a kid, but these days I might simply type the term into my search engine query bar and see what pops up. When I have a couple of definitions, I go back to the text I am reading and decide how the dictionary’s entries would make the most sense in the context of the thing I am reading. If the definitions do not instantly clarify things, I know that I have to make decisions about which definition makes the most sense in, and of, the context. I also may need to decide whether a definition that combines aspects of several dictionary entries or follows a tangent implied by some entry would make more sense than the specific entries the dictionaries yielded. Even after getting some general ideas from a dictionary, in other words, I still have to make crucial decisions about what I think this word should mean in the specific place I encounter it. One might say I treat a dictionary definition as a springboard, a reminder, or a pointer—but not as an answer that I insist the text must fit.

---

In contrast, when courts use dictionaries, they often proceed as though one could—or even must—simply transplant a dictionary definition into a text without making any decisions about how the word at issue would make the most sense. That is, courts sometimes treat the dictionary as obviating the court’s own obligation to give meaning to the statute—as though a court’s work were to make the statute conform to whatever dictionary entries it collects, rather than to use the dictionary entries to help decide what the statute should mean.

The oddity of this approach comes through nicely in Chapman v. United States, where the Supreme Court interpreted a statutory provision that imposed a mandatory minimum prison sentence on a person who distributed more than one gram of a “mixture or substance” containing LSD. Unlike drugs such as heroin or cocaine, which can be sold pure or cut with some substance that dilutes them, LSD in its pure form weighs so little that it must be sold on a carrier. The drug is “dissolved in a solvent such as alcohol, and . . . the solution” is then dripped or sprayed onto something that is easier to divide and distribute—often paper or sugar cubes, but potentially many other things. The alcohol eventually evaporates, leaving pure LSD in the paper, sugar cube, or other carrier. A user ingests the carrier to get the LSD dose.

Should the paper, sugar cube, or other thing onto which the LSD and alcohol solution has been dripped count as part of the drug “mixture or substance” when determining whether a defendant had distributed more than one gram? Chapman’s majority opinion sought the answer in dictionaries. One dictionary defined “mixture” as combining components that “however thoroughly commingled[,] are regarded as retaining a separate existence.” Another described it as “two substances blended together so that the particles of one are diffused among the particles of the other.” Rather than using these dictionaries as a springboard for deciding how this provision fit into the statute’s overall aims, effects, and other provisions, the opinion treated the definitions as boxes into which the statutory text must be stuffed. The majority decided that the carrier formed part of a mixture

109. See id. at 455, 459–60.
110. See id. at 457.
111. Id. at 462 (quoting WEBSTER’S THIRD NEW INTERNATIONAL DICTIONARY 1449 (1986)).
112. Id. (quoting 9 OXFORD ENGLISH DICTIONARY 921 (2d ed. 1989)).
with LSD. Since LSD stays “inside” the carrier, “commingled” but not “chemically combine[d] with [it],” and could “be released by dropping the [carrier] into a liquid or by swallowing [it],” the drug “retain[ed] a separate existence,” just like the first dictionary entry said. But since “LSD is diffused among the fibers of the paper,” the carrier “cannot be distinguished” from the LSD “nor easily separated from it,” as the second definition held.

The majority’s two conclusions—that LSD (1) stays separate from its carrier to be released upon contact with water, and (2) is not distinguishable or easily separated from its carrier—are mutually contradictory. But the opinion betrayed no recognition of the incompatible qualities it attributed to the drug. Instead of treating the dictionary definitions as an indication of the types of things the word mixture might mean, considering which versions might make the most sense in the context of the statute, and then justifying its conclusions on legal or policy grounds, the majority treated the dictionary definitions it had chosen as criteria that the statutory words had to fit.

This is not how most people use dictionaries in everyday life. We understand that texts are not written to fit dictionary definitions. And we recognize that definitions in dictionaries are general and illustrative, while the meanings of words in the texts we interpret depend on their specific contexts.

Courts using dictionaries to interpret legal language thus often use techniques that are analogous to those used in legal corpus analysis. They take a word out of its statutory context, put it in the context of illustrative examples of general usage, find commonalities in those examples, and then attribute those commonalities back to the specific use at issue. And they treat the word (or short word-bundle) as the primary unit of meaning, to the exclusion of paradigm relations, genre conventions, pragmatic contexts, and so on.

By allowing users to render invisible the numerous determinative contexts in which legal language takes shape and force, dictionaries do indeed share some of legal corpus work’s thin, limited image of how meaning is produced. Because dictionaries provide much less information than a corpus search, though, and because they do not provide a lot of raw data, they may give users less of a sense of false mastery than a corpus search. That is, the presuppositions they build in may be less powerful and the entailments less totalizing. Plus, some dictionaries provide a picture of language change over time

113. Id.
114. Id.
when they give examples from different historical periods, showing clearly how even solo words continually change. So there are plusses and minuses to each one. Dictionaries may provide less, but they claim less too.

B. Agencies

The preceding section put legal corpus linguistics in the context of its close cousin, dictionary consultation. But given their close relationship, leaving the comparison there would be unnecessarily limiting. What we want, after all, is not a better dictionary. What we want are good ways of attributing meaning to legal texts that suit our ideas about both law and text. As Daryl Levinson has pointed out in another context, how we draw the borders of our analysis determines what we see.\textsuperscript{115} So it makes sense to look further afield. In this section, I consider a pervasive but quite different technology of legal meaning-production: administrative rulemaking.

One thing that legal corpus linguistics shares with dictionary consultation is that both are generally offered as technologies for use by individual judges. A judge pondering a legal decision can go to the dictionary and look up some key terms, or go to the computer and type a few entries into the corpus. The prototypical image of statutory interpretation echoes the wonderful Wallace Stevens poem: “The house was quiet and the world was calm. / The reader became the book.”\textsuperscript{116} This is a beautiful and inspiring image. But Stevens probably

\textsuperscript{115} See Daryl J. Levinson, *Framing Transactions in Constitutional Law*, 111 YALE L.J. 1311, 1313–14 (2002) (arguing that the viability of a public law claim of right infringement depends on how observers decide to “frame” the relevant “transaction,” a choice that precedes decisions about how to apply or interpret the right itself).

\textsuperscript{116} Wallace Stevens, *The House Was Quiet and the World Was Calm*, in *The Palm at the End of the Mind: Selected Poems and a Play* 279, 279 (Holly Stevens ed., 1990). The poem, well worth reproducing in full:

The house was quiet and the world was calm.
The reader became the book; and summer night

Was like the conscious being of the book.
The house was quiet and the world was calm.

The words were spoken as if there was no book,
Except that the reader leaned above the page,

Wanted to lean, wanted much most to be
The scholar to whom his book is true, to whom
did not have statutes in mind. Statutes are speech acts: they accomplish things in the world. We may strive to understand what they really say—to become, as the poem says, the book and the truth. But in the end, the effect of interpreting is to determine how statutes get implemented, that is, how it is that they accomplish things in the world.

Moreover, 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 to interpret the true meaning of the statute. This inherent multiplicity arises not only because judicial interpretations of statutes only happen within the adversarial context of litigation. It also comes about 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

The summer night is like a perfection of thought.
The house was quiet because it had to be.

The quiet was part of the meaning, part of the mind:
The access of perfection to the page.

And the world was calm. The truth in a calm world,
In which there is no other meaning, itself

Is calm, itself is summer and night, itself
Is the reader leaning late and reading there.


suffices to disrupt the house-is-quiet image of a reader reaching a true understanding of a text.\textsuperscript{120}

Administrative rulemaking is a vast and complex enterprise with different agencies taking different approaches at different times. But there are some key similarities worth considering in a comparison to legal corpus linguistics. One striking characteristic: Agency interpretation involves a lot of people. Indeed, even the basic requirements that structure rulemaking require agencies to take and respond to input from anyone who cares to comment on proposed rules.\textsuperscript{121} Agencies also routinely gather numerous people to develop rules based on statutes, involving discussions over substantial periods of time, as multiple administrators collect information, do research, outline policy priorities, consider legal requirements, and assess Congressional reactions to potential rules.\textsuperscript{122} Moreover, decisions about rulemaking routinely involve many actors within the administrative state, taking input from an agency’s program offices, policy experts, legal professionals, legislative liaisons, White House representatives, and more. A single decision might travel up and down this multifarious hierarchy as the agency decides whether to pursue a


\textsuperscript{121}See 5 U.S.C. § 553; see also United States v. N.S. Food Prods. Corp., 568 F.2d 240, 247 (2d Cir. 1977).

policy, drafts a proposed rule, responds to comments, and publishes the final rule.\textsuperscript{123}

This process poses a stark contrast to the calm, quiet world in which interpreting is often imagined. It speaks instead of the cacophony that attends most statutory interpretation in practice. Legal interpretation usually involves many—perhaps hundreds—of people, working in different ways on different aspects of an issue, iteratively confronting and resolving problems, disagreements, incoherencies, and values choices.

This work can be said to produce an emergent \textit{institutional consolidation} around a statutory meaning. Despite the complex and multi-participant process that goes into producing this institutional consolidation, moreover, no one can be under many illusions about its finality or permanence. The potential for change is baked into the system, as agencies wield ongoing authority over the statutes they implement, rendering every interpretation potentially subject to a new and different institutional consolidation down the line. Such stability within change resembles the common law system’s built-in flexibility. If anything, given the widespread recognition that agency interpretations can and do change, it lacks the common law system’s illusory claims to finality.\textsuperscript{124}

This image also suggests that, in the statutory realm, the question of interpretation may be inseparable from the question of implementation. Statutes, after all, are not just things that mean. They are things that \textit{do}.\textsuperscript{125} The idea of “understanding” a statute may be incoherent if it is divorced from the idea of effectuating or implementing it. And since lawsuits determine how a statute will act

\textsuperscript{123} In separate research, Cristina Rodríguez and I have been conducting interviews on statutory interpretation and policy formulation practices with federal agency personnel. This description is drawn from some of those interviews.

\textsuperscript{124} For example, the potential for changed interpretation has become a mainstay of judicial doctrine. \textit{See, e.g.}, Chevron U.S.A., Inc. v. Nat. Res. Def. Council, 467 U.S. 837, 866 (1984) (holding, in a challenge to an agency’s changed statutory interpretation, that courts should defer to agencies’ reasonable interpretations of ambiguous statutory provisions); Nat’l Cable & Telecomms. Ass’n v. Brand X Internet Servs., 545 U.S. 967, 1000–01 (2005) (holding that courts should defer to an agency’s reasonable reinterpretation of an ambiguous statutory provision even when a court had previously upheld a different interpretation by that agency as reasonable).

\textsuperscript{125} \textit{See supra} notes 21–24 and accompanying text; \textit{see also} Austin, \textit{supra} note 23, at 6–7; Searle, \textit{supra} note 23, at 16; MARIANNE CONSTABLE, OUR WORD IS OUR BOND: HOW LEGAL SPEECH ACTS 18–19 (2014).
on the ground, effectuating or implementing is what judges do, too, whether they phrase it that way or not.

The multiplicity inherent in agency legal interpretation finds echoes in the multiplicity of legal corpus work. Legal corpus work, too, seeks paths away from the centrality of individual judges and their internal states and intuitions. It, too, looks instead to multiple voices. But it looks not to the voices of people participating in the ongoing social practice of giving statutes meaning, but to those in the corpus. And it treats this multiplicity differently than agency interpreters do. Legal corpus work does not provide for debate or disagreement the way that the multiparty discursive process of agency interpretation does. Instead, legal corpus analysts look to multiplicity as a path to averaging out differences and providing clear answers about enduring underlying meanings—meanings that already inhere in statutory language waiting to be discovered.

Both agency procedures and legal corpus linguistics are machines for giving meaning to law. Both take into account factors beyond a judge’s intuitions. Yet, we see how the images of law, language, and technology they rest on are quite dissimilar. Legal corpus linguistics presents itself as a fast, simple, automated technology of meaning-discovery. Agency interpretation is a slow, complex, messy, and very human process of meaning-making. Legal corpus linguistics presents law itself as primarily a text with predetermined meanings that can be deduced from its relationship to surrounding text in a confined data space. Agency interpretation treats legal meaning-making as a social event in a data-diversified environment, inherently subject to development and change. Perhaps most strikingly, legal corpus linguistics presents legal texts as always already having whatever meanings they were meant to have. Agency interpretation, in contrast, treats legal text as a social phenomenon that must be given meaning through the discursive interaction of many individuals, interests, viewpoints, and preferences. It imagines law as a social force enacted through social practices.

Of course, judges do not have the personnel or the authority to undertake the kind of ongoing research, debate, and decision-making that agencies do. Yet choosing a different model for legal interpretation could affect judicial interpretive choices. Rather than seeking the real meaning of a word in a database, judges can consider how the legal text has developed, what goals it furthers, and what effects different interpretations would have on the world. These kinds of considerations are not foreign to legal interpretation; on the contrary, they permeate the work of agencies, our primary interpreters
of law. Of course, not every agency interpretation process takes every perspective into account or fulfills its promise of rationality. But the model of interpretation, our image of what prototypical interpretation looks like, makes a difference in how we choose and evaluate the technologies available for it.

Those turning to legal corpus linguistics may strive to emulate the kind of empirical, grounded, and realistic approach to meaning-making that agencies have developed. Yet choosing which technology to favor also commits us, to some extent, to the premises that technology builds in. It is thus worth asking what image of legal process rings most true, and most desirable, as we evaluate our options. Looking for the meaning of law in words isolated from their broader legal context, as legal corpus work does, seems to evacuate the law-ness from law. In contrast, understanding that legal meaning comes about through the interplay of collaborating and conflicting social forces comports better to our experience of law as it wends its way through iterative judicial decisions and ongoing administrative procedures—not to mention as it is negotiated on the ground in discretionary interactions and debated in various forums. Viewing judicial legal interpretation in a broader setting can thus provide insight into the range of things legal interpretation can be. It can also raise productive questions about what legal interpretation should be, pushing commentators out of easy assumptions based on current judicial practice and ideology.

CONCLUSION

Legal corpus linguistics, a newly popular technology of legal interpretation, promises a quick, simple answer to the puzzle of legal meaning. It is worth noting what a very odd promise this is. The social and political processes that go into producing, enforcing, litigating, and understanding the law are extremely complex. Why would a quick, simple answer be the best fit for that complexity? Rather than looking for clear, ultimate conclusions, we should prefer technologies of legal meaning-making that take into account law’s inherent complexity.

126. See Anya Bernstein, Differentiating Deference, 33 YALE J. REG. 1, 52 (2016) (arguing that courts reviewing agency action should take into account whether the agency utilized its competence rather than assuming that agencies will always use all the competence they potentially have at their disposal).
Taking that complexity into account might mean giving up the notion that legal meaning inheres in individual words or small word clumps whose meaning does not change over time. But that notion is not necessary to the project of legal interpretation. Rather, it is a preference of some specific normative approaches to law, most forcefully expressed in the theories of textualism and originalism. These theories may be attractive because of the view of law, politics, or normative order that they present. But we should be clear that they do not describe some quality of language or communication that inheres in legal meaning. Rather, they express a normative position about how we ought to go about giving meaning to law. Just so, legal corpus linguistics, for all its reliance on databases, does not tell us what legal language really means. It too expresses a normative position about how we ought to go about giving meaning to law.

As Latour argues, truths—that is, those things that come to look like truths—emerge from complex, multi-participant processes of messy clash and cooperation. It is the participants of communicative events who actively create meanings for linguistic utterances. Legal language benefits from being involved in many such meaning-creating contexts. From legal drafting and enactment to implementation to popularization to adjudication, the process of giving meaning to law involves hundreds, sometimes thousands, of people. In some sense, legal corpus linguistics, with its penchant for large data sets with millions of exemplars, seems like a natural fit for analyzing the kind of complex, multilateral social processes that go into the creation of legal meaning.

Yet, as currently practiced, legal corpus linguistics instead brackets the social practices that actually give meaning to law in our society. It looks to words that live in unrelated social worlds—novels, newspapers, radio shows, and so on. In doing so, it ignores both the contexts and the participants that give law meaning, and it treats legal language as somehow inherently meaningful. If we see law not as a collection of separable words with a clear right answer but as an actant whose effect on the world requires the ongoing participation of different groups—those who enforce, interpret, obey, obstruct, and so on—we may be more attracted to meaning-making technologies that place law in more relational and practical contexts. We often look to technology to simplify our work. But sometimes we may be better served by technologies that recognize its true complexity.