August 2017

Corpus Linguistics and the Criminal Law

Carissa Byrne Hessick

Follow this and additional works at: https://digitalcommons.law.byu.edu/lawreview

 Part of the Applied Linguistics Commons, and the Criminal Law Commons

Corpus Linguistics and the Criminal Law

Carissa Byrne Hessick[*]

This brief response to Ordinary Meaning and Corpus Linguistics, *an article by Stefan Gries and Brian Slocum, explains why corpus linguistics represents a radical break from current statutory interpretation practice, and it argues that corpus linguistics ought not be adopted as an interpretive theory for criminal laws. Corpus linguistics has superficial appeal because it promises to increase predictability and to decrease the role of judges' personal preferences in statutory interpretation. But there are reasons to doubt that corpus linguistics can achieve these goals. More importantly, corpus linguistics sacrifices other, more important values, including notice and accountability.*

CONTENTS

1503

INTRODUCTION

When read in isolation, *Ordinary Meaning and Corpus Linguistics*, an article by Stefan Gries and Brian Slocum,[1] appears quite modest. It examines two prominent criminal law cases, and it uses a linguistics database to demonstrate how the judicial interpretations in those cases about the "ordinary meaning" of two statutory terms do not align with the most frequently used meanings of the terms.

But the Gries and Slocum article should not be read in isolation. It should be read as part of a broader attempt to change modern statutory interpretation. A small group of legal scholars (and an even smaller group of judges) are advancing a new interpretive theory called "corpus linguistics." Corpus linguists advocate treating the "ordinary meaning" inquiry in statutory interpretation as an empirical question: the ordinary meaning should be ascertained by consulting a linguistics database to determine how frequently a term is used in a certain manner.[2]

---

1. Stefan Th. Gries & Brian G. Slocum, *Ordinary Meaning and Corpus Linguistics*, 2017 BYU L. REV 1417.

2. The leading proponents of corpus linguistics as an interpretive theory are Thomas Lee, a justice on the Utah Supreme Court, and Stephen Mouritsen. *See, e.g.*, Justice Thomas Lee & Stephen Mouritsen, *The Path Forward for Law and Corpus Linguistics*, WASH. POST (Aug. 11, 2017), https://www.washingtonpost.com/news/volokh-conspiracy/wp/2017/08 /11/the-path-forward-for-law-and-corpus-linguistics/?utm_term=.0d9ec9a77315 [hereinafter Lee & Mouritsen, *The Path Forward*] ("[W]e see corpus linguistic analysis playing a central role in legal interpretation going forward. When problems of ambiguity in legal language arise, we expect judges and lawyers to turn increasingly to data from linguistic databases and less to mere dictionaries. Corpus linguistics will form an important part of the future of legal interpretation."); Justice Thomas Lee & Stephen Mouritsen, *Linguistic Corpora and the Evolving Study of Language and Meaning*, WASH. POST (Aug. 9, 2017), https://www. washingtonpost.com/news/volokh-conspiracy/wp/2017/08/09/linguistic-corpora-and-the-evolving-study-of-language-and-meaning/?utm_term=.f56cb676f54b [hereinafter Lee & Mouritsen, *Evolving Study*] (arguing that "*a complete theory* of ordinary meaning requires us to take into account the comparative frequency of different senses of words, the (syntactic, semantic and pragmatic) context of an utterance, its historical usage and the speech community in which it was uttered" (emphasis added)); *see also* Stephen C. Mouritsen, *The Dictionary is Not a Fortress: Definitional Fallacies and a Corpus-Based Approach to Plain Meaning*, 2010 BYU L. REV. 1915, 1953–54 (describing an argument to determine the "ordinary meaning" of a statutory term by reference to "common usage" as "a theory of ordinary meaning").

This brief response explains why corpus linguistics represents a radical break from current interpretive theories. It also argues that corpus linguistics ought not be adopted as an interpretive theory for criminal laws.[3] Corpus linguistics has superficial appeal because it promises to increase predictability and to decrease the role of judges' personal preferences in statutory interpretation. But there are reasons to doubt that corpus linguistics can achieve these goals. More importantly, in attempting to achieve these goals, corpus linguistics sacrifices other, more important values, including notice and accountability.

I. "ORDINARY MEANING" AND EMPIRICISM

It is easy to overlook that corpus linguistics is an interpretive theory, rather than simply an interdisciplinary methodology, because it bills itself as providing an answer to a question that many current interpretive theories ask: What is the "plain" or "ordinary" meaning of the statutory text?[4] Indeed, some who advocate for the use of

---

But there are others who have written about corpus linguistics who have advocated its use as an interdisciplinary tool, rather than as a legal theory. *See, e.g.*, Lawrence M. Solan & Tammy Gales, *Corpus Linguistics as a Tool in Legal Interpretation*, 2017 BYU L. REV. 1311, 1356 ("At all times, it should be kept in mind that linguistic corpora provide a tool for those engaged in statutory interpretation, but they say nothing about when this tool is most useful. That judgment must come from a combination of legal decisions about whether the distribution of word usage should determine the outcome of a case, a further legal decision about what 'ordinary meaning' should mean in a legal context, and appropriate use of a corpus when one is used."); Lawrence B. Solum, *Originalist Methodology*, 84 U. CHI. L. REV. 269, 285 (2017) (identifying corpus linguistics as one of a number of tools that can be used to "check" the meanings given by others).

    3. I limit my arguments in this Article to the appropriateness of using corpus linguistics to interpret criminal laws. One of my major criticisms of corpus linguistics is its impact on notice, and the need for notice is particularly acute when it comes to criminal laws. *See* Paul H. Robinson, *Fair Notice and Fair Adjudication: Two Kinds of Legality*, 154 U. PA. L. REV. 335, 347–48 (2005) (explaining that concerns about punishment without notice have resulted in "special rules for the construction of penal statutes"). But not all of the objections that I raise are necessarily limited to criminal laws.

    4. Different theories ask this question for different purposes. For textualists, the plain meaning may control. *See* William Baude & Ryan D. Doerfler, *The (Not So) Plain Meaning Rule*, 84 U. CHI. L. REV. 539, 541 (2017). For purposivists, the plain meaning serves as a limit on possible interpretations. *See* HENRY M. HART, JR. & ALBERT M. SACKS, THE LEGAL PROCESS: BASIC PROBLEMS IN THE MAKING AND APPLICATION OF LAW 112 (William N. Eskridge, Jr. & Philip P. Frickey eds., 1994).

corpus linguistics describe it as a tool that can be used by textualists. And I agree that corpus linguistics may be a useful tool for identifying permissible meanings of a statutory term. But not everyone who promotes corpus linguistics in the law does so as a mere tool for identifying permissible meanings. Some advocate the use of corpus linguistics as a new theory about how statutes ought to be interpreted.[5]

To understand how corpus linguistics functions as an interpretive theory, it is helpful to contrast it with textualism.[6] Textualism is a method of statutory interpretation in which statutes are construed, at least initially, according to the "plain" or "ordinary" meaning of the text.[7] Textualists will consider other sources—such as legislative history, policy considerations, practice, or canons of statutory construction—only if the meaning of the statutory language cannot be ascertained from the text itself.[8] In other words, textualists will look first at the ordinary meaning of a statute, and they will not consult nontextual sources if the meaning of the statute is apparent from the text itself.

Corpus linguistics reframes the "plain" or "ordinary" meaning inquiry in two ways. First, it claims that ordinary meaning is an empirical question. Second, it tells us that this empirical question ought to be answered by how frequently a term is used in a particular way. Both of these analytical moves represent significant departures from current theories of statutory interpretation, including textualism, and they render statutory interpretation essentially unrecognizable.

---

5. *See supra* note 2.

6. Corpus linguistics is often discussed in terms of how it compliments textualism. *See, e.g.*, *Corpus Linguistics and Legal Interpretation–Event Audio/Video*, FEDERALIST SOC'Y (Jan. 23, 2017) (statement of Justice Thomas Lee at 23:20–25:00), https://fedsoc.org/commentary/videos/corpus-linguistics-and-legal-interpretation-event-audio-video?autoplay=1 (discussing how corpus linguistics allows the law to deliver on the promise of textualism).

7. *See* Baude & Doerfler, *supra* note 4, at 543–44.

8. *See id.*

Courts do not usually treat ordinary meaning as an empirical question.[9] Sometimes courts use the term "ordinary meaning" to refer to whether a meaning is permitted,[10] sometimes to refer to whether the meaning is obvious,[11] and sometimes to refer to the meaning that the hypothetical reasonable person would give to the statutory language.[12] None of these usages is empirical—or at least they are not readily quantifiable. Nor does the academic literature on textualism treat ordinary meaning as an empirical question subject to quantification.[13]

Gries and Slocum are actually quite careful about the empiricism of corpus linguistics. They make clear that "the proper standard for designating some permissible meaning as the ordinary meaning" of a statutory term is a question for the courts, not a question that corpus analysis itself can answer.[14] And they take pains not to assume that, when judges seek the ordinary meaning of a term, the answer they seek is necessarily empirical.[15]

---

9. A fact that corpus linguists admit. *See, e.g.*, Stephen C. Mouritsen, *Hard Cases and Hard Data: Assessing Corpus Linguistics as an Empirical Path to Plain Meaning*, 13 COLUM. SCI. & TECH. L. REV. 156, 163–80 (2011) [hereinafter Mouritsen, *Hard Cases*] (describing the various nonempirical approaches that courts have taken regarding the "plain" or "ordinary meaning" question over time).

10. *See, e.g.*, Smith v. United States, 508 U.S. 223, 230–31 (1993).

11. *See, e.g.*, United States v. Ron Pair Enters., Inc., 489 U.S. 235, 242 (1989).

12. *See, e.g.*, Bond v. United States, 134 S. Ct. 2077, 2090–91 (2014).

13. *See* Richard H. Fallon Jr., *The Meaning of Legal "Meaning" and Its Implications for Theories of Legal Interpretation*, 82 U. CHI. L. REV. 1235, 1244–45 (2015) ("In debates about legal meaning and interpretation, participants' references to legal meaning sometimes invoke or appeal to each of the following: (1) semantic or literal meaning; (2) contextual meaning as framed by shared presuppositions of speakers and listeners, including shared presuppositions about application and nonapplication; (3) real conceptual meaning; (4) intended meaning; (5) reasonable meaning; and (6) interpreted meaning."); *id.* at 1239 (noting that "proponents of leading interpretive theories sometimes appear to" "question whether statutory and constitutional provisions have uniquely correct meanings that exist as a matter of prelegal, linguistic fact").

14. Gries & Slocum, *supra* note 1, at 1470.

15. *E.g., id.* at 1434–35 (observing that in deciding the ordinary meaning question the "question *might be framed* as empirical in nature and based on an accurate understanding of language" (emphasis added)); *id.* at 1435 ("*To the extent that ordinary meaning is an empirical question*, it would seem that courts and others should consider the possibility that other determinants might be better suited to the task." (emphasis added)). The closest that Gries and Slocum come to endorsing the idea that ordinary meaning ought to be treated as an

Gries and Slocum's modesty about whether textualism's quest for ordinary meaning is an empirical question is not shared by all corpus linguists. The first law review article to promote corpus linguistics in legal interpretation claimed that ordinary meaning ought to be treated as an empirical question—in particular a question about frequency.[16] This claim has been repeated multiple times since then.[17] But corpus linguists' explanation about why "ordinary meaning" ought to be the most frequently used meaning is not very persuasive. Proponents of corpus linguistics believe that the frequency with which words are used in various nonlegal publications can tell us how a member of the public would understand those words when they appear in statutes.[18] While there are good reasons for courts to interpret statutes—especially criminal statutes—in a fashion that is consistent with how an ordinary member of the public would understand the statute,[19] the frequency with which a term is used does not give us that information.

To be clear, corpus linguists think that the context in which a statutory term is used also matters in determining "ordinary meaning."[20] But that is not what corpus linguistics contributes to

empirical question is when they state that "[ordinary meaning] would seem to be subject in some way to the empirical verification that corpus analysis can provide." *Id.* at 1470.

16. Mouritsen, *supra* note 2, at 1953–54. Mouritsen relies, in part, on a dictionary definition to argue that "ordinary" ought to be interpreted to mean most commonly used. (This reliance on the dictionary seems almost ironic given that the article is devoted to criticizing dictionaries as a source for legal meaning.)

17. *E.g.*, Thomas R. Lee & Stephen C. Mouritsen, *Judging Ordinary Meaning*, 127 YALE L.J. (forthcoming 2017) (manuscript at 5) (available at https://www.ssrn.com/abstract =2937468) [hereinafter Lee & Mouritsen, *Judging Ordinary Meaning*] ("When we speak of *ordinary* meaning we are asking an empirical question—about the sense of a word or phrase that is most likely implicated in a given linguistic context."); Lee & Mouritsen, *Evolving Study*, *supra* note 2 ("[A] complete theory of ordinary meaning requires us to take into account the comparative frequency of different senses of words, the (syntactic, semantic and pragmatic) context of an utterance, its historical usage and the speech community in which it was uttered.").

18. *E.g.*, Lee & Mouritsen, *Judging Ordinary Meaning*, *supra* note 17, at 24.

19. These reasons include notice and accountability. *See infra* text accompanying notes 48–49.

20. *See, e.g.*, Lee & Mouritsen, *Judging Ordinary Meaning*, *supra* note 17.

statutory interpretation—everyone thinks that context matters.[21] What corpus linguistics adds to statutory interpretation is the idea that the ordinary meaning ought to be determined according to the frequency with which a word is used a particular way.

How often a term appears in newspapers, magazines, or other publications is a separate inquiry from how members of the public would understand that term when used in a statute. Imagine, for example, a dispute over the scope of a statute that provides relief for flood victims. The dispute centers around how much water must have accumulated in an area in order for an event to be considered a flood. A database search of how often the word "flood" is used to refer to very large amounts of accumulated water will doubtlessly be skewed by the fact that instances of extensive flooding—such as New Orleans after Hurricane Katrina in 2005 or Houston during Hurricane Harvey in 2017—will receive far more media coverage than other events. Indeed, a corpus analysis may demonstrate that seventy percent of all mentions of the word "flood" occur in the context of these superstorms. But that does not tell us whether the average American would understand the statutory term "flood" to include three inches of water in a homeowner's basement after a neighboring water main burst.[22]

As the flood example illustrates, the frequency with which a word appears in print is separate from how an ordinary citizen would understand that word. As Oliver Wendell Holmes explained, when a criminal statute uses a word or a phrase, that word or phrase "calls up the picture" of something—and it is that picture that the law

---

21. *See, e.g.*, FAA v. Cooper, 566 U.S. 284, 308 (2012) ("Context, of course, is relevant to statutory interpretation; it may provide clues that Congress did not employ a word or phrase in its ordinary meaning."); Nken v. Holder, 556 U.S. 418, 426 (2009) (stating that "statutory interpretation turns on 'the language itself, the specific context in which that language is used, and the broader context of the statute as a whole'" (quoting Robinson v. Shell Oil Co., 519 U.S. 337, 341 (1997))).

22. Lawrence Solan and Tammy Gales make a similar point in their contribution to this issue—that corpus data may reflect the fact that a given sense of a certain term is a more factually common iteration of that term in the real world. Solan & Gales, *supra* note 2, at 1314, 1356. My point is not simply about real world occurrence, but also about how often an event receives media coverage.

should be interpreted to mean.[23] This mental picture is what some call a "prototype."[24] As Gries and Slocum tell us, corpus linguistics "can provide clues" about prototypes.[25] But, as the flood example demonstrates, sometimes the "clues" that corpus linguistics provides will point us in the wrong direction.[26]

Although the Supreme Court has been far from consistent in its use of the term "ordinary meaning," textualists often use the term to refer to the meaning that is intended by a reasonable speaker or the meaning that is understood by a reasonable listener.[27] In other words, textualism's use of "ordinary meaning" is intended to answer the prototype question. But corpus linguists do not view the reasonable person as an appropriate alternative to empirical questions of frequency. Instead they denounce the reasonable person standard as nothing more than "judicial intuition,"[28] which is insufficiently

---

23.   McBoyle v. United States, 283 U.S. 25, 26–27 (1931).

24.   *See* Solan & Gales, *supra* note 2, at 1354 (explaining why the giraffe is the prototypical tall animal).

25.   Gries & Slocum, *supra* note 1, at 1441 ("[Corpus linguistics] can provide clues as to what the most prototypical meaning of an expression might be based on various factors (e.g., highest frequency, most even dispersion, earliest acquired, central in network of senses, meaning with the highest number of features with the highest cue validities).")

26.   At times, corpus linguists appear to acknowledge that frequency analysis does not necessarily identify a linguistic prototype. *See, e.g.*, Lee & Mouritsen, *Judging Ordinary Meaning*, *supra* note 17, at 10–11 (treating a linguistic prototype as a different "sense" of the term "ordinary meaning" than the frequency with which a term is used); Mouritsen, *Hard Cases*, *supra* note 9, at 181–201 (treating linguistic prototypes and corpus frequency analyses as separate inquiries).

27.   For example, Judge Easterbrook has said:
We should look at the statutory structure and hear the words as they would sound in the mind of a skilled, objectively reasonable user of words. Words appeal to the reasonable man of tort law; private language and subjective intents should be put aside. The meaning of statutes is to be found not in the subjective, multiple mind of Congress but in the understanding of the objectively reasonable person.
Frank H. Easterbrook, *The Role of Original Intent in Statutory Construction*, 11 HARV. J.L. & PUB. POL'Y 59, 65 (1988). Similarly, Justice Scalia referred to the interpretive process as the search for "the intent that a reasonable person would gather from the text of the law." ANTONIN SCALIA, A MATTER OF INTERPRETATION: FEDERAL COURTS AND THE LAW 17 (1997).

28.   *See* Lee & Mouritsen, *Judging Ordinary Meaning*, *supra* note 17, at 15; Mouritsen, *Hard Cases*, *supra* note 9, at 161, 176.

objective[29] and does not result in predictable outcomes.[30] Some corpus linguists have been quite clear that they prefer corpus linguistics to the reasonable person standard because corpus linguistics minimizes judicial discretion.[31] In particular, these corpus linguists object to the reasonable person standard because the standard is "idealized" and therefore it "has little or nothing to do with the actual meaning intended by a legislator or understood by the public."[32]

It is true that judges will not always be able to accurately determine what the majority of Americans believe is the linguistic prototype of a statutory term. Instead, judges will likely identify their own linguistic prototypes. And it is also true that judges will—and do—disagree about the ordinary meaning of statutory terms.[33] But the idea that judges are not competent to say what a statute means is a radical—and in my mind dangerous—doctrinal path to take. The idea that the traditional ordinary meaning inquiry is somehow inferior to a frequency analysis because judges might rely on their intuition—that is, their professional judgment—rejects the very foundation of the judicial role.

The Constitution assigns "the judicial power" to the courts. The judicial power has been understood for centuries to include the

---

29. *See* Mouritsen, *Hard Cases*, *supra* note 9, at 174–80; *see also* Lee & Mouritsen, *Judging Ordinary Meaning*, *supra* note 17, at 5 (stating that "empirical methods utilized by linguists . . . may help us do a better job of delivering on the promise of an objective inquiry into ordinary meaning").

30. *See* Mouritsen, *Hard Cases*, *supra* note 9, at 205 (stating that corpus linguistics "improve[s] the predictability and consistency of judicial decision making").

31. *See, e.g.*, *Corpus Linguistics and Legal Interpretation–Event Audio/Video*, FEDERALIST SOC'Y (Jan. 23, 2017) (statement of Justice Thomas Lee at 24:10), https://fedsoc.org/commentary/videos/corpus-linguistics-and-legal-interpretation-event-audio-video?autoplay=1 (praising corpus linguistics because it minimizes judicial discretion).

32. Lee & Mouritsen, *Judging Ordinary Meaning*, *supra* note 17, at 21–22. In an earlier article, Mouritsen expressed concern about judges employing the reasonable person test because they are unlikely to be very good at estimating how often a term is used a particular way. Mouritsen, *Hard Cases*, *supra* note 9, at 175–76. But that concern assumes that frequency is an appropriate measure for original meaning.

33. *See generally* Ward Farnsworth, Dustin F. Guzior & Anup Malani, *Ambiguity About Ambiguity: An Empirical Inquiry into Legal Interpretation*, 2 J. LEGAL ANALYSIS 257 (2010) (showing disagreement over whether text is ambiguous).

1

power to interpret laws.[34] And the power to interpret statutes has always been understood to require more than simply reading the text.[35] In recent years, judges have often used their interpretive power to read statutes as they would be understood by the public. But in so doing, judges have not doubted their authority to rely on their own judgment about what the public understanding is likely to be. Nor have they shied away from using interpretation to further other goals when they believed those goals were warranted.[36]

Seeking to prevent judges from relying on their own judgment in statutory interpretation is, in a very real sense, a rejection of the judicial power to interpret the laws. Judges have long relied on their professional judgment when interpreting laws. Oliver Wendell Holmes endorsed this role of judgment in interpretation when he said: "Behind the logical form [of statutory interpretation] lies a

34. *See, e.g.*, 5 MATHEW BACON, A NEW ABRIDGMENT OF THE LAW 217 (3d ed., London, J. Worrall & Co. 1770) ("It is the [p]rovince of the [j]ustices to determine, what the [m]eaning of a [w]ord or [s]entence in an [a]ct of Parliament is."); THE FEDERALIST NO. 78, at 381 (Alexander Hamilton) ("The interpretation of the laws is the proper and peculiar province of the courts."); 1 THE RECORDS OF THE FEDERAL CONVENTION OF 1787, at 98 (Yale Univ. Press 1966) (statement of Rufus King) ("Judges ought to be able to expound the law as it should come before them"); *see also* Philip Hamburger, *A Tale of Two Paradigms: Judicial Review and Judicial Duty*, 78 GEO. WASH. L. REV. 1162, 1171–72 (2010) (arguing that courts historically exercised independent judgment in interpretation); Robert J. Pushaw, Jr., *A Neo-Federalist Analysis of Federal Question Jurisdiction*, 95 CALIF. L. REV. 1515, 1542 (2007) ("By 1787, 'judicial power' had acquired a core meaning that has lasted to this day: rendering a binding judgment after impartially interpreting and applying the law in light of the facts presented in litigation.").

35. *See, e.g.*, Comm'r v. Ickelheimer, 132 F.2d 660, 662 (2d Cir. 1943) (Hand, J., dissenting) ("Compunctions about judicial legislation are right enough as long as we have any genuine doubt as to the breadth of the legislature's intent; and no doubt the most important single factor in ascertaining its intent is the words it employs. But the colloquial words of a statute have not the fixed and artificial content of scientific symbols; they have a penumbra, a dim fringe, a connotation, for they express an attitude of will, into which it is our duty to penetrate and which we must enforce ungrudgingly when we can ascertain it, regardless of imprecision in its expression."); *Nine Speeches by Justice Roger J. Traynor*, *in* 8 CALIFORNIA LEGAL HISTORY 211, 218 (Selma Moidel Smith ed., 2013) ("'The law is not an encyclopedia to which lawyers may rush,' . . . but rather, it thrives on 'conflict and fresh interpretation.'").

36. For example, the courts will construe statutes to avoid constitutional doubts, Clark v. Martinez, 543 U.S. 371, 380–82 (2005), promote federalism, BFP v. Resolution Tr. Corp., 511 U.S. 531, 544–45 (1994), and achieve a series of other goals, *see* William N. Eskridge, Jr. & Philip P. Frickey, *Quasi-Constitutional Law: Clear Statement Rules as Constitutional Lawmaking*, 45 VAND. L. REV. 593, 598–611 (1992) (cataloguing various interpretive goals).

judgment as to the relative worth and importance of competing legislative grounds, often an inarticulate and unconscious judgment, it is true, and yet the very root and nerve of the whole proceeding."[37] Similarly, Felix Frankfurter said that "the final rendering of the meaning of a statute is an act of judgment."[38]

Of course, disagreements have arisen over *how* judges ought to interpret laws. There are differing opinions over, for example, what sources of meaning judges ought to consult and whether judges ought to consider the policy implications of their interpretive choices.[39] But so long as judges are declaring what they believe the statute says—as opposed to declaring what they believe it ought to say—they are well within their constitutional sphere.

One suspects that this negative attitude towards judicial judgment is related to the omnipresent concern in modern legal thought about the counter-majoritarian difficulty—that judges might use their interpretive powers to usurp the policy powers of the elected branches.[40] Is it possible that judges will allow their personal preferences to color their interpretations of laws? Of course it is. But that possibility can be—and is—already guarded against in the system through various accountability mechanisms. The practice of opinion writing creates accountability because it requires judges to explain their decisions.[41] The norm of respecting precedent provides further accountability because judges know that their decisions will have effects beyond the particular case.[42] Because judges remain accountable for their decisions, they are more likely to decide cases

---

37. Oliver Wendell Holmes Jr., *The Path of the Law*, 10 HARV. L. REV. 457, 466 (1897).

38. Felix Frankfurter, *Some Reflections on the Reading of Statutes*, 47 COLUM. L. REV. 527, 531 (1947).

39. *See* Fallon, *supra* note 13, at 1287–88.

40. ALEXANDER M. BICKEL, THE LEAST DANGEROUS BRANCH: THE SUPREME COURT AT THE BAR OF POLITICS 16–23 (2d ed. 1986).

41. *See* Louis Michael Seidman, *Ambivalence and Accountability*, 61 S. CAL. L. REV. 1571, 1574 (1988) (discussing these two types of accountability).

42. *See* William L. Reynolds & William M. Richman, *The Non-Precedential Precedent– Limited Publication and No-Citation Rules in the United States Courts of Appeals*, 78 COLUM. L. REV. 1167, 1204 (1978) (explaining how the precedential value of an opinion increases judicial accountability).

in a disinterested fashion. And by both explaining why they interpreted a statute a particular way and by being bound by that decision in future cases, judges ensure a significant amount of predictability in the law. Even if we want to further maximize predictability and minimize the possibility that interpretation will be driven by idiosyncratic judicial preferences, there are several reasons why performing a frequency analysis using a corpus linguistics database is not the path to take.

II. THE PROBLEMS WITH FREQUENCY

Corpus linguistics tells us that the ordinary meaning of a statutory term ought to be resolved by looking to the frequency with which a term is used a certain way.[43] This is a problematic theory for the interpretation of criminal laws because it creates problems of notice and accountability.

Notice is one of the bedrock principles of modern criminal law.[44] Criminal laws must give us fair notice of what conduct is permitted and what conduct is prohibited.[45] Without such notice, an individual may accidentally engage in illegal conduct,[46] or she may choose to

---

43. To be clear, corpus linguists do not look at frequency only. They say that "a complete theory of ordinary meaning requires us to take into account the comparative frequency of different senses of words, the (syntactic, semantic and pragmatic) context of an utterance, its historical usage and the speech community in which it was uttered." Lee & Mouritsen, *Evolving Study*, *supra* note 2. As noted above, everyone agrees that context is important. *See supra* text accompanying note 20. To the extent that corpus linguists constrain their frequency analyses according to "its historical usage and the speech community in which [a term] was uttered," that does not alleviate the notice and accountability concerns addressed in this section. Lee & Mouritsen, *Evolving Study*, *supra* note 2.

44. *See* Robinson, *supra* note 3, at 364.

45. Laws that fail to give such notice are frequently invalidated under the vagueness doctrine. *E.g.*, Smith v. Goguen, 415 U.S. 566, 572 (1974) ("The [vagueness] doctrine incorporates notions of fair notice or warning.").

46. *See* Grayned v. City of Rockford, 408 U.S. 104, 108 (1972) ("[B]ecause we assume that man is free to steer between lawful and unlawful conduct, we insist that laws give the person of ordinary intelligence a reasonable opportunity to know what is prohibited, so that he may act accordingly. Vague laws may trap the innocent by not providing fair warning." (footnote omitted)).

avoid large swaths of legal conduct because she does not know for certain whether such conduct is permitted.[47]

Corpus linguistics is not well designed to provide notice because the meaning of a statute depends on a database search and an analysis of those results. When it comes to notice, we care about the notice that is provided to members of the general public.[48] But members of the general public cannot be expected to perform their own corpus searches and analyses. The process described in the corpus linguistics literature appears quite involved, and it hardly seems accessible to the average American.[49]

Indeed, there is a debate within the corpus linguistics field over whether attorneys and judges are qualified to perform corpus searches and analyses.[50] Even those who believe that attorneys and judges are able to perform corpus searches and analyses "concede that corpus linguistics is not 'plug and play' analysis" and that "[c]orpus data can be gathered and analyzed properly only with care and a little background and training in the underlying methodology."[51] But they nonetheless conclude that attorneys and judges will be able to effectively use corpus linguistics because of their legal training.[52]

This debate over the competence of attorneys and judges makes clear that the average American would be unable to conduct an

---

47. *See* United States v. Nat'l Dairy Prods. Corp., 372 U.S. 29, 36 (1963) ("[W]e are concerned with the vagueness of the statute 'on its face' because such vagueness may in itself deter constitutionally protected and socially desirable conduct." (citing Thornhill v. Alabama, 310 U.S. 88, 98 (1940))).

48. *See, e.g.*, Johnson v. United States, 135 S. Ct. 2551, 2556 (2015) (stating that the Due Process Clause is violated if the government deprives a person of "life, liberty, or property under a criminal law so vague that it fails to give ordinary people fair notice of the conduct it punishes").

49. For an example of how one conducts such a search and the resulting analysis of *one thousand* randomly returned results, see Mouritsen, *Hard Cases*, *supra* note 9, at 194–201.

50. For example, Gries and Slocum do not believe that attorneys or judges are qualified. Instead, Gries and Slocum advocate the involvement of expert linguists during litigation. Gries & Slocum, *supra* note 1, at 1469–71. In contrast, Lee and Mouritsen argue that attorneys and judges are competent to conduct these analyses. Though they agree that experts will improve judicial corpus analysis, they do not think that linguistics experts are necessary. Lee & Mouritsen, *Judging Ordinary Meaning*, *supra* note 17, at 69–72.

51. Lee & Mouritsen, *Judging Ordinary Meaning*, *supra* note 17, at 69.

52. *Id.* at 70–71.

appropriate search or analysis on her own. Most Americans do not have legal training, nor do they have any background or training in linguistic methodology. Thus, even corpus linguistics' most ardent defenders must concede that the general public cannot competently use corpus linguistics in order to determine the meaning of a statutory term. And if the public does not know the meaning of a term in a criminal statute, then the statute does not give sufficient notice to satisfy due process.[53]

In addition to failing to provide sufficient notice, the corpus methodology creates accountability concerns. If the meaning of a statute depends on a database search and an analysis of those results, then how will legislators know the scope of the laws that they enact and how will voters hold legislators responsible for their actions? Perhaps legislators could have their staff perform corpus searches and analyses, which would allow them to predict how courts will interpret the legislation that they enact. But unless voters conduct the same searches and analyses, they will not be able to hold their representatives accountable for the legislation that is passed.

One might respond to these notice and accountability concerns by saying that members of the general public and legislators need not conduct their own corpus analyses because the frequency of usage reflected in the corpus database is likely to correspond to their own intuitions about the meaning of language.[54] But this is at odds with a fundamental premise of corpus linguistics—that *judicial* intuition ought to be replaced with corpus analyses precisely because that intuition is unreliable.[55] We have no reason to think that nonjudicial intuition will be more reliable.[56] Also, the response either

---

53. *See supra* notes 44–47 and accompanying text.
54. Indeed, corpus linguists do not appear to contemplate members of the public performing these searches. *See, e.g.*, Lee & Mouritsen, *Judging Ordinary Meaning*, *supra* note 17, at 69–72 (arguing that judges and lawyers have the necessary skills to perform competent corpus searches and analyses).
55. *See, e.g.*, Mouritsen, *Hard Cases*, *supra* note 9, at 175–78.
56. *See* Lawrence Solan, Terri Rosenblatt & Daniel Osherson, *False Consensus Bias in Contract Interpretation*, 108 COLUM. L. REV. 1268, 1285–94 (2008) (demonstrating that both judges and laypeople overestimate whether others would agree with their interpretation of language).

proves too much or too little. If the results will always be the same as intuition, then corpus linguistics is unnecessary. If there are cases where corpus linguistics returns a different result than would the intuitions of lawmakers, judges, or voters, then corpus linguistics is a real threat to notice and accountability.

Legislative accountability is not the only accountability problem posed by corpus linguistics. Telling judges to rely on frequency analyses rather than their professional judgment also creates an accountability deficit for judges. Corpus linguistics is attractive, in part, because it taps into our respect for science and data. Because human judgment is opaque, contested, and sometimes messy, when a person can point to objective facts to justify a decision, they will often do so.[57] Those facts ease some of the pressure to justify an otherwise contestable decision.[58] But as Kate Stith and José Cabranes have explained, justice may only be meted out by people. "By replacing the case-by-case exercise of human judgment with a mechanical calculus, we do not judge better or more objectively, nor do we judge worse. Instead, we cease to judge at all."[59]

Corpus linguists appear to assume that the personal judgment of judges is necessarily arbitrary, and thus indefensible from a rule of law perspective.[60] But that is not so. Judgment is not "a matter of subjective 'feeling.'"[61] Judicial decisions have to be explained and justified. And judges know that they have to grapple with their past decisions in future cases. That is why, when judges interpret the meaning of a statute, they do not simply write "because I say so" in their opinions.

---

57. *Cf.* Dan. M. Kahan, *The Cognitively Illiberal State*, 60 STAN. L. REV. 115, 143-45 (2007) (identifying (and criticizing) the tendency in liberal democracies to conduct policy debates by appealing to "objective procedures and criteria," such as cost-benefit analysis, rather than to criteria related to morals or virtue).

58. *Cf.* Fallon, *supra* note 13, at 1306 (noting that "any interpretive theory that requires normatively based case-by-case judgments cannot escape moral controversy").

59. KATE STITH & JOSÉ A. CABRANES, FEAR OF JUDGING: SENTENCING GUIDELINES IN THE FEDERAL COURTS 82 (1998).

60. *See supra* notes 28–30.

61. STITH & CABRANES, *supra* note 59, at 82.

But if we adopt corpus linguistics, then judges could replace their explanations and justifications with a short "because the corpus says so." If judges' interpretive task is reframed as conducting corpus searches or acting as a referee between competing experts, then judges will not bear the same responsibility to explain why and how a particular interpretation makes sense.

Our constitutional system assigns judges the responsibility of interpreting laws for a reason: the Constitution divides authority between the three branches of government in order to promote individual liberty.[62] Especially in the criminal law context, the separation of powers is assumed to promote individual liberty because all three branches must agree that conduct is deserving of punishment in order for a defendant to be convicted. If the legislature does not believe certain conduct should be punished, then it will exclude the conduct from its definition of crimes. If prosecutors do not believe that the particular circumstances of a case warrant punishment, then they will decline to bring charges in that case. And the judiciary will dismiss charges brought by prosecutors unless the defendants' conduct is clearly prohibited by the criminal statute.[63]

Because the separation of powers assumes that statutory interpretation is not merely a ministerial task, but rather one of several checks on the state's power, judges must take responsibility for their decisions. Corpus linguistics mutes that constitutional role. It allows judges to see their role as counting database returns or evaluating expert credentials, rather than what it is: a separate and independent judgment about the legality or illegality of conduct.

---

62. *See, e.g.*, INS v. Chadha, 462 U.S. 919, 959 (1983) ("[W]e have not yet found a better way to preserve freedom than by making the exercise of power subject to the carefully crafted restraints spelled out in the Constitution."); THE FEDERALIST NO. 47, at 313 (James Madison) ("The accumulation of all powers, legislative, executive, and judiciary, in the same hands, whether of one, a few, or many, and whether hereditary, self-appointed, or elective, may justly be pronounced the very definition of tyranny."); Rebecca L. Brown, *Separated Powers and Ordered Liberty*, 139 U. PA. L. REV. 1513 (1991).

63. *See* Shima Baradaran Baughman, *Subconstitutional Checks*, 92 NOTRE DAME L. REV. 1071, 1083–84 (2017) (articulating these constitutional checks and noting that they have, at times, been underenforced).

III. THE CORPUS LINGUISTICS CRITIQUE OF *COSTELLO*
AND *SMITH*

Part of the obvious appeal of corpus linguistics is that it promises us right answers—answers that are derived from data, rather than fallible humans, and thus answers "that convey the impression of scientific precision and objectivity."[64] But as with any data analysis, corpus linguistics requires humans to make decisions that will affect—perhaps conclusively—the results of the data-based inquiry. The corpus linguistics analysis that Gries and Slocum conduct appears to illustrate the centrality of human decision making in corpus linguistics analysis.

The two cases that Gries and Slocum analyze—*United States v. Costello*[65] and *Smith v. United States*[66]—are familiar to those who study law and corpus linguistics.[67] *Costello*, a Seventh Circuit opinion by Judge Posner, involved the scope of a federal statute that criminalized the harboring of undocumented aliens.[68] In *Costello*, the court held that, in order to "harbor" an alien in violation of the statute, a defendant must deliberately safeguard or conceal the alien from authorities; merely sheltering an alien was not enough.[69] *Smith* is a U.S. Supreme Court decision. It involved a statute that set a mandatory minimum sentence if a defendant "uses" a firearm "during and in relation to . . . [a] drug trafficking crime."[70] The *Smith* Court held that a defendant who trades a gun for drugs "uses" a firearm and thus is subject to the mandatory minimum.[71]

---

64. STITH & CABRANES, *supra* note 59, at 82.
65. United States v. Costello, 666 F.3d 1040 (7th Cir. 2012).
66. Smith v. United States, 508 U.S. 223 (1993).
67. *See, e.g.*, Lee & Mouritsen, *Judging Ordinary Meaning*, *supra* note 17, at 13–14, 19–20, 52–53, 68 (analyzing *Costello*); Jennifer L. Mascott, *The Dictionary as a Specialized Corpus*, 2017 BYU L. REV. 1557, 1583–86 (analyzing *Costello*); Daniel Ortner, *The Merciful Corpus: The Rule of Lenity, Ambiguity and Corpus Linguistics*, 25 B.U. PUB. INT. L.J. 101, 124–27 (2016) (analyzing *Smith*); Solan & Gales, *supra* note 2, at 1314–15, 1352–53 (analyzing *Smith*).
68. 8 U.S.C. § 1324(a)(1)(A)(iii) (2012).
69. *Costello*, 666 F.3d at 1050.
70. 18 U.S.C. § 924(c)(1)(A) (2012).
71. *Smith*, 508 U.S. at 241.

In their critique of *Costello*, Gries and Slocum criticize Judge Posner's decision to use a Google search to bolster his decision to interpret the statutory term "harbor" to mean conceal or safeguard rather than to shelter. One of their criticisms is that Posner did not account for the meanings of the term "harbor" that did not include humans as an object.[72] The most common uses of the term "harbor" involved people harboring emotions or perceptions.[73] Judge Posner's Google searches focused only on those usages that involve a person harboring another person, such as "harboring fugitives."[74]

Gries and Slocum do not fully explain why Judge Posner should have paid more attention to the usages that involved nonhuman objects.[75] Given that the statute spoke in terms of harboring "aliens"—that is, noncitizens—one would expect that results about harboring emotions or perceptions were not relevant to the interpretive task: namely whether the statute included merely sheltering an alien. Gries and Slocum may have a perfectly defensible reason for why Judge Posner ought to include these other nonhuman object uses of "harbor" in his analysis. But those reasons are not obvious to lawyers. Indeed, when Lee and Mouritsen perform their own corpus analysis of "harbor," they exclude all uses of the term that referred to feelings, perceptions, and other nonhuman, nonanimate, or nonconcrete objects.[76]

To be clear, Gries and Slocum perform a secondary analysis of "harbor" that limits the analysis to usages in which humans are the

---

72.   Gries & Slocum, *supra* note 1, at 1447–51.

73.   *Id*. at 1449 (noting that such uses made up 44.3% of their search results).

74.   *Costello*, 666 F.3d 1040, 1044.

75.   Gries and Slocum say that because harboring emotions and perceptions are the most frequently used meaning of the word, "the burden of proof was on Posner to show that the narrower and more specific sense of *to harbor* he stipulated was in fact required to define 'harbor [such] alien[s].'" Gries & Slocum, *supra* note 1, at 1449. They tell us that Posner could have done this either by providing "data that show that our frequency-based approach to 'ordinary meaning' is not borne out by other/more/better data," or by providing "data showing that *to harbor*, when used with the human direct objects he uses as search terms, *requires* a new sense, one that is different from the general 'providing space/a habitat for,' because this general one is not good enough to cover *to harbor*'s use in the statute." *Id*. at 1449–50.

76.   Lee & Mouritsen, *Judging Ordinary Meaning*, *supra* note 17, at 52–53.

direct object.[77] But that analysis also is different than the analysis Lee and Mouritsen conduct. Gries and Slocum focus on the frequency with which "harbor" is used to refer to justifiable action that presents the harborer in a positive rather than a negative light.[78] Lee and Mouritsen do not examine whether the harboring was justified. Instead, they look only at whether the search results showed that "harbor" was being used to mean "conceal" or "shelter."[79]

Gries and Slocum's critique of *Costello* is even more puzzling because it looks quite different than their critique of *Smith*. The Court in *Smith* had to decide whether a defendant who traded a firearm for drugs fell within the scope of a criminal law that increased penalties for using a firearm during and in relation to a drug crime. Like the word "harbor," the word "use" has a number of meanings. But Gries and Slocum's primary analysis of the word "use" looked only at corpus data that involved weapons.[80] They then analyzed that data to determine whether the word "use" includes the concept of trade or barter. Gries and Slocum also conduct a

---

77. Gries & Slocum, *supra* note 1, at 1447–49.

78. Gries and Slocum focus on the positive or negative portrayal because

> [Posner] defines *to harbor . . .* and then searches for *to harbor* + direct objects such as *fugitives*, *refugees*, and *Jews*, which are direct objects of the type that he stipulates in his definition, i.e. groups of humans that are justifiably harbored, and which portray the agent of harboring as positive. This, however, means that his 'analysis' also misses the fact that there are uses of *to harbor* that differ in their potential evaluative prosody: the use of *harbor enemies* is one in which harboring that kind of group of humans is probably not justified and portrays the harborer as negative. In our data, for instance, we found eighteen cases that are arguably 'not to be protected, not worthy of protection, or harboring.' The cases include *terrorists*, *bandits*, *rebels*, *militants*, *pedophile priest*, and more. In contrast, we found only nine patients of the type Posner restricted his analysis to, namely ones that are 'to be protected/worthy of protection/harboring,' which include *Jews*, *some dedicated people*, *many of the priests*, and *a minor*. Only four of our 453 instances involved *harboring fugitives*.

Gries & Slocum, *supra* note 1, at 1450. Reading the *Costello* opinion, I do not see how Posner's analysis turned on the justifiable or unjustifiable nature of the harboring. Someone who is harboring a fugitive—Posner's most frequent result—is not necessarily acting justifiably. But that result does support his conclusion "that 'harboring,' as the word is actually used, has a connotation—which 'sheltering,' and *a fortiori* 'giving a person a place to stay'—does not, of deliberately safeguarding members of a specified group from the authorities, whether through concealment, movement to a safe location, or physical protection." *Costello*, 666 F.3d at 1044.

79. Lee & Mouritsen, *Judging Ordinary Meaning*, *supra* note 17, at 68.

80. Gries & Slocum, *supra* note 1, at 1457–62.

secondary analysis of a small sample of search results in which "use" appeared in connection with a concrete object and find that none of those results employed "use" to mean trade or barter.

As someone who is trained in law, not linguistics,[81] I see the approach that Gries and Slocum take in their analyses of *Costello* and *Smith* as inconsistent. Why would the primary analysis of "to harbor" not be limited to humans when the primary analysis of "use" was limited to weapons? Even accounting for the secondary analyses, the approaches still seem inconsistent. The secondary analysis of "use" is limited to concrete objects; the analysis never sweeps as broadly as the analysis for "harbor," which includes many nonconcrete objects.

It is quite possible that the analyses Gries and Slocum performed conform to the best practices of their discipline. I am unable to judge whether the choices that they made would be accepted by linguistic experts as an appropriate or correct method to obtain corpus results. But whether their approach is defensible as a matter of linguistics methodology does not tell us how the law ought to view those results.

At the very least, the different approaches that Gries and Slocum take to determine the most frequent meaning of "harbor" and the most frequent meaning of "use" demonstrate that humans must make these choices and that true linguistic experts will sometimes take different approaches to limiting their search results. That is extremely important for lawyers, judges, and academics who are evaluating corpus linguistics as a potential theory of statutory interpretation. It tells us that corpus linguistics will give us different answers depending on how we conduct our searches and that the experts who are qualified to undertake corpus searches and analyze the results do not have a single methodology for limiting those searches. In other words, corpus linguistics does not produce a single "right answer." Instead, it requires judgment calls that must be made

---

81. I received a B.A. in Linguistics from Columbia University in 1999, and I thoroughly enjoyed my undergraduate coursework. But the papers that I wrote on initial consonant clusters and the history of English letter names hardly give me the sort of specialized background that would be necessary to perform a competent corpus linguistics analysis of statutory terms.

by people. As a result, if corpus linguistics is adopted as a statutory interpretation theory, litigants will submit dueling corpus searches. Then judges—the people that Gries and Slocum tell us are not competent to undertake this task—will have to choose between those competing expert opinions.

How to conduct the search is not the only subjective judgment involved in corpus linguistics. The analysis of search results is also fraught with choices. Take, for example, the corpus linguistics analysis that Justice Thomas Lee conducts in *State v. Rasabout*.[82] The *Rasabout* case presented the question whether the term "discharge" in a criminal statute ought to be interpreted to mean firing a single shot from a firearm or to mean firing multiple shots if those shots all occurred at the same time.[83] Justice Lee conducted a corpus search of the term "discharge" "within five words of the nouns *firearm, firearms, gun,* and *weapon*."[84] The search returned eighty-one results, and twelve of those results "clearly linked *discharge* to a single bullet."[85] Despite the fact that fewer than fifteen percent of the results clearly indicated that "discharge" was being used to mean "fire a single shot," Justice Lee concluded "that the single shot sense of this verb is overwhelmingly the ordinary sense of the term."[86] This conclusion was based on how he interpreted the term "discharge" being used in other, more ambiguous results.[87] In other words,

---

82. State v. Rasabout, 2015 UT 72, ¶ 40, 356 P.3d 1258, 1271 (Lee, J., concurring in part and concurring in the judgment).
83. *Id.* ¶¶ 88–93, 356 P.3d at 1282.
84. *Id.* ¶ 88, 356 P.3d at 1282.
85. *Id.* ¶ 89, 356 P.3d at 1282.
86. *Id.* ¶ 88, 356 P.3d at 1282 ("By examining the instances of *discharge* in connection with these nearby nouns, I confirmed that the single shot sense of this verb is overwhelmingly the ordinary sense of the term in this context.")
87. Justice Lee explained his interpretive choices as follows:

> Twelve of the eighty-one hits in the COCA search clearly linked *discharge* to a single bullet. In sixteen other hits, the discharge was accidental. I deemed those hits as also consistent with the single shot sense of *discharge,* as it seems highly unlikely if not impossible that an accidental trigger-pull could result in a release of all of the bullets in a gun's magazine. Fifteen other hits were a bit more ambiguous; but on closer examination, the *discharge* in question seemed to imply a single shot (based on the nature of the weapon, the circumstance of the discharge, or description of the resulting damage).

Id. ¶ 89, 356 P.3d at 1282.

Justice Lee had to use his professional judgment—his judicial intuition—to interpret the corpus search results.

Justice Lee is hardly the only person to use his subjective judgment in analyzing the results of a corpus search. Subjective judgment is a necessary feature of any corpus analysis. A recent paper on the meaning of the word "emolument" in the Constitution provides another example of how large a role subjective judgment plays.[88] In their methodological section, the two authors explicitly note that, because they were likely to code results differently, they practiced in order to facilitate consistent coding. They independently coded material and then met to discuss why they had arrived at particular decisions.[89] After multiple rounds of these practice coding sessions, they were agreeing only on how to code a particular result seventy percent of the time.[90] If people who have specifically trained with one another to achieve consistent results nonetheless disagree with one another thirty percent of the time, then it is quite clear that corpus linguistics cannot give us "right answers."

Given the prominent role that personal judgment plays in analyzing corpus results, I find it difficult to understand the claim by corpus linguists that frequency analyses are "replicable and falsifiable" findings.[91] When it is not immediately apparent how the statutory term is being used in a corpus search result, then the person analyzing the search result will have to rely on context and personal judgment to determine meaning. Context and personal judgment are precisely what judges currently use to decide the meaning of a statutory term. If corpus linguists think that this is an

---

88.   James Cleith Phillips & Sara White, *The Meaning of the Three Emoluments Clauses in the U.S. Constitution: A Corpus Linguistic Analysis of American English, 1760–1799*, 59 S. TEX. L. REV. (forthcoming 2018), https://papers.ssrn.com/sol3/papers.cfm?abstract_id =3036938.

89.   *Id.* at 26–27.

90.   *Id.* at 27.

91.   *See, e.g.*, Justice Thomas Lee & Stephen Mouritsen, *Judging Ordinary Meaning with Corpus Linguistics*, WASH. POST (Aug. 8, 2017), https://www.washingtonpost.com /news/volokh-conspiracy/wp/2017/08/08/judging-ordinary-meaning-with-corpus-linguisti cs/?utm_term=.738f3033613f ("Through corpus analysis we can test our hypotheses about language through rigorous experimentation with observable and quantifiable data and arrive at results that are replicable and falsifiable.").

improper method for determining statutory meaning, then it is unclear why they want to reintroduce that very method of determining meaning many times over in the analysis of corpus search results.

Of course, corpus linguistics is hardly alone as an interpretive theory in failing to give us a single right answer. Nor is it alone in relying on context and personal judgment to determine meaning. The difference is that the very *raison d'être* of corpus linguistics is that it supposedly minimizes subjectivity and discretion in the interpretive process.[92] Because corpus searches do not have a single, accepted methodology, and because corpus analyses necessarily involve subjective judgment, corpus linguistics reintroduces the very flaws that its proponents seek to guard against.[93] Corpus linguistics requires judges to make judgment calls, but it masks the import and the consequences of those judgment calls by changing the dispute to one about database searches and analysis rather than one about legal meaning and its consequences.

IV. ALTERNATIVE INTERPRETIVE TOOLS

Perhaps because they realize that corpus linguistics cannot give definitive answers, Gries and Slocum are careful to frame the corpus linguistics inquiry as something that could help judges determine a term's meaning rather than as definitive proof of meaning.[94] And I

---

92. *See supra* notes 28–32; *see also* Mouritsen, *Hard Cases*, *supra* note 9, at 178 ("If statutory terms are to be applied consistently and predictably, their ordinary meaning cannot be determined introspectively.").

93. *See* Lee & Mouritsen, *Judging Ordinary Meaning*, *supra* note 17, at 34 (praising corpus linguistics because "the results of a corpus-based conclusion will be replicable and falsifiable"); Mouritsen, *Hard Cases*, *supra* note 9, at 202 (praising corpus linguistics because it "remove[s] the question of ordinary meaning from the black-box of the judicial interpreter's mind" and employs "a method that is quantifiable and verifiable"); *Corpus Linguistics and Legal Interpretation—Event Audio/Video*, FEDERALIST SOC'Y (Jan. 23, 2017) (statement of Justice Thomas Lee at 24:10), https://fedsoc.org/commentary/videos/corpus-linguistics-and-legal-interpretation-event-audio-video?autoplay=1 (praising corpus linguistics because it minimizes judicial discretion).

94. *E.g.*, Gries & Slocum, *supra* note 1, at 1423 (stating that corpus linguistics "would be useful to interpreters"); *id*. (stating that "linguistic insights from corpora" are "relevant to the questions courts ask (even if implicitly) when interpreting legal texts").

imagine there are others who might balk at adopting corpus linguistics as an interpretive methodology, but who would nonetheless support its use as an additional tool for determining "ordinary meaning."[95] As I discuss more fully below, however, using corpus linguistics as a tool, rather than an interpretative methodology, may also have negative consequences. That is because, if corpus linguistics is used by textualists, then it may crowd out other interpretive tools. In particular, it could crowd out the rule of lenity and clear statement rules, which could deliver on corpus linguistics' promise of increasing predictability and decreasing arbitrariness without sacrificing notice or accountability.

By framing their frequency inquiry as a question about ordinary meaning, corpus linguists ensure that their methodology can be used in all statutory interpretations.[96] Recall, textualists will look for the ordinary meaning of a statute first, and they will consider nontextual sources such as legislature history, policy considerations, practice, or canons of statutory construction only if the meaning of the statutory language is not plain from the text itself. Textualist judges may always consult a corpus frequency analysis to decide the meaning of a statute because corpus linguistics purports to tell us what the "plain" or "ordinary" meaning of a word is. One assumes that corpus linguistics will prove decisive in at least some cases, and thus some cases that might have moved to the second step of statutory construction will end at ordinary inquiry meaning. As a consequence, corpus linguistics will decrease the number of cases in which the courts turn to other nontextual sources of meaning.[97]

---

95.   *See supra* note 2. I distinguish between those who would use corpus linguistics as a tool to determine the *permissible* meanings of a statutory term from those who would use it to determine the "ordinary meaning." Corpus searches give us information about how terms have been used in the past; that information seems like reliable evidence about how a term *may be* used by an ordinary English speaker, as opposed to how an ordinary English speaker is likely to understand the term to have been used in a particular statute.

96.   Indeed, those who advocate corpus linguistics as a theory of interpretation "see corpus linguistic analysis playing a central role in legal interpretation going forward." Lee & Mouritsen, *The Path Forward*, *supra* note 2.

97.   Some might argue that corpus linguistics could be a tool for declaring more statutes ambiguous. And if that were the case, then corpus linguistics could result in more applications of the rule of lenity. (I am indebted to Stephen Mouritsen for this insight.) However, that is

This is unfortunate because there are nontextual sources of interpretation that, unlike corpus linguistics, could increase predictability without sacrificing notice or accountability. Specifically, courts could employ a more robust rule of lenity or they could employ clear statement rules in order to promote predictability. These rules actually strengthen notice and accountability while increasing predictability.[98]

The rule of lenity tells judges that they should resolve ambiguity in criminal statutes in favor of defendants.[99] And there are several cases in which judges have insisted on a clear, affirmative statement by the legislature in order to dispense with criminal mens rea.[100] But courts do not use these interpretive tools consistently. The rule of lenity is, in many cases, a tool of last resort; courts will employ it only if all other interpretive tools have been exhausted.[101] And while

---

not how corpus linguistics has played out in practice. When faced with only twelve clear usages of the term "discharge," for example, Justice Lee did not conclude that the statutory term was ambiguous in *Rasabout*. State v. Rasabout, 2015 UT 72, ¶ 56, 356 P.3d 1258, 1275 (Lee, J., concurring in part and concurring in the judgment). Instead, he found reasons to interpret less clear usages in a manner that confirmed his initial intuition about the meaning of the term. *See supra* text accompanying notes 81–87. Nor do I think that such an outcome is likely to be uncommon. Behavioral economists have shown that giving people a little bit of information may cause them to be (unwarrantedly) certain about decisions involving uncertainty and to make worse decisions than with less information. *See, e.g.*, Daniel Kahneman & Amos Tversky, *On the Psychology of Prediction*, 80 PSYCHOL. REV. 237 (1973) (demonstrating that people will make irrational decisions if given information that confirms their priors).

98.   *See* Fallon, *supra* note 13, at 1305 ("Pressed to articulate a prescriptive theory, I would provisionally offer this: in any case in which there is more than one linguistically and legally plausible referent for claims of legal meaning . . . , interpreters should choose the best interpretive outcome as measured against the normative desiderata of substantive desirability, consistency with rule of law principles, and promotion of political democracy, all things considered.").

99.   *See, e.g.*, United States v. Bass, 404 U.S. 336, 347 (1971); Rewis v. United States, 401 U.S. 808, 812 (1971).

100.   *E.g.*, Staples v. United States, 511 U.S. 600, 618 (1994).

101.   *See* Moskal v. United States, 498 U.S. 103, 108 (1990) ("[W]e have always reserved lenity for those situations in which a reasonable doubt persists about a statute's intended scope even *after* resort to 'the language and structure, legislative history, and motivating policies' of the statute." (quoting Bifulco v. United States, 447 U.S. 381, 387 (1980))); *Bass*, 404 U.S. at 347 (stating that a court should rely on lenity only if, "[a]fter 'seiz[ing] every thing from which aid can be derived,'" it is "left with an ambiguous statute" (quoting United States v. Fisher, 6 U.S. (1 Cranch) 358, 386 (1805))); *see also* Ortner, *supra* note 67, at 106–20 (tracing the decreasing force of lenity in court opinions); Zachary Price,

courts have occasionally required clear statements for mens rea, requiring such statements is hardly the norm. Whether courts will require a clear statement in order to dispense with a mens rea requirement depends on "the nature of the statute and the particular character of the items regulated."[102] There are also many other situations in which courts have not required clear statements.[103]

Strengthening and expanding both lenity and clear statement rules could be used to improve notice, accountability, and predictability in the interpretation of criminal statutes.[104] The rule of lenity improves notice because it ensures that defendants will not be convicted of violating a statute that could have reasonably been read to permit their conduct.[105] Clear statement rules improve notice because they require legislatures to affirmatively and unambiguously indicate when they intend to impose criminal liability under particular circumstances.

Both the rule of lenity and clear statement rules promote democratic accountability because they clarify the content of legislation. When legislatures are forced to better articulate their policies, voters are better able to understand the content of those policies and to hold their representatives accountable for those decisions.[106] These rules also promote judicial accountability because they make clear that judges remain as a check on government power.

The rule of lenity and clear statement rules—especially if they always apply—also help to constrain judicial discretion, thus ensuring

---

The Rule of Lenity as a Rule of Structure, 72 FORDHAM L. REV. 885, 891 (2004) (noting that prevailing doctrine "ranks lenity dead last in the interpretive hierarchy").

102.	*Staples*, 511 U.S. at 607.

103.	For example, they have not required clear statements regarding whether harsh penalties apply to individuals under the age of eighteen, even though they have recognized that such penalties appear to be unintended byproducts of statutory schemes rather than the intended consequences of legislative action. *See, e.g.*, Graham v. Florida, 560 U.S. 48, 66–67 (2010).

104.	*See, e.g.*, Joseph E. Kennedy, *Making the Crime Fit the Punishment*, 51 EMORY L.J. 753 (2002) (arguing for a clear statement rule for punishment statutes in cases where culpability is unclear); Price, *supra* note 101 (arguing for a more robust rule of lenity).

105.	*See Bass*, 404 U.S. at 347.

106.	*See generally* Price, *supra* note 101 (arguing that the rule of lenity promotes democratic accountability because it ensures that voters have a clear sense of what criminal legislation prohibits).

consistency and predictability. That is not to say that these rules are neutral—they push the interpretation of criminal statutes to be narrower. But the rules help ensure that judges decide cases independently of their own idiosyncratic preferences.

CONCLUSION

Corpus linguistics is, at first glance, an attractive alternative to traditional statutory interpretation. It promises to use objective data and quantitative analysis to solve questions that have often prompted judicial disagreement.

But a closer look at corpus linguistics reveals both its limitations and its drawbacks. Corpus linguistics does not eliminate the need for human judgment in conducting searches and analyzing results.[107] More importantly, as an interpretive theory, corpus linguistics fundamentally changes the nature of judicial interpretation. It tells us to interpret a statute based on how often a word is used a certain way rather than what an ordinary speaker would have understood that word to mean. In so doing, it creates a real risk that members of the public will not have notice about the scope of the law or be able to hold their representatives responsible for their policy choices. Corpus linguistics also minimizes the role that judges play in our constitutional system by making their interpretations appear ministerial rather than significant substantive actions for which judges should take responsibility.

Our system could do more to increase the predictability of statutory interpretation. Especially when it comes to the interpretation of criminal laws, statutory construction should not seem like an arbitrary or idiosyncratic endeavor. But there are other methods courts could employ to increase the predictability of statutory interpretations—methods like more robust enforcement of

---

107.   "[Q]ualitative judgments form an integral part of [corpus linguistic's] primarily data-driven approach. There are human beings at both ends of the corpus—the architect and the user—and both are subject to the same systematic errors and biases that affect every other member of the species." Mouritsen, *Hard Cases*, *supra* note 9, at 203.

the rule of lenity or clear statement rules—that could increase predictability without decreasing notice or accountability.

In our continuing quest to improve statutory interpretation, we should not forget that judicial statutory interpretation is a feature, rather than a bug, of our constitutional system. Judges interpret statutes not simply because someone has to do it but rather because it creates another opportunity for our system to check government power and promote individual liberty. There is value in requiring the legislature to speak especially clearly when writing criminal laws and in pushing back on creative executive enforcement decisions. Substantive judicial review helps to achieve these goals. But if we tell judges that they ought to perform that role by counting database returns, it will obscure the necessarily normative character of their task.[108]

---

108.   *Cf.* Fallon, *supra* note 13, at 1306 ("[O]nce it is recognized that reasonably disputed claims of legal meaning cannot be resolved by appeal to any prelegal, linguistic fact of the matter, there is no alternative to the exercise of legally constrained normative judgment. Suggestions to the contrary are either false or misleading.").