THE YALE LAW JOURNAL

THOMAS R. LEE & STEPHEN C. MOURITSEN

Judging Ordinary Meaning

ABSTRACT. Judges generally begin their interpretive task by looking for the ordinary meaning of the language of the law. And they often end there — out of respect for the notice function of the law or deference to the presumed intent of the lawmaker.

Most everyone agrees on the primacy of the ordinary meaning rule. Yet scholars roundly bemoan the indeterminacy of the communicative content of the language of the law. And they pivot quickly to other grounds for interpretation.

We agree with the diagnosis of important scholars in this field — from Richard Fallon and Cass Sunstein to Will Baude and Stephen Sachs — but reject their proposed cures. Instead of setting aside the threshold question of ordinary meaning, we seek to take it seriously. We do so through theories and methods developed in the scholarly field designed for the study of language: linguistics.

We identify theoretical and operational deficiencies in our law's attempts to credit the ordinary meaning of the law and present linguistic theories and tools to assess it more reliably. Our framework examines iconic problems of ordinary meaning — from the famous "no vehicles in the park" hypothetical to two Supreme Court cases (*United States v. Muscarello* and *Taniguchi v. Kan Pacific Saipan, Ltd.)* and a Seventh Circuit opinion by Judge Richard Posner (in *United States v. Costello*). We show that the law's conception of ordinary meaning implicates empirical questions about language usage. And we present linguistic tools from a field known as corpus linguistics that can help to answer these empirical questions.

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. Linguists have developed computer-aided means of answering such questions. We propose to import those methods into the law's methodology of statutory interpretation. And we consider and respond to criticisms of their use by lawyers and judges.

ARTICLE CONTENTS

INTRODUCTION

A key component of the meaning we ascribe to law concerns its "communicative content." Professor Lawrence Solum has spoken of such content as consisting of the "linguistic meaning" of the words of a statute or regulation.[1] We can also think of it as encompassing the "intended" meaning of the lawmaker, to use the words of Professor Richard Fallon,[2] or the "contextual meaning" understood by the public, as framed by Professors Will Baude and Stephen Sachs.[3] This is the threshold question for the "standard picture" of legal interpretation, which starts with a search for the "ordinary communicative content" of the words of the law.[4] That search is the focus of this article. We highlight deficiencies in the law's search for ordinary meaning and introduce a tool imported from linguistics — corpus linguistic analysis — that can help overcome some of those deficiencies.

Most everyone — not just textualists anymore — agrees that "[t]here are excellent reasons for the primacy of the ordinary meaning rule."[5] Most of the reasons stem from the purported determinacy of the ordinary meaning inquiry. We

---

1. Lawrence B. Solum, *Communicative Content and Legal Content*, 89 NOTRE DAME L. REV. 479, 480 (2013) (distinguishing the "communicative content" of a legal text from its "legal content," or in other words "the legal norms the text produces").

2. Richard H. Fallon, Jr., *The Meaning of Legal "Meaning" and Its Implications for Theories of Legal Interpretation*, 82 U. CHI. L. REV. 1235, 1249-50 (2015) (speaking of these and other conceptions of the communicative or "conversational" content of the words of the law).

3. William Baude & Stephen E. Sachs, *The Law of Interpretation*, 130 HARV. L. REV. 1079, 1106 (2017).

4. *See id.* at 1086 (speaking of the "Standard Picture," or the "view that we can explain our legal norms by pointing to the ordinary communicative content of our legal texts," in other words "an instrument's meaning as a matter of language"); *see also* Mark Greenberg, *The Standard Picture and Its Discontents*, *in* 1 OXFORD STUDIES IN PHILOSOPHY OF LAW 39, 48 (Leslie Green & Brian Leiter eds., 2011) (describing the "Standard Picture"). Here and elsewhere we sometimes conflate "communicative content" and "ordinary meaning." Yet we acknowledge that some legal terms are used in an *extra*ordinary sense — as with legal terms of art. And we recognize that legal language may be viewed as a distinct dialect, and thus that "communicative content" may sometimes be understood to encompass "extraordinary" (specialized legal) meaning. *See* John O. McGinnis & Michael B. Rappaport, *The Constitution and the Language of the Law* 4-5 (Univ. of San Diego Sch. of Law Legal Studies Research Paper Series, No. 17-262, 2017), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2928936 [http://perma.cc/2V4V-C69M] (asserting that the Constitution is written in the "language of the law," not ordinary English, and thus that its interpretation should account for the canons and legal conventions that would have been accepted by the legal community at the time of the founding).

5. WILLIAM N. ESKRIDGE, JR., INTERPRETING LAW: A PRIMER ON HOW TO READ STATUTES AND THE CONSTITUTION 35 (2016).

speak of a search for meaning "not in the subjective, multiple mind of Congress but in the understanding of the objectively reasonable person."[6] And we generally conclude that the search for such meaning "matches up well with our understanding of what the *rule of law* entails"[7]: it assures notice to the public, protects reliance interests, assures consistency of application, and respects the will of the legislative body.[8] So although we recognize that "ordinary meaning does not always yield predictable answers to statutory issues," we tend to accept that it "yield[s] greater predictability than any other *single* methodology."[9]

This premise has taken hold in our courts: "[W]e're all textualists now."[10] That holds true at least in the sense that most judges begin the interpretive inquiry with the words of a statute — and even end there if they find the meaning of those words to be "plain."[11]

Yet the academy has been less sure of the premises of this trend. Scholars like Fallon and Cass Sunstein generally have endorsed the value of determinacy but roundly doubted the judge's ability to find it in the mere "communicative content" or "ordinary meaning" of statutory text.[12] There are two dimensions to this

---

6. Frank H. Easterbrook, *The Role of Original Intent in Statutory Construction*, 11 HARV. J.L. & PUB. POL'Y 59, 65 (1988).

7. ESKRIDGE, *supra* note 5, at 35.

8. *See id.* ("A polity governed by the rule of law aspires to have legal directives that are known to the citizenry, that are predictable in their application, and that officials can neutrally and consistently apply based upon objective criteria.").

9. *Id.* at 36.

10. Elena Kagan, *The Scalia Lecture: A Dialogue with Justice Kagan on the Reading of Statutes*, HARV. L. TODAY (Nov. 17, 2015), http://today.law.harvard.edu/in-scalia-lecture-kagan-discusses -statutory-interpretation [http://perma.cc/3BCF-FEFR].

11. *See* JOHN F. MANNING & MATTHEW C. STEPHENSON, LEGISLATION AND REGULATION: CASES AND MATERIALS 60 (2d ed. 2013) ("Over the last quarter-century, textualism has had an extraordinary influence on how federal courts approach questions of statutory interpretation. When the Court finds the text to be clear in context, it now routinely enforces the statute as written."); Abbe R. Gluck, *The States as Laboratories of Statutory Interpretation: Methodological Consensus and the New Modified Textualism*, 119 YALE L.J. 1750, 1756-58 (2010) (concluding, based on a comprehensive study of state court approaches to statutory interpretation, that state courts are engaged in an "effort[] to increase predictability in statutory interpretation," and that they give primacy to text and decline to look to external sources of meaning if they find the text "plain").

12. *See* Fallon, *supra* note 2, at 1255-63, 1272 (exploring a range of possible meanings of communicative or "conversational" meaning, including "semantic" or "literal" meaning, "contextual" meaning embraced by "shared presuppositions of speakers and listeners," "intended meaning," and others, and asserting that there accordingly is "no single, linguistic fact of the matter concerning what statutory or constitutional provisions mean"); Cass R. Sunstein, *There Is*

skepticism—questions about the *meaning* of the law's search for "ordinary meaning" and concerns about a judge's ability to *measure* or assess it with any degree of determinacy.[13]

As Baude and Sachs say, "we can't treat the meaning of [a given text's] language as the only source of its legal effect."[14] Our law of interpretation may have good reasons to depart from the "standard picture"—to substitute "fake" answers to linguistic questions for real ones.[15] It is doubtless true, moreover, that some of our rules of interpretation dictate a "process" that "often looks nothing like a straight-forward search for linguistic meaning."[16]

We share these commentators' concerns but offer a different solution. In this Article, we show that the law has done a poor job conceptualizing the notion of ordinary meaning, and we ultimately agree that "[u]ncertainty and division" in assessing such meaning "seem inevitable" under the methods currently resorted to by judges.[17] But we do not see these problems as an invitation to *abandon* the search for the ordinary communicative content of the law in favor of case-by-case "interpretive eclecticism."[18] Nor do we find in the indeterminacy of the search for ordinary meaning a broad license for "normative judgments" about whatever "interpretation" "makes our constitutional system better rather than

---

Nothing that Interpretation Just Is, 30 CONST. COMMENT. 193, 194-95 (2015) (identifying possible notions of meaning, including authorial intention, public meaning, moral reading, and others).

13. *See* Fallon, *supra* note 2, at 1272 (noting that "there can be a multitude of linguistically pertinent facts, generating different senses of meaning, which in turn support a variety of claims"); *id.* at 1268-69 (asserting that "[u]ncertainty and division" in measuring ordinary meaning are "inevitable," that evidence of "communicative or assertive content, understood as a matter of linguistic fact, is often sparse, minimal, or indeterminate as applied to particular cases," and that we "cannot proceed by taking or imagining the outcome of an opinion poll" about ordinary meaning).

14. Baude & Sachs, *supra* note 3, at 1088; *see also id.* at 1096 ("We see this as one of the most important functions of a legal system: to replace real answers with fake ones. There may be real answers out there to lots of important normative and policy questions, such as how fast we should drive on the highway, what tax policy is best, and so on. But people persistently disagree on the real answers, and the legal system helpfully offers fake answers instead—answers that hopefully are somewhat close to the real ones, but on which society (mostly) agrees and which allow us (mostly) to get along.").

15. *Id.* at 1082, 1096.

16. *Id.* at 1088.

17. Fallon, *supra* note 2, at 1268.

18. *Id.* at 1305, 1308 (describing "interpretive eclecticism" as involving the choice of 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").

worse."[19] This kind of "interpretation" overrides—rather than protects—the values served by the ordinary meaning rule. It undermines reliance and fair-notice interests and gives voice to the will of judges, not lawmakers.

We may eventually throw up our hands and conclude that some questions of ordinary meaning have no good answers. Or we may conclude that the law has good reason to substitute a nonlinguistic answer that vindicates policies more important than the ones advanced by the "standard picture."[20] But we cannot skip or assume away the threshold question of ordinary meaning. While the search for ordinary meaning is hard, the premises of this inquiry are too deeply embedded in our law and too clearly rooted in important policy considerations to give up at the first sight of difficulty or indeterminacy, or to judge the enterprise on the fuzzy premises or mistaken methodologies of the past. So we take up the inquiry here.

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. 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.[21] Linguists have developed computer-aided means of answering such questions. We propose to import those methods into the modern theory and practice of interpretation, and we identify problems in the methods that the law has been using to address these issues.

Our proposed methodology is a set of tools utilized in a field called corpus linguistics. Corpus linguists study language through data derived from large bodies—*corpora*—of naturally occurring language. They look for patterns in meaning and usage in large databases of actual written language. And we think their methods may easily be adapted in a manner that will allow us to conceptualize and measure the "standard picture" in a much more careful way.[22]

---

19. Sunstein, *supra* note 12, at 193-94.

20. In other words, maybe the "standard picture" doesn't claim to be a picture of American law. *Cf.* Baude & Sachs, *supra* note 3, at 1089 (arguing that there may be "real trouble for the standard picture, at least if it claims to be a picture of American law").

21. Judge Posner framed the ordinary meaning question in this (empirical) way in his opinion in *United States v. Costello*, 666 F.3d. 1040, 1044 (7th Cir. 2012). There he proposed to answer this question using the results of a Google search. We think Judge Posner's instincts were right but his methods fell a bit short, as explained below. *See* discussion *infra* Section I.B.3.

22. Corpus linguistics is not the only linguistic discipline that relies on empirical observation and experimentation. Empirical observation is a vital component of a variety of linguistic disciplines, including sociolinguistics, historical linguistics, phonetics, discourse analysis, field linguistics, computational linguistics, cognitive linguistics, and psycholinguistics. As we will discuss below, this Article focuses on corpus linguistics, but we do not mean to suggest that other

In Part I, we begin by noting the circumstances in which the "standard picture" controls under statutory interpretation, highlighting exemplary cases where the ordinary communicative content of the words of a statute seems to dictate the court's holding. Next, we identify shortcomings in the law's attempt to give effect to that communicative content—shortcomings in both the theory of ordinary meaning and in attempts to operationalize (or measure) it. In Part II, after outlining these two sets of problems, we introduce theories and empirical methods from the field of corpus linguistics that may help us deliver on the promise of an objective inquiry into ordinary meaning.[23] In Part III, we apply these tools to our exemplary cases. We close, in Part IV, by responding to actual and anticipated criticisms of our approach and by highlighting unresolved issues that must be addressed going forward.

I. ORDINARY MEANING IN THE LAW OF INTERPRETATION

Everyone agrees that our sense of the ordinary communicative content of legal language is an important starting point for interpretation. All agree, moreover, that the law should credit that content *at least sometimes*. This holds even for those who doubt our ability to settle on a single notion of meaning or to assess it with any degree of consistency.[24]

---

empirical linguistic disciplines could not be brought to bear on questions of ordinary meaning. We briefly discuss a few of these approaches below.

23. Some judges (present company included) are beginning to take note of the deficiencies we highlight here and to try to address them. In a few recent cases, judges have made a studied effort to define the inquiry into ordinary meaning more precisely. And, importantly, they have presented empirical analysis in support of their conclusions. *See, e.g.*, People v. Harris, 885 N.W.2d 832, 838-39 & n.29 (Mich. 2016) (citing a Utah Supreme Court opinion in support of the methodology of corpus linguistics and relying on corpus linguistic data to buttress the court's interpretation of the term "information" in a Michigan statute forbidding the use of "information" provided by a law enforcement officer if compelled under threat of employment sanction); *id*. at 850-51 n.14 (Markman, J., concurring in part and dissenting in part) (citing another Utah Supreme Court opinion and relying on corpus linguistic data, but drawing a different inference from the data); State v. Rasabout, 2015 UT 72, ¶¶ 68-75, 356 P.3d 1258 (Lee, Associate C.J., concurring in part and concurring in the judgment) (advancing corpus linguistic data in support of his interpretation of the phrase "discharge[] a firearm" in a state statute); State v. Canton, 2013 UT 44, ¶ 27 & n.6, 308 P.3d 517 (presenting corpus linguistic data in support of the court's construction of the phrase "out of the state" in a tolling provision for criminal statutes of limitations under Utah law); J.M.W. v. T.I.Z. (*In re* Adoption of Baby E.Z.), 2011 UT 38, ¶ 89 & nn.23-24, 266 P.3d 702 (Lee, J., concurring in part and concurring in the judgment) (advocating the use of corpus linguistic data in support of his interpretation of "custody" proceeding under the federal Parental Kidnapping Protection Act, 28 U.S.C. § 1738A (2006)).

24. *See* Baude & Sachs, *supra* note 3, at 1089; Fallon, *supra* note 2, at 1239.

Judges are generally even more sanguine about the matter. The case law in this field is marked by numerous references to the "standard picture." Judges routinely advert to the idea of crediting the "ordinary meaning" of statutory text.[25] Where such meaning is viewed as "plain," moreover, judges consistently declare the interpretive enterprise to be at an end.[26] The general rule is to credit the communicative content of statutory text where it is "plain," and in that event, to close the door to the consideration of extratextual sources of meaning or intent.[27]

A variation on the theme applies in the realm of substantive canons of construction or principles of deference. The rule of lenity, for example, says that genuine ambiguities in criminal laws are resolved in favor of the defendant;[28] the converse is the notion that "the rule of lenity has no application when the statute is clear."[29] *Chevron* deference is similar: the courts defer to agencies only where the terms of the statute are ambiguous.[30]

---

25. *See, e.g.*, Clark v. Rameker, 134 S. Ct. 2242, 2246 (2014) ("[W]e give the term its ordinary meaning."); Bond v. United States, 134 S. Ct. 2077, 2091 (2014) ("In settling on a fair reading of a statute, it is not unusual to consider the ordinary meaning of a defined term . . . ."); Mohamad v. Palestinian Auth., 566 U.S. 449, 454 (2012) ("Because the [Act] does not define the term 'individual,' we look first to the word's ordinary meaning."); Mac's Shell Serv., Inc. v. Shell Oil Prods. Co., 559 U.S. 175, 182 (2010) ("We . . . give [the relevant] terms their ordinary meanings.").

26. *See* William Baude & Ryan D. Doerfler, *The (Not So) Plain Meaning Rule*, 84 U. Chi. L. Rev. 539, 539 (2017) (characterizing the "plain meaning rule" as a "compromise" in which "other information can't be considered" if "the statute's meaning is plain," but in which other information "comes in" if "it isn't plain").

27. *See, e.g.*, Kent Greenawalt, Legislation: Statutory Interpretation: 20 Questions 35 (1999) ("No one seriously doubts that interpretation of statutes turns largely on textual meaning."); Gluck, *supra* note 11, at 1758 (stating that the "modified textualism" approach embraced in most state courts "ranks interpretive tools in a clear order—textual analysis, then legislative history, then default judicial presumptions—and it includes legislative history in the hierarchy").

28. How much ambiguity, of course, is a difficult question. *See* Abramski v. United States, 134 S. Ct. 2259, 2272 n.10 (2014) (asserting that the rule applies only if "there remains a grievous ambiguity or uncertainty in the statute" that cannot be resolved—if the Court is left to "simply guess as to what Congress intended" (quoting Maracich v. Spears, 133 S. Ct. 2191, 2209 (2013)); *id*. at 2281 (Scalia, J., dissenting) (suggesting that the rule should apply if "after all legitimate tools of interpretation" have been employed "a reasonable doubt persists" (quoting Moskal v. United States, 498 U.S. 103, 108 (1990))); Antonin Scalia & Bryan A. Garner, Reading Law: The Interpretation of Legal Texts 299 (2012) (decrying the "multiplicity of expressed standards" for invoking the rule of lenity, "leav[ing] open the crucial question . . . of how much ambiguousness constitutes an ambiguity" (quoting United States v. Hansen, 772 F.2d 940, 948 (D.C. Cir. 1985))).

29. *See* Scalia & Garner, *supra* note 28, at 301.

30. *See* Michigan v. EPA, 135 S. Ct. 2699, 2707 (2015).

However, commentators are undoubtedly right to question the determinacy of the inquiry into ordinary meaning. The problem, as noted, is twofold—going both to the law's conception of ordinary meaning and to our judges' attempts to measure it. First is a problem of theory: ironically, we have no ordinary meaning of "ordinary meaning."[31] The same goes for "plain meaning." "Courts and scholars sometimes use the phrase 'plain meaning' to denote something like *ordinary* meaning," or in other words, "the meaning one would normally attribute to [the] words" of a statute "given limited information about their context."[32] Other times "plain meaning" is used to denote *obvious* meaning—i.e., "the meaning that is clear."[33] This is the sense at work in the "plain meaning rule."[34]

Second is a problem of operationalization or measurement. The concern here is that even if we could settle on a theory of ordinary or plain meaning, we are unsure how to assess it. "Uncertainty and division seem inevitable."[35] That is true because the question of intended or understood meaning is an empirical one, and judges cannot "proceed by taking or imagining the outcome of an opinion poll" as to intended or perceived meaning.[36] The problem is underscored by the tools (mis)used by judges to try to answer this empirical question (resort to dictionary definitions or even a word's etymology, for example, as explored below).

The theoretical and measurement problems plaguing the ordinary meaning inquiry are even bigger than most have acknowledged. The depth of the problem is best illustrated by reference to concrete examples in the case law. Throughout this article we consider the following:

- Is a person guilty of *carrying a firearm* (under a federal sentencing enhancement provision) in connection with a drug crime if he merely

---

31. WILLIAM N. ESKRIDGE, JR. ET AL., CASES AND MATERIALS ON LEGISLATION: STATUTES AND THE CREATION OF PUBLIC POLICY 792-93 (4th ed. 2007) (noting the irony that "'plain meaning' is . . . a deeply ambiguous term" and highlighting differences in the ways courts use the terms "plain meaning," often to refer to a sense that is "quite clear in a literal sense," and "ordinary meaning," which may mean "the best (most coherent) textual understanding that emerges after close textual analysis"); Richard A. Posner, *Statutory Interpretation—in the Classroom and in the Courtroom*, 50 U. CHI. L. REV. 800, 808 (1983) (observing, as to the "'start with the words' canon," that "[i]t is ironic that a principle designed to clarify should be so ambiguous").

32. Baude & Doerfler, *supra* note 26, at 545.

33. *Id.*

34. *Id.*

35. Fallon, *supra* note 2, at 1268.

36. *Id.*

transports it to a drug deal in a locked glove compartment of the car he is driving? This was the question presented in *United States v. Muscarello*.[37] The *Muscarello* Court was sharply divided. All nine Justices agreed that the question came down to the "ordinary meaning" of the notion of *carrying a firearm*. Yet they divided 5-4 on whether the ordinary sense of that phrase encompassed the conveyance of a gun in a glove compartment. Each side proffered varying senses of the meaning of "ordinary meaning" and claimed support for their view in sources ill-suited to providing a reliable answer to the empirical question presented — looking to dictionaries, to isolated examples of language from literature, and even to the etymology of the verb *carry*.

- Is a litigation expert who is paid to translate written documents from one language to another an *interpreter* under a statute authorizing an award of costs for prevailing parties who utilize such an expert in litigation? This question arose in *Taniguchi v. Kan Pacific Saipan, Ltd.*[38] The Court agreed that the case came down to a matter of ordinary meaning of the term *interpreter*. Yet again the Court was divided, this time 6-3. In *Taniguchi*, the majority and dissent agreed that the more common sense of *interpreter* referred to a person engaged in simultaneous oral translation. But again, they resorted only to dictionaries and similar sources for their conclusions. They also disagreed about what the search for *ordinary meaning* ultimately entails, with the majority insisting that only the more common sense of the term is covered and the dissent asserting that a permissible sense should also count.

- Is a woman who allows her boyfriend — an undocumented immigrant — to sleep at her apartment guilty of *harboring an alien* under a federal statute criminalizing that act? This question arose in *United States v. Costello*.[39] Like *Muscarello* and *Taniguchi*, *Costello* involved a statutory term broad enough to encompass both parties' positions. Sometimes *harbor* refers to the mere act of providing shelter, but it may also indicate the sort of sheltering that is aimed at concealment. How is the court to decide which sense is the ordinary one? Writing for the majority, Judge Posner recognized the deficiencies of standard methods — principally, dictionaries — in answering that question. So he proceeded to a search

---

37. 524 U.S. 125 (1998).
38. 566 U.S. 560 (2012).
39. 666 F.3d 1040 (7th Cir. 2012).

for data, and he did so using the tool that is perhaps most familiar to us today. He performed a Google search.

Is this the best we can do? Below, we use these cases to highlight the theoretical and operational deficiencies in the law's search for ordinary meaning.

A. Theoretical Shortcomings

The case law embraces a startlingly broad range of senses of ordinary meaning. When judges speak of ordinary meaning, they often seem to be speaking to a question of relative frequency—as in a point on the following continuum:

POSSIBLE → COMMON → MOST FREQUENT → EXCLUSIVE

At the left end of the continuum is the idea of a possible or linguistically permissible meaning—a sense of a word or phrase that is attested in a known body of written or spoken language. A meaning is a possible one if we can say that "you *can* use that word in that way" (as attested by evidence that other people have used the word in that way in the past). Yet a possible meaning may be an uncommon or unnatural sense of a given term. In that case, we might note that a given sense of a term is not common in a given linguistic setting, even if it is possible to speak that way. And even a common sense of a term might not be the most frequent use of it in a certain context.

The notion of plain meaning adds the final point to the continuum. When courts speak of plain meaning (as a concept distinct from ordinary meaning) they generally mean to "denote *obvious* meaning" or "meaning that is clear."[40] A plain—obvious or clear—meaning would be more than most frequent. It would be nearly exclusive.

The four points on the continuum can be illustrated by a range of senses of the term *vehicle* in the hypothetical "no vehicles in the park" provision.[41] One attested sense of *vehicle* is the notion of a "carrier" or "agent of transmission."[42] That sense could sweep broadly. If we are thinking of the *carrier* sense of *vehicle*, the "no vehicles in the park" prohibition could possibly be viewed as covering a

---

40. Baude & Doerfler, *supra* note 26, at 545.

41. H.L.A. Hart, *Positivism and the Separation of Law and Morals*, 71 HARV. L. REV. 593, 606-15 (1958). In this Section, we make some suppositions on the points on the continuum—on which senses of *vehicle* are possible, common, and most frequent. We do so to illustrate the range of senses of ordinary meaning. We will move from supposition to empirical analysis of these questions later. *See infra* Section III.C.2.a.

42. WEBSTER'S THIRD NEW INTERNATIONAL DICTIONARY 2538 (1961).

dog or cat, which could be referred to as a *vehicle* (as a carrier of infection). Yet that sense of *vehicle* would not be viewed as a natural or common one in this linguistic setting. If the "no vehicles" ordinance is aimed at only common senses of *vehicle*, we likely would not deem it to prohibit pets. Alternatively, we could say that the meaning of *vehicle* in this setting is plain or clear, meaning that the exclusive sense of *vehicle* is one that eliminates the possibility of its extension to pets.

Is a bicycle a *vehicle* covered by the ordinance? Perhaps so — as encompassed by the sense of *vehicle* as "a means of carrying or transporting something: conveyance."[43] This sense of *vehicle* could easily be viewed as a common sense of vehicle — certainly more common than the sense of an infection *carrier* noted above. But conveyance may not be the *most* common — the statistically most frequent — sense of vehicle in this linguistic setting (an outdoor public park). If we are looking for the most frequent sense of *vehicle* in this context,[44] we might understand the term to encompass only *motor vehicles*, and thus not to cover the bicycle.

The four points on the frequency continuum do not completely capture the range of senses of ordinary meaning embraced by our courts. Sometimes judges seem to have reference to a fifth notion of ordinary — a notion of linguistic *prototype*.[45] A *prototype* is a sense, or example of a sense, that is viewed as most strongly associated with a given term in a given context. And that may jibe with the way we separate senses or definitions in our minds. A difference in word meaning may "be represented in cognition not as a set of criterial features with clear-cut boundaries" the way a dictionary would represent things, but instead "in terms

---

43. *Id.*

44. This notion of "ordinariness" is attested in the *Oxford English Dictionary*. *See* 10 THE OXFORD ENGLISH DICTIONARY 912 (2d ed. 1989) (defining "ordinary" as "Of language, usage, discourse, etc.: that most commonly found or attested . . . ."); *see also* Lawrence M. Solan & Tammy Gales, *Finding Ordinary Meaning in Law: The Judge, the Dictionary or the Corpus?*, 1 INT'L J. LEGAL DISCOURSE 253, 263 (2016) ("'Ordinary meaning,' especially as applied to particular words and phrases, is a distributional fact. A usage is 'ordinary' when it predominates.").

45. *See* McBoyle v. United States, 283 U.S. 25, 27 (1931) (Holmes, J.) (determining whether an "airplane" was a "vehicle" for the purposes of the National Motor Vehicle Theft Act of 1919, and stating: "When a rule of conduct is laid down in words that *evoke in the common mind only the picture of vehicles moving on land*, the statute should not be extended to aircraft . . . ." (emphasis added)). This notion of a "picture" "evoke[d] in the common mind" maps very well onto the concept of prototype.

of prototypes (the clearest cases, best examples) of the category."[46] Thus, prototype analysis has shown that people consider *chair* to be a more prototypical example of *furniture* than *stool*,[47] *automobile* to be a more prototypical *vehicle* than *yacht*,[48] and *robin* to be a more prototypical bird than *ostrich*.[49]

Prototype is another way to conceive of the notion of ordinary meaning in the law.[50] A judge who approaches the question of ordinary meaning by attempting to determine the most prototypical example of a given sense of a term is searching for a linguistic prototype.[51] Under this approach, the ordinary (prototype) sense of *vehicle* would be the one that is most "vehicle-like," perhaps encompassing a *passenger vehicle* with four wheels and an engine. If that is our sense of the ordinary meaning of *vehicle*, we might conclude that the hypothetical ordinance prohibits cars and trucks but not motorized scooters.[52]

This range of meaning can also be illustrated through our three feature cases. We turn to them here.

---

46. Eleanor Rosch, *Cognitive Representation of Semantic Categories*, 104 J. EXPERIMENTAL PSYCHOL. 192, 193 (1975).

47. *Id.* at 229.

48. *Id.* at 230.

49. *Id.* at 232.

50. *See* Lawrence M. Solan, *Law, Language, and Lenity*, 40 WM. & MARY L. REV. 57, 67-68 (1998) [hereinafter Solan, *Law, Language, and Lenity*] ("In the realm of statutory interpretation, judges often evoke the canon that they are to give words in a statute their 'ordinary' meaning. Prototype analysis tells us that the notion of ordinary meaning has a cognitive basis."); Lawrence M. Solan, *Why Laws Work Pretty Well, but Not Great: Words and Rules in Legal Interpretation*, 26 L. & SOC. INQUIRY 243, 258 (2001) ("Some Supreme Court cases concerning statutory interpretation can be seen as battles among the justices over definitions versus prototypes." (citing Smith v. United States, 508 U.S. 223 (1993))); *see also* ESKRIDGE ET AL., *supra* note 31, at 850 (discussing prototypical meaning in the context of statutory interpretation); Lawrence M. Solan, *The New Textualists' New Text*, 38 LOY. L.A. L. REV. 2027, 2042-55 (2005) [hereinafter Solan, *The New Textualists' New Text*] ("One function of the ordinary meaning approach is to use prototypical experiences as a proxy for contextualization."); *The Supreme Court, 1997 Term—Leading Cases*, 112 HARV. L. REV. 355, 362 (1998) ("[W]hen a legislature uses non-technical terms . . . it is likely that both the legislature and the general public interpret the term in accordance with its prototypical meaning.").

51. Solan, *Law, Language, and Lenity*, *supra* note 50, at 67-68.

52. This conclusion, however, cannot be derived with mere intuition. The discovery of a prototype for a given word in a given context requires the application of empirical methods, as we will discuss below.

1. Muscarello v. United States

In *Muscarello* the Court was asked to interpret a statute calling for a five-year mandatory prison term for a person who "uses or carries a firearm" "during and in relation to" a "drug trafficking crime."[53] Frank Muscarello was convicted on drug charges after he was shown to have transported a gun in the locked glove compartment of his car to a drug deal. The question presented was whether that counted as "carrying" under the statute. Both the majority (Justice Breyer) and dissenting (Justice Ginsburg) opinions agreed that the proper interpretation of "carries a firearm" came down to the "ordinary English meaning" of that "phrase."[54] Yet neither opinion settled on a single sense of "ordinary." Instead, both opinions slide back and forth along the continuum, without acknowledging that they are doing so.

At some points Justice Breyer seems to employ a merely "common" sense of ordinary. For example, he asserts that the *transport in a vehicle* sense of *carry* is ordinary given that "many" — "perhaps more than one-third" — of the instances of *carrying a firearm* in the *New York Times* and *U.S. News* databases reflect that sense,[55] and he concludes that "the word 'carry' in its ordinary sense *includes* carrying in a car."[56] Yet elsewhere Justice Breyer seems to speak of the car-carrying sense as most frequent. He reasons that 1) the "ordinary English" sense of *carry* is to *transport* it in a vehicle; 2) the *bear personally* sense is "special"; and 3) "we believe Congress intended to use the word in its primary sense and *not* in this latter, special way."[57]

Justice Ginsburg's dissent is also inconsistent. In concluding that the personally *bearing* sense is ordinary, Justice Ginsburg asserts that it is "hardly implausible, nor at odds with *an accepted* meaning" of the statutory terms.[58] That is the language of possibility or commonality.[59] Elsewhere, however, Justice Ginsburg seems to speak in terms of personally *bearing* as the most frequent sense of the

---

53. Muscarello v. United States, 524 U.S. 125, 126 (1998) (quoting 18 U.S.C. § 924(c)(1)(A) (2012)).

54. *Id.* at 127; *id.* at 139 (Ginsburg, J., dissenting).

55. *Id.* at 129 (majority opinion).

56. *Id.* at 131 (emphasis added).

57. *Id.* at 128 (emphasis added).

58. *Id.* at 149 (Ginsburg, J., dissenting) (emphasis added).

59. *Id.* at 143-44 (asserting "that 'carry' is a word commonly used to convey various messages," and that it "could mean" either personally bear or transport in a vehicle).

term—in noting, in response to Justice Breyer's statistics, "what meaning showed up some two-thirds of the time."[60]

2. Taniguchi v. Kan Pacific Saipan, Ltd.

Taniguchi was a personal injury case.[61] The plaintiff was a Japanese baseball player suing for medical expenses and lost income from contracts he was unable to honor as a result of injuries at the defendant's resort.[62] The defense "paid to have various documents translated from Japanese to English,"[63] and when the district court dismissed Taniguchi's case on summary judgment, the defense submitted a request for compensation for the amounts it paid for document translation. As in *Muscarello*, the *Taniguchi* case came down to ordinary meaning. Here the operative language was from a statute allowing the prevailing party in federal litigation to recover certain costs, including those incurred by an "interpreter."[64]

This case also seems to turn on the operative notion of ordinary meaning. Writing for the majority, Justice Alito concludes that the ordinary sense of *interpreter* is *oral translator*: "an interpreter is normally understood as one who translates orally from one language to another."[65] While Justice Alito says that *written translator* is possible, he concludes that this is "hardly a common or ordinary meaning."[66] Indeed, Justice Alito characterizes the *written translator* notion of *interpreter* as "obsolete," citing dictionaries to support that conclusion.[67]

Justice Ginsburg's dissent acknowledges that *interpreter* "commonly refers to translators of oral speech" but concludes that the term "more than occasionally" is "used to encompass those who translate written speech as well."[68] This is the core basis of the view of the *Taniguchi* dissenters. They do not expressly disagree with Justice Alito's assertion that the *oral translator* notion is most common; they

---

60. *Id.* at 143.

61. Taniguchi v. Kan Pac. Saipan, Ltd., 566 U.S. 560 (2012).

62. *Id.* at 562.

63. *Id.* at 563.

64. *Id.* at 562.

65. *Id.* at 569.

66. *Id.*

67. *Id.*

68. *Id.* at 576 (Ginsburg, J., dissenting).

are simply saying that both common senses of the term should count as ordinary.[69]

3. United States v. Costello

The defendant in *Costello* was charged with knowingly "conceal[ing], harbor[ing], or shield[ing] from detection" an "alien in any place, including any building or any means of transportation."[70] Her alleged crime was "having permitted [her] boyfriend to live with her,"[71] knowing that he was an "illegal alien."[72] The principal question presented was whether the ordinary meaning of the verb *harbor* required proof of concealment.

As in *Taniguchi*, the difference between the majority and dissent in *Costello* seems to come down largely to the conception of the meaning of ordinary meaning. Judge Posner, writing for the majority, warns of the perils of overreliance on the dictionary to resolve questions of ordinary meaning. And he directs the ordinary meaning analysis to an empirical inquiry, which he proposes to resolve by means of a Google search.[73]

Judge Posner's reliance on his Google results places his sense of ordinary meaning on the frequency continuum. He uses Google to look for relative numbers of "hits" for phrases like "harboring fugitives" and "harboring guests."[74] Because Judge Posner found more hits for the former than for the latter, he concludes that "'harboring,' as the word is actually used, has a connotation . . . of deliberately safeguarding members of a specified group from the authorities."[75] This is a "most frequent" sense of ordinariness — and a blatantly empirical example of that inquiry.

The *Costello* dissent takes a different tack. In concluding that the *providing shelter* notion of *harbor* falls within the statute, Judge Manion asserts that "the ordinary meaning of 'harboring' certainly *includes* 'providing shelter to.'"[76] In support of this point, Judge Manion cites definitions from dictionaries in print

---

69. *Id.* (asserting that the *written translator* sense is an "acceptable usage" even if it is "not 'the most common usage'" (quoting *id.* at 568 (majority opinion))).

70. United States v. Costello, 666 F.3d 1040, 1041 (7th Cir. 2012) (quoting 8 U.S.C. § 1324(a)(1)(A)(iii) (2012)).

71. *Id.* at 1043.

72. *Id.* at 1042.

73. *Id.* at 1044-45.

74. *Id.* at 1044.

75. *Id.*

76. *Id.* at 1052 (Manion, J., dissenting) (emphasis added).

at the time of the statute's enactment. He says that these dictionaries show that "[t]his was *a common understanding* of the term when the term 'harbor' was first added to the statute in 1917, and when the statute was amended and the term retained in 1952."[77]

<div align="center">∗∗∗</div>

What can we learn from these cases? Our judges purport to be speaking of a consistent, common sense of ordinary meaning. But they switch back and forth between different senses of ordinary meaning, usually without acknowledging the inconsistency. Sometimes (as in *Muscarello*) judges embrace varying senses of ordinary meaning within a single opinion. Elsewhere (as in *Taniguchi* and *Costello*) the seemingly nuanced distinction between different senses of ordinary meaning becomes outcome-determinative. This is problematic—not just for statutory interpretation, but also for the rule of law.[78]

B. Operational Shortcomings

The theoretical deficiencies identified above are one element of the problem. Another is operational—in the way we seek to identify or measure the ordinary meaning of statutory terms. Typically, this assessment is made at a gut level, on the basis of a judge's linguistic intuition, without recognition of the empirical nature of the question.

A judge considering the prohibition on *vehicles* in the park, for example, would reject out of hand the notion that the ordinance extends to pets, insisting (without further analysis or support) that the *infection carrier* sense of *vehicle* is an outlier—an extraordinary meaning. A parallel conclusion would be likely in response to an attempt to extend the *no vehicles* ordinance to bicycles. We understand *vehicle* to encompass a *conveyance on wheels*, but again a court seems likely to jump to the conclusion that the ordinary sense of *vehicle* is *motor vehicle*, and that a bicycle does not count.

These conclusions seem uncontroversial. But the judge who makes them is making an empirical assessment. Gut-level empirics probably will not bother us if they go only to a holding that a pet or bicycle is not a *vehicle* prohibited in the

---

77. *Id.* (emphasis added) (first citing WEBSTER'S NEW INTERNATIONAL DICTIONARY OF THE ENG-LISH LANGUAGE 981 (1917) (defining "harbor" as "[t]o afford lodging to; to entertain as a guest; to shelter; to receive; to give refuge to"); and then citing WEBSTER'S NEW COLLEGIATE DICTIONARY 376 (John P. Bethel et al. eds., 1953) (defining "harbor" as "to entertain as a guest; to shelter; to give a refuge to")).

78. *See* Baude & Sachs, *supra* note 3, at 1089-90 (noting that "we have to decide *which* meaning, produced by which *theory* of meaning, we ought to pick").

park. But what about a motorized scooter or a golf cart? These are harder questions. And here we may have more cause for concern about the lack of transparency and determinacy.

With this in mind, judges sometimes turn to other grounds for their assessment of ordinary meaning, looking up a word in a dictionary or even turning to the word's etymology. A common use of a dictionary involves simple cherry-picking. "Instead of acknowledging and rejecting contrary senses of a statutory term, judges tend to ignore them — identifying only the sense of a word they deem ordinary without acknowledging any others."[79] As to *vehicle*, for example, a judge might simply cite a definition referring to an automobile and assert, without more, that the term's ordinary meaning does not encompass a motor scooter, or maybe even a golf cart. That is troubling — a judge who cherry-picks a preferred dictionary definition while ignoring an alternative is misusing the dictionary.

Some judges, to their credit, are more transparent. Instead of ignoring a contrary definition — the *conveyance on wheels* notion of vehicle, for example — a judge may acknowledge competing senses but find a basis for embracing one as ordinary. For example, a judge might prefer the definition that appears first in a dictionary's list of senses, or cite the etymology of the statutory term. Neither of these approaches is defensible, however, for reasons explained immediately below in our critique of the Court's ordinary meaning analysis in *Muscarello* and *Taniguchi*. *Costello*, on the other hand, acknowledges some of the problems we identify and turns to Google, albeit in a manner that raises a new set of problems.

1. Muscarello v. United States

The *Muscarello* majority invokes both sense ranking and etymology in support of its holding. Justice Breyer acknowledges that *carry* can be understood to mean either *transport* in a vehicle or *bear* on your person.[80] But he embraces the

---

79. State v. Rasabout, 2015 UT 72 ¶ 53, 356 P.3d 1258, 1274 (Lee, Associate C.J., concurring in part and concurring in the judgment); *see also, e.g.*, Kovach v. Zurich Am. Ins. Co., 587 F.3d 323, 346 (6th Cir. 2009) (McKeague, J., dissenting) (criticizing the majority for ignoring other definitions in basing its presentation of the "ordinary meaning" of "accidental" on one definition without regard to others); Konop v. Hawaiian Airlines, Inc., 302 F.3d 868, 878 (9th Cir. 2002) (ignoring broader definitions in favor of a narrow definition as "ordinary meaning" of "intercept"); United States v. Warner Bros. Well Drilling, No 89-5494, 1990 WL 37610, at *2-3 (6th Cir. Apr. 3, 1990) (citing only one definition of "operator" in determining the ordinary meaning, even though opposing definitions existed).

80. Muscarello v. United States, 524 U.S. 125, 128 (1998).

former sense as "primary" and dismisses the latter as "special."[81] His first argument in support of that conclusion is that "[t]he Oxford English Dictionary gives as its *first* definition 'convey, originally by cart or wagon, hence in any vehicle, by ship, on horseback, etc.'"[82] The italicized emphasis on "first" is Justice Breyer's. His opinion takes a similar tack in citing the "*first* definition" in *Webster's Third*—"move while supporting (*as in a vehicle* or in one's hands or arms)"[83]—and the "*first* definition" in the *Random House Dictionary*—"to take or support from one place to another; convey; transport."[84]

Justice Breyer reinforces his reliance on sense ranking in his reference to the personally *bear* sense of carry in the *Oxford English Dictionary*, noting that this is the "*twenty-sixth* definition" in the dictionary.[85] This is the threshold basis of Justice Breyer's conclusion that "[t]he relevant linguistic fact[]" is "that the word 'carry' in its ordinary sense includes carrying in a car."[86]

Justice Breyer also turns to etymology, in asserting that "[t]he origin of the word 'carries' explains why the first, or basic, meaning of the word 'carry' includes conveyance in a vehicle."[87] Justice Breyer states that *carry* traces from "Latin 'carum,' which means 'car' or 'cart,'" and from "Old French 'carier' and late Latin 'carricare,' which meant to 'convey in a car.'"[88] The precise premises of Justice Breyer's analysis are left implicit. But the point seems clear: the etymology of the verb *carry* confirms that the *transport* sense of the term is ordinary and the personally *bear* sense is unusual.

This is problematic. If the ordinary meaning question in *Muscarello* is an empirical question of frequency or prototype analysis, neither the dictionary nor etymology is useful. The dictionaries typically cited by our courts (including those cited by Justice Breyer) make no claims about the relative frequency of the listed senses of a given word.[89] Many commonly used, unabridged dictionaries

---

81. *Id.*

82. *Id.* (quoting 2 THE OXFORD ENGLISH DICTIONARY, *supra* note 44, at 919).

83. *Id.* (quoting WEBSTER'S THIRD NEW INTERNATIONAL DICTIONARY 343 (1986)).

84. *Id.* (quoting THE RANDOM HOUSE DICTIONARY OF THE ENGLISH LANGUAGE—UNABRDIGED 319 (2d ed. 1987)).

85. *Id.* at 130 (citing 2 THE OXFORD ENGLISH DICTIONARY, *supra* note 44, at 921).

86. *Id.* at 131.

87. *Id.* at 128 (citing THE BARNHART DICTIONARY OF ETYMOLOGY 146 (Robert K. Barnhart ed., 1988)).

88. *Id.* (first citing THE BARNHART DICTIONARY OF ETYMOLOGY 146 (Robert K. Barnhart ed., 1988); and then citing 2 THE OXFORD ENGLISH DICTIONARY, *supra* note 44, at 919).

89. *See* Stephen C. Mouritsen, *The Dictionary Is Not a Fortress: Definitional Fallacies and a Corpus-Based Approach to Plain Meaning*, 2010 BYU L. REV. 1915, 1924-29 (discussing problems with dictionary usage by courts and identifying the "sense-ranking fallacy"). The *Random House*

arrange their definitions based on evidence of historical usage.[90] *Webster's Third* expressly disavows any attempt to establish a "hierarchy of importance" among different senses and admits that "[s]ometimes an arbitrary arrangement or rearrangement is the only reasonable and expedient solution to the problems of ordering senses."[91]

A similar problem undermines Justice Breyer's use of etymology. As the philologist Henry Sweet observed:

> The meaning of a word in a given period of a given language is a matter of usage, and the fact of its having had a certain meaning at some earlier period or in some cognate language does not necessarily afford any help in determining, and still less in remembering, its present meaning.[92]

If this were not true, then *December* would mean the tenth month, and an *anthology* would mean a bouquet of flowers.[93] Yet, so common is the assumption

---

Dictionary of the English Language appears to be an exception. Its front matter states "a general policy of putting the most frequently used meanings . . . at the beginning of the entry, followed by other senses in diminishing frequency of usage, with archaic, and obsolete senses coming last." RANDOM HOUSE DICTIONARY OF THE ENGLISH LANGUAGE — UNABRIDGED, at viii (2d ed. 1987). But we see grounds for skepticism of these sorts of claims. *See infra* notes 99-102, 137-140 and accompanying text. When unabridged dictionaries assembled their citation files, they were concerned about possible usage, not about making a representative, scientific sample of the speech community. So, their claims about frequency and obsolescence are suspect. *Random House* acknowledges that its sense ranking based on frequency holds only "generally." RANDOM HOUSE DICTIONARY OF THE ENGLISH LANGUAGE — UNABRIDGED, *supra*, at xxii. Without more (and there is no more in this dictionary), the reader is left to guess about which senses are ordered according to frequency and which ones follow some other organizing principle.

90. *See* 1 THE OXFORD ENGLISH DICTIONARY, *supra* note 44, at xxix ("That sense is placed first which was actually the earliest in the language: the others follow in order in which they have arisen."); WEBSTER'S THIRD NEW INTERNATIONAL DICTIONARY 19a (1971) (indicating that the order of senses is "historical," in that "the one known to have been first used in English is entered first"; also stating that its "system of separating senses" is "only a lexical convenience," and not an "enduring hierarchy").

91. WEBSTER'S THIRD NEW INTERNATIONAL DICTIONARY OF THE ENGLISH LANGUAGE, *supra* note 90, at 17a.

92. HENRY SWEET, THE PRACTICAL STUDY OF LANGUAGES: A GUIDE FOR TEACHERS AND LEARNERS 88 (1900).

93. THE BARNHART CONCISE DICTIONARY OF ETYMOLOGY 29 (Robert K. Barnhart ed., 1995) (providing the etymology of "anthology" as "1640, collection of the 'flowers' of verse (i.e. small, choice poems) by various authors; borrowed, perhaps by influence of French *anthologie*, from Greek *anthología* flower-gathering (*ánthos* flower + *légein* gather)."); *id.* at 188 (providing the etymology of "December" as "1122, borrowed from Old French *decembre*, from Latin *December*, from *decem* TEN, this being originally the tenth month of the early Roman calendar (which began with March)").

that a word's etymology shows its true meaning that the assumption has been given a name: the "etymological fallacy."[94] For this reason, Justice Breyer's analysis of the etymology of *carry* tells us nothing about its ordinary meaning.[95]

2. Taniguchi v. Kan Pacific Saipan, Ltd.

The *Taniguchi* opinion appears, at first glance, to employ dictionaries in a less arbitrary way. Justice Alito does not turn to sense ranking or etymology. He presents an informal "survey" of dictionary definitions, asserting that "only a handful" of dictionaries include the *written translator* sense of *interpreter*, but "all" of them speak of the *oral translator* sense.[96] And he says that the "sense divider[s]" in the cited dictionaries confirm the Court's holding in designating the *oral translator* notion as one that is "especially" indicated and flagging the *written translator* sense as "obsolete."[97]

Yet Justice Alito's approach is still problematic. The "survey" of dictionaries is far from systematic. Justice Alito presents his own set of preferred dictionaries. And within the cited dictionaries, the Court sometimes cites a definition of the noun *interpreter* and sometimes cites a definition of the verb *interpret*. We cannot tell from the opinion whether the *written translator* sense of *interpreter* is less often listed in a real "survey" of dictionaries because we are not presented with an actual *survey* of dictionaries. We have only the definitions that Justice Alito has presented for our review.[98]

---

94. SWEET, *supra* note 92, at 88; *see also* RANDOLPH QUIRK, STYLE AND COMMUNICATION IN THE ENGLISH LANGUAGE 86 (1982) (characterizing as "one of the most pernicious of popular *idées fixes*" the notion that a word's etymology "gives you the 'real' meaning").

95. It is also worth noting that neither the ordinary legislator nor the ordinary citizen are likely to have a working knowledge of the etymology of most words. Interpreting a statute according to a long-lost meaning that neither the drafter nor the citizen is aware of seems a far cry from searching for ordinary meaning.

96. Taniguchi v. Kan Pac. Saipan, Ltd., 566 U.S. 560, 568-69 (2012).

97. *Id.* at 568-69; *see id.* at 567-68 & n.2 (noting that the *Concise Oxford Dictionary of Current English*, *Webster's Third New International Dictionary*, the *World Book Dictionary*, and *Cassell's English Dictionary* designate the oral translator meaning as "especially" indicated); *id.* at 569 (noting that the *Oxford English Dictionary* "designated [the written translator] meaning as obsolete").

98. Even a documented survey of every known dictionary might not be sufficient, moreover, for reasons explained below. *See infra* notes 99-102, 137-140 and accompanying text. Dictionaries are not trying to show ordinary meaning. But even if they were, the methods that they use to sample language use don't create a reliable sample — aggregating dictionaries isn't going to accomplish anything if none of them has a reliable sample of language usage.

Justice Alito's sense dividers are also insufficient. First, not all dictionaries designate *written translator* as obsolete or *oral translator* as special. At least one definition mentioned in the majority opinion explicitly encompasses the written sense of the term, without any indication of obsolescence.[99]

Second, sense dividers are not reliable measures. Dictionaries tell us very little about the basis for the "obsolete" sense designation. Ultimately, such a designation must be made on the basis of some underlying data that is unavailable to the reader of the dictionary. So the "obsolete" designation tells us only that the lexicographers who compiled the dictionary in question deemed a particular sense to no longer be in use; but, without more, such designation gives us only the opinion of those lexicographers and not a hard basis for an empirical conclusion.[100]

An "especially" designation may be even more unreliable. Such a designation suffers from all of the problems inherent in the "obsolete" designation, and it also masks another deficiency, going to the arbitrariness of the distinction between two senses listed in a dictionary (described further below). The fact that a given sense, or subsense, of a term is a *special* application of another highlights the interrelationship between the two senses.[101] It suggests that the two senses are not highly distinct from each other, but instead are exemplars or prototypes of a broader category. That is what the *Webster's* definition cited in *Taniguchi* seems to convey. The cited *Webster's Third* definition of *interpreter* is "'one that translates; *esp*: a person who translates orally for parties conversing in different tongues.'"[102] This is an indication that the lexicographers who formulated this definition for *Webster's* viewed the *especially* designated notion *not* as a separate sense but as an exemplar of it—perhaps a common, prototypical example.

For these reasons the *Taniguchi* opinion also employs inadequate tools of measurement. Justice Alito's "survey" and sense designations seem more sophisticated, but ultimately they are also poor tools for assessing empirical questions of ordinary meaning.

---

99. *Taniguchi*, 566 U.S. at 567 (citing *Interpreter*, BALLENTINE'S LAW DICTIONARY 655 (3d ed. 1969) (defining "interpreter" as "[o]ne who interprets, particularly one who interprets words written or spoken in a foreign language")).

100. *See* DOUGLAS BIBER ET AL., CORPUS LINGUISTICS: INVESTIGATING LANGUAGE STRUCTURE AND USE 22 (1998) (observing that "citation slips" used by lexicographers represent only those contexts "that [human] readers happen to notice").

101. *See infra* Part III.

102. *Taniguchi*, 566 U.S. at 567-68 (quoting WEBSTER'S THIRD NEW INTERNATIONAL DICTIONARY 1182 (1976)).

3. United States v. Costello

Judge Posner rejects a dictionary-based approach to ordinary meaning in *Costello*. He rightly notes that "[d]ictionary definitions are acontextual, whereas the meaning of sentences depends critically on context, including all sorts of background understandings."[103] And for that reason, Judge Posner turns to Google to get a "rough index of the frequency of [*harboring*'s] use."[104] This approach is innovative. But it is far from perfect.

Google might seem to be a good source for data-driven analysis of language usage. "The World Wide Web is enormous, free, immediately available, and largely linguistic."[105] And it is "appealing to use the Web as a data source" because "language analysis and generation benefit from big data."[106] Google has low entry costs, moreover. Even the most Luddite lawyer or judge is likely to be able to perform a basic Google search. Yet we still see a range of problems in Judge Posner's approach.

First is the black box of the Google algorithm. Google searches "are sorted according to a complex and unknown algorithm (with full listings of all results usually not permitted) so we do not know what biases are being introduced."[107] Google returns can vary by geography, by time of day, and from day to day.[108] Google search results are thus rather unscientific, if we understand good science as including replicability.

Second are problems with the Google search engine: the fact that it does not allow us to search only for verb forms of *harbor* and that it will not allow us to look at a particular speech community or period of time (only contemporary web pages, even if their content was first published in the past). If we are interested in knowing the ordinary use of *harbor* as a verb among ordinary English speakers at the time of the enactment of the statute at issue (1917), Google cannot give us that kind of parsed data.

In light of these search engine problems, Judge Posner formulated his own set of search terms — comparing hit counts for phrases like "harboring fugitives"

---

103. United States v. Costello, 666 F.3d 1040, 1044 (7th Cir. 2012).

104. *Id.*

105. Adam Kilgarriff, *Googleology Is Bad Science*, 33 COMPUTATIONAL LINGUISTICS 147, 147 (2007) (discussing the limitations of Google as a corpus).

106. *Id.*

107. *Id.* at 148.

108. *Id.*

and "harboring guests."[109] But this innovation introduces a third set of problems: Judge Posner gives no basis for his chosen set of search terms, and the terms he chose seem likely to affect the outcome.

Finally, even setting aside the problems discussed above, the hit counts that Judge Posner relies on may not be indicative of ordinariness in the sense of frequency of usage. Judge Posner implies that relative hit counts are an indication of frequency of usage in our ordinary language. But that may not hold. Google hit counts are based on the total number of web pages, not the total number of occurrences of a given phrase.[110] A single web page may have tens, hundreds, or thousands of uses of an individual word or phrase that would only register in a Google search as a single hit. So hit counts may not be a reliable indication of ordinariness, even if we could overcome the other problems identified here.

We think Judge Posner was onto something in seeking an empirical method of measurement, but we also think his Google search was inadequate.

II. THEORIZING ORDINARY MEANING

The deficiencies in the courts' approaches to ordinary meaning are also reflected in legal scholarship. Here we outline some of the approaches to ordinary meaning reflected in the scholarly literature, in an attempt to expand on the themes discussed in Part I.

Legal scholarship posits a range of conceptions of ordinary meaning. Professor Richard Fallon's catalog is perhaps the most extensive. He speaks of the "semantic" or "literal" meaning of the words of the law; the "contextual" meaning informed by "shared presuppositions" of speakers and listeners (which we take to align with Sunstein's notion of "public meaning"[111] and Baude and Sachs's idea of the "reader's understanding"[112]); the "intended meaning" of the lawmaker; the "reasonable" or "imputed" meaning attributed to "hypothetical, reasonable legislators"; and the "interpreted meaning" of laws in judicial precedent.[113]

Not all of these conceptions of meaning are applicable to our analysis here. Certainly there is a case for respecting statutory meaning embedded in judicial precedent. If judges have deemed a statute to have a certain meaning in the past,

---

109. *Costello*, 666 F.3d at 1044.

110. Kilgarriff, *supra* note 105, at 147.

111. Sunstein, *supra* note 12, at 198.

112. Baude & Sachs, *supra* note 3, at 1090 (distinguishing the "author's intent" and the "reader's understanding").

113. *See* Fallon, *supra* note 2, at 1255-63.

the law of interpretation—informed by principles of stare decisis—can (and should) yield due deference to the "interpreted meaning" established by precedent.[114] But our reasons for respecting such meaning have nothing to do with the rule of law premises behind the law's search for ordinary communicative content.

Fallon's notion of "reasonable" or "imputed" meaning is also, but less obviously, a conception of extra-ordinary meaning. This construct is related to the "fair reading" method advanced by Justice Scalia and Bryan Garner in *Reading Law*.[115] The fair reading inquiry is framed in objective-sounding terms—in a search for "objectified intent."[116] But on closer review this notion of meaning has nothing to do with actual communicative content of the words of the law (or of intentions attributable to lawmakers). It is an idealized, constructive inquiry aimed at an impossibly well-informed legislator—one with "aptitude in language, sound judgment, the suppression of personal preferences regarding the outcome, and, with older texts, historical linguistic research"; "an ability to comprehend the *purpose* of the text, which is a vital part of its context," and to glean it "only from the text itself"; and even an understanding of "a word's historical associations acquired from recurrent patterns of past usage" and "a word's immediate syntactic setting—that is, the words that surround it in a specific utterance."[117]

This notion of meaning has little or nothing to do with the actual meaning intended by a legislator or understood by the public. We may well have reasons to credit this sort of idealized, constructive meaning. But if we do, it will not be because we think that any actual legislator is likely to have read the words of a law and understood it in this "reasonable" way, much less that an ordinary member of the public gleaned that understanding. It will be because we deemed other policies—policies having nothing to do with vindicating linguistic meaning—to be of greater significance.[118]

---

114. *See id.* at 1251 (articulating stare decisis arguments in support of the law's acceptance of "interpreted meaning").

115. SCALIA & GARNER, *supra* note 28, at 428.

116. ANTONIN SCALIA, Common-Law Courts in a Civil-Law System: The Role of United States Federal Courts in Interpreting the Constitution and Laws, in A MATTER OF INTERPRETATION: FEDERAL COURTS AND THE LAW 17 (Amy Gutmann ed., 1997).

117. *Id.* at 33.

118. *See* Gary Lawson & Guy Seidman, *Originalism as a Legal Enterprise*, 23 CONST. COMMENT. 47, 48 (2006) (stating that the "touchstone" of this approach to interpretation "is not the specific thoughts in the heads of any particular historical people . . . but rather the hypothetical understandings of a reasonable person who is artificially constructed by lawyers").

That leaves, in Fallon's taxonomy, (a) semantic meaning, (b) contextual meaning (public meaning or the reader's understanding), and (c) intended meaning. Semantic meaning is meaning that the language of the law would have "for someone operating solely with dictionary definitions, rules of grammar, and other general propositions bearing on how the meaning of a sentence emerges from the combination of its elements."[119] Fallon attributes this sort of meaning to a sort of "literalist" textualism, asserting that "[p]articipants in legal discourse frequently assume or argue that a legal provision's semantic or literal meaning determines its legal meaning."[120]

The point here is the notion that "literalist" textualism excludes nonsemantic context. In contrasting his notion of "contextual" meaning, for example, Fallon distinguishes "semantics, which is concerned with the context-independent meaning of words, phrases, and sentences, and pragmatics, which involves the meaning of utterances in particular contexts."[121] Here he cites an example from *Reading Law*: "*Nail* in a regulation governing beauty salons has a different meaning from *nail* in a municipal building code."[122] He also distinguishes semantic meaning from "contextual meaning" because the latter is "framed by the shared presuppositions of speakers and listeners."[123]

We agree that judges often frame their discussion of ordinary meaning in terms of literalist versus contextual meaning. But our theory of ordinary meaning parts company with Fallon at his suggestion that what he calls the "semantic meaning" of an utterance should be distinguished from its "contextual meaning," or that these two competing notions of meaning allow for a "choice among multiple candidates to supply legal or conversational meaning."[124]

Whenever we engage in the act of communication—whenever a speaker speaks and a hearer hears—our minds take in the relevant interpretative information at once. We take account of the formal aspects of an utterance (its lexical, syntactic, and semantic content), as well as the pragmatic (in the linguistic sense of the term) aspects of the utterance (for example, the physical or social setting in which it is uttered). We interpret an utterance as part of a community of

---

119. Fallon, *supra* note 2, at 1245.

120. *Id.*

121. *Id.* at 1246.

122. *Id.* (quoting SCALIA & GARNER, *supra* note 28, at 20). We use the phrase "semantic meaning" more specifically to refer to the formal linguistic features of an utterance (i.e., syntax and argument structures, semantic features, and functional roles), but we do not suggest that semantic meaning can ever be derived with reference to pragmatics or the "meaning of utterances in particular contexts."

123. *Id.*

124. *Id.* at 1266.

speakers of a language (with shared linguistic conventions and a shared understanding of different linguistic registers), and we necessarily interpret the utterance according to the shared linguistic conventions that exist at the time of the utterance.

If we are to have a theory of ordinary meaning that tracks the way we actually use and interpret language, we cannot artificially separate out formal and pragmatic considerations. Literalist semantic meaning alone is not an indication of ordinary communicative content. Real human beings do not derive meaning from dictionary definitions and rules of grammar alone. Everyone takes nonsemantic context—pragmatics—into account in deriving meaning from language.[125] And for that reason we see no basis to credit semantic meaning without consideration of pragmatic context.[126] If no lawmaker would read the text that is voted into law purely semantically—devoid of pragmatic context—then there is no reason to credit that kind of meaning as a means of vindicating the intent of a lawmaker. The same goes for the public governed by the law. If no one reads

---

125. As Lawrence Solum has noted:

> The word "pragmatic" is ambiguous. In contemporary legal theory, "pragmatism" refers to an antifoundationalist approach that is strongly associated with Judge Richard A. Posner. Legal pragmatism is related to the philosophical pragmatism that is associated with philosophers Professor John Dewey, Professor William James, and Charles Sanders Peirce. As used in the philosophy of language and theoretical linguistics, pragmatics is a technical term with a contested and evolving meaning.

Lawrence B. Solum, *Originalist Methodology*, 84 U. CHI. L. REV. 269, 286 n.58 (2017) (citations omitted). Often in linguistics, pragmatics is used to refer to the study of specific linguistic phenomena like conversational implicature or deixis. ALAN CRUSE, MEANING IN LANGUAGE: AN INTRODUCTION TO SEMANTICS AND PRAGMATICS 332-37, 355-94 (2d ed. 2004). But pragmatics is also "sometimes defined as being concerned with the role or effects of context." *See* Solum, *supra*, at 286 n.59. Unless otherwise specified, we use the term pragmatics in this paper to refer to nonverbal context that may affect meaning.

126. As we will demonstrate below, a significant amount of contextual information may be derived from corpus data. In this respect, the corpus can provide insight into the context of an utterance. Yet there are some aspects of nonverbal context that are harder to derive from corpus data. These include notions of contextual enrichment like implicature, impliciture, presupposition, and modulation. *See* Solum, *supra* note 125, at 288-91 (discussing types of contextual enrichment). The use of linguistic corpora to analyze these linguistic phenomena has only recently begun to be explored by linguists. *See* Christoph Rühlemann & Karin Aijmer, *Introduction: Corpus Pragmatics: Laying the Foundations*, in CORPUS PRAGMATICS: A HANDBOOK 1, 1 (Karin Aijmer & Christoph Rühlemann eds., 2015) ("For a long time pragmatics and corpus linguistics were regarded as 'parallel but often mutually exclusive.' However, in recent years corpus linguists and pragmaticists have actively begun exploring their common ground." (citation omitted)).

laws literally by pure semantics, we have no reason to protect reliance interests or notice concerns rooted in that kind of understanding.

This does not mean that it is never worthwhile to consider the formal aspects of an utterance. Humans do take verbal, semantic context into account in interpreting language. It is just that humans also take nonverbal, pragmatic context into account. In developing a more robust theory of ordinary meaning, we think it important to highlight each of these elements of context that might affect our understanding—and to clarify the determinants of ordinary meaning that our law might seek to measure.

Before we turn to that endeavor, however, we first finish our treatment of Fallon's taxonomy by addressing "intended" meaning. Is the law's search for meaning aimed at finding the "public" meaning inferred by a "reader" of the law or a more private "intended" sense of a lawmaker? On this we agree with Baude and Sachs. "There may be good reasons for a legal system to prefer" either public meaning or intended meaning.[127] And "neither has to win every time," because the "right" answer "depends on our reasons" for the resort to ordinary meaning "in the first place."[128]

Intended meaning is an appropriate construct to the extent we are aiming to vindicate the preferences of lawmakers. This is a viable, distinct basis for crediting ordinary meaning. We may say, as does Professor Larry Alexander, "that the reason we should seek the actual authors' intended meaning is that the actual authors possessed the legal authority to promulgate norms, and their texts just *are* their communications of the norms they intended to promulgate."[129] If that is our premise for looking to the ordinary communicative content of the law then we will certainly look to the intended meaning of lawmakers (informed by relevant elements of context, as presented below). Even the reader, at least arguably, would seek this meaning.[130] But if we are to seek the intended meaning of the authors of the law, we must have some objective means of doing so.

---

127. Baude & Sachs, *supra* note 3, at 1091.

128. *Id.* at 1090.

129. Larry Alexander, *Telepathic Law*, 27 CONST. COMMENT. 139, 140 (2010).

130. *See* Larry Alexander, *Originalism, the Why and the What*, 82 FORDHAM L. REV. 539, 540 (2013) (asserting that "our job is to determine the uptake the legislator(s) intended us to have"). We also agree with Professor Ryan D. Doerfler, however, that legislative intent is ultimately a fiction—not only because "Congress is a 'they,' not an 'it,'" or because "Members of Congress . . . share no . . . intention to treat as authoritative the views of a statute's 'principal sponsors' or 'others who worked to secure enactment,'" but also because language must be understood in light of context consisting of "information salient to both author and audience." Ryan Doerfler, *Who Cares How Congress Really Works?*, 66 DUKE L.J. 979, 982–83 (2017). This suggests that the line between intended meaning and public meaning is thin or perhaps nonexistent, a point we return to below. *See infra* Section III.C.1.

There is also a case for the public or "reader's" understanding. This sort of meaning makes sense to the extent we are seeking to vindicate the notice rationale for the "standard picture"—the protection of reliance interests and the avoidance of unfair surprise.[131] Enforcing "hard-to-find intentions" of lawmakers "would make the law unpredictable or arbitrary."[132] So to the extent our search for ordinary meaning is aimed at protecting these interests, we should seek to assess the public's understanding of the law at the time it was passed.

In summary, before framing the theory of meaning in a manner that may allow us to measure it, we must first delineate the components of such meaning. At a broad level, those components encompass semantic meaning and pragmatic meaning. To assess meaning, linguists would tell us that we must also take into account the relevant speech community (*whose* meaning?) and the relevant timeframe (meaning *as of when*?). We explore each of these components below.

A. Semantic Meaning

Semantic meaning encompasses several components: lexicography, syntax, and semantics.

1. Lexicography

The search for "semantic" meaning often distills to a question of *word sense*. In *Muscarello*, *Taniguchi*, and *Costello*, for example, the courts were considering a problem of competing word senses—senses numbered separately from each other in the cited dictionaries.

Judges tend to assume that a dictionary's division of senses (by numbers and letters) represents an immutable linguistic fact about the universe. We tend to "ignore the fact," as Professor Larry Solan has put it, "that someone sat there and wrote the dictionary, and we speak as though there were only one dictionary,

---

131. *See* Michael Herz, *Purposivism and Institutional Competence in Statutory Interpretation*, 2009 MICH. ST. L. REV. 89, 102 (arguing that "the case for textualism" is in part "[t]he claim . . . that if legal rules are embedded in publicly available texts, affected persons will be able to know, understand, and comply with those rules . . . . [T]he fair notice argument for textualism in statutory interpretation presupposes, and seeks to ensure the full benefit of, a shift from the common law to statutes"); Note, *Textualism as Fair Notice*, 123 HARV. L. REV. 542, 542 (2009) ("Perhaps the most intuitive and straightforward argument for textualism is that it promotes fair notice of the law."); *see also* SCALIA, *supra* note 116, at 17 (asserting that it is "incompatible with democratic government, or indeed, even with fair government, to have the meaning of a law determined by what the lawgiver meant, rather than by what the lawgiver promulgated").

132. Baude & Sachs, *supra* note 3, at 1091.

whose lexicographer got all the definitions 'right' in some sense that defies analysis."[133] But that is not the case. Dictionaries may differ sharply in the number of senses they assign to a given term or in the divider they use to distinguish senses. "And human beings, try as they may, bring their prejudices and biases into the dictionaries they make."[134]

The question of "what is a word sense" turns out to be a very challenging one in lexical semantics. Linguists and lexicographers lack "decisive criteria for defining word senses and clearly discriminating between them."[135] And linguists also acknowledge that the sense distinctions reflected in dictionaries are "more of a descriptive device rather than a claim about psycholinguistic reality."[136]

In traditional lexicography, words are defined first by determining the class of things to which they belong (their *genus*) and second by distinguishing them from all other things in their class (their *species*).[137] Words are then divided into senses based on a variety of factors, including their part of speech, pronunciation, inflection, etymology, and shades of meaning.[138] This approach to defining words and dividing them into senses can be highly impressionistic and has a number of limitations. There is no agreed-upon formula for sense division—some lexicographers make very fine-grained distinctions between senses (they are sometimes called *splitters*), while others tend to make broader, more coarse-grained distinctions (they are sometimes called *lumpers*).[139] Moreover, the citation or quotation files from which many dictionaries are derived were collected without the benefit of modern sampling methods. Accordingly, these files cannot be relied upon for information about the frequency of a given word or word sense.[140]

133. Lawrence Solan, *When Judges Use the Dictionary*, 68 AM. SPEECH 50, 50 (1993) ("[O]ur society's reverence for dictionaries is not driven by the latest discoveries in psycholinguistic research. Rather, it is deeply embedded in our culture.").

134. JONATHAN GREEN, CHASING THE SUN: DICTIONARY MAKERS AND THE DICTIONARIES THEY MADE xiv (1996).

135. Nikola Dobrić, *The Predictive Power of the (Micro)Context Revisited—Behavioral Profiling and Word Sense Disambiguation*, 57 ZBORNIK MATICE SRPSKE ZA FILOLOGIJU I LINGVISTIKU 77 (2014).

136. Stefan Th. Gries, *Polysemy*, in HANDBOOK OF COGNITIVE LINGUISTICS 482 (Ewa Dabrowska & Dagmar Divjak eds., 2015); *see also* Dylan Glynn, *Polysemy and Synonymy: Cognitive Theory and Corpus Method*, in CORPUS METHODS FOR SEMANTICS: QUANTITATIVE STUDIES IN POLYSEMY AND SYNONYMY 7, 10 (Dylan Glynn & Justyna A. Robinson eds., 2014).

137. SIDNEY I. LANDAU, DICTIONARIES: THE ART AND CRAFT OF LEXICOGRAPHY 153 (2d ed. 2014).

138. BO SVENSÉN, PRACTICAL LEXICOGRAPHY: PRINCIPLES AND METHODS OF DICTIONARY-MAKING 204–05 (John Sykes & Kerstin Schofield trans., 1993).

139. Elizabeth Walter, *Using Corpora To Write Dictionaries*, in THE ROUTLEDGE HANDBOOK OF CORPUS LINGUISTICS 433-34 (Ann O'Keeffe & Michael McCarthy eds., 2010).

140. LANDAU, *supra* note 137, at 153.

Contemporary lexicographers have moved past relying on citation files alone and have begun to rely on electronic "corpora" — large bodies or databases of naturally occurring language — to gather linguistic data.[141] Corpus analysis has allowed lexicographers to address the problem of sense division with greater granularity. Lexicographers can now view a more complete range of potential uses of a given word and collect statistical information about the likelihood of a given word appearing in a particular semantic environment.[142]

Linguistic corpora allow us to make observations about the way that language is (and was) used through a less arbitrary and more readily measurable methodology than resort to dictionaries. And because language is the output of the mind, it is reasonable to assume that we can learn something about the way the mind perceives language by examining the way language is used in natural language environments. As mentioned above, there may be other ways to measure the way that language utterances are perceived.[143] Linguists in other linguistic disciplines use a variety of experimental methods to account for human perception of sense and meaning. But for now, our focus is on corpus linguistic analysis.

The challenge of sense division can be illustrated by reference to competing senses of *carry* in *Muscarello* or the alternative notions of *interpreter* in *Taniguchi*. Lexicographers may disagree about where to draw the lines between senses of these terms, or whether the two alternatives are distinct from each other.[144] But if the question of the dividing line is in some sense arbitrary and not reflective of the way in which the mind perceives and interprets language, then we ought to seek to measure — to the extent it is possible to measure — whether the mind perceives a sense distinction between two occurrences of the same word, rather than relying on the sense-divisions in the dictionaries before us.

This problem is most acute as to two senses that are viewed as closely related to each other. The two notions of *interpreter* in *Taniguchi* are illustrative. We can

---

141. *See, e.g.*, OXFORD DICTIONARY OF ENGLISH xi (3d ed. 2010) ("The general principle on which the senses in the Oxford Dictionary of English are organized is that each word or part of speech has at least one core sense or core meaning, to which a number of subsenses may be attached. . . . Core meanings represent typical, central uses of the word in question in modern standard English, as established by analysis of the Oxford English Corpus and our other language databases."). Below we will discuss in greater detail the nature of the language databases in question and why the language in those databases is properly characterized as "naturally occurring."

142. JOHN LYONS, LINGUISTIC SEMANTICS: AN INTRODUCTION 80 (1995) (noting that another way to think about word senses is as "the set, or network, of sense-relations that hold between [an expression] and other expressions of the same language").

143. *See supra* note 22.

144. *See* BIBER ET AL., *supra* note 100, at 40 (documenting the differences in the definition and sense distribution of the noun "deal" as recorded in five general-use dictionaries).

find distinct definitions encompassing *oral translator* and *written translator*. But that may not tell us how these two senses are perceived, or that one sense would be viewed as excluding the other. The notion of *oral translator* could simply be perceived as a more common "prototype" of the more general notion of "one who translates." The *written translator* idea could certainly be viewed as an *atypical* example. That may be all that dictionaries are telling us by indicating that *written translator* is "obsolete." And if so, that sort of obsolescence would not tell us that an ordinary person would not understand text providing for compensation for an *interpreter* to cover a *written translator*.

A dodo, after all, is an obsolete bird. But it is still a bird. And a person who happened to discover a remaining dodo on a remote island would certainly be understood to be in possession of a bird. Such a person would be covered, for example, by the terms of a rental agreement prohibiting tenants to keep "dogs, cats, birds, or other pets" in their apartments. If you are found in possession of a caged dodo, you are not likely to escape the wrath of the landlord by insisting that a dodo is an "obsolete" sort of a bird.

2. *Syntactic and Semantic Context*

The need to consider context is a staple element of the judicial inquiry into ordinary meaning. Courts often reference the notion of context when they invoke the ordinary meaning canon.[145] Yet they rarely say what they mean by context. Linguistic theory can help identify which elements of context may matter, and thereby offer some discipline for what has been haphazard judicial practice.

Context can be viewed as encompassing both verbal and non-verbal components.[146] The verbal context of a word or phrase in a statute includes its *syntactic* and *semantic* environments. Syntax is a set of rules and principles that governs sentence formation and determines which sentences will convey meaning to

---

145. *See* Johnson v. United States, 559 U.S. 133, 139 (2010) (observing that in the ordinary meaning inquiry, "[u]ltimately, context determines meaning"); Chisom v. Roemer, 501 U.S. 380, 404 (1991) (Scalia, J., dissenting) (arguing that the "regular method for interpreting the meaning of language in a statute" was to "first, find the ordinary meaning of the language in its textual context"); Braunstein v. Comm'r, 374 U.S. 65, 70 (1963) (defining the ordinary meaning of "gain" in a particular context). We are referring to linguistic context, which is a somewhat different concept than seeing if the statutory context precludes turning to ordinary meaning altogether. *See, e.g.*, Gonzales v. Carhart, 550 U.S. 124, 152 (2007) ("In interpreting statutory texts courts use the ordinary meaning of terms unless context requires a different result.").

146. Charles Goodwin & Alessandro Duranti, *Rethinking Context: An Introduction*, in RETHINKING CONTEXT: LANGUAGE AS AN INTERACTIVE PHENOMENON 1, 6-9 (Alessandro Duranti & Charles Goodwin eds., 1992).

members of the same speech community.[147] One aspect of syntax is *argument structure*, a linguistic term of art that refers to the participants in the action of a verb.[148] A transitive verb, like *carry*, has two arguments—the subject and the object. If we are interested in examining the meaning of phrases like *carries a firearm*, we would look for phrases that have a similar argument structure. And those arguments may affect our understanding of the meaning of *carry*.

Semantic context may also affect our perception of meaning. *Semantics* is the study of meaning at the word or phrase level.[149] Embedded within the words and phrases we use are a number of concepts that are sometimes referred to as the semantic features or semantic components of a word.[150] These features include concepts like number, animacy, gender, humanness, and concreteness (i.e., tangibleness).[151] In semantic theory, words can also be understood with reference to their functional role. A word has an *agentive* function if it is an instigator of the action of a verb, or an *objective* function if it is the entity that is affected by the action of the verb. A word may also serve an *instrumental* function if it is a force or object involved in, but not instigating, the action.[152]

An illustration of these linguistic concepts may be made by reference to the *Muscarello* statute's requirement of a mandatory minimum sentence for "anyone who . . . carries a firearm." Our understanding of *anyone who carries a firearm* is informed by the syntactic arguments—with *who* as the subject and *firearm* as the object of the verb. With respect to semantic features, we can characterize the relevant subject of the statute in *Muscarello* as *animate*, *human*, and *concrete*. Similarly, we can characterize the relevant object as *inanimate*, *concrete*, *non-human*, and even *weapon*. With respect to functional roles, *who* performs the agentive

---

147. 17 OXFORD ENGLISH DICTIONARY 487 (2d ed. 1989). Syntax is also the study of these rules and principles. NOAM CHOMSKY, SYNTACTIC STRUCTURES 11 (1957) ("Syntax is the study of the principles and processes by which sentences are constructed in particular languages.").

148. CRUSE, *supra* note 125, §§ 14.1-5, at 281-90 (3d. ed. 2011). An intransitive verb (*fall, die, yawn*) has one argument—the subject. *Id.* § 14.4.1, at 283. A di-transitive verb (*throw, send*) has three arguments—the subject, the direct object, and the indirect object. *Id.* § 14.4.3, at 287-88.

149. MICHAEL MORRIS, AN INTRODUCTION TO THE PHILOSOPHY OF LANGUAGE 152 (2006) ("Semantics is the attempt to give a systematic explanation of how the meaning of sentences depends upon the meaning of their parts.").

150. JOHN I. SAEED, SEMANTICS 260, 265 (4th ed. 2015).

151. *Id.*

152. Following Charles Fillmore, Alan Cruse lists a number of functional roles for words, including *agentive*—the instigator of the action; *instrumental*—the force or object involved in the action of the verb; *dative/experiencer*—the animate being affected by the action of the verb; *factitive*—the result of the action of the verb; *locative*—location or spatial orientation of the verb; and *objective*—the inanimate entity affected by the action of the verb. CRUSE, *supra* note 125, § 14.5, at 288-290 (3d. ed. 2011).

function, instigating the action of *carry*, and *firearm* serves an *objective* function. In the phrase *anyone who carries a firearm in a car*, *car* serves an *instrumental* function—it is involved in the action of the verb, but does not instigate it and does not receive it.

By looking to the argument structure and semantic features of the relevant statute, we are able to perform a more targeted search for language data to inform our inquiry into ordinary meaning. When we seek to measure language usage, we may wish to limit our search to uses of the verb *carry* that share the above-noted features—the syntax of a transitive verb, with the semantic features of a human subject and a weapon object.[153]

Limiting our search in this way will also inform the utility of information about frequency. We may find, for example, that the most frequent use of a given word is in fact quite rare in the syntactic and semantic context that is most similar to the statute. For example, we could find that the most common use of *carry* is to *carry on one's person*, but that the more prevalent use is different in the context of a human agent *carrying* an inanimate, weapon object—there, the *carry in a car* meaning may be most prevalent. Thus, when we search for ordinary meaning, we ought to be looking for the most common use of a given word in the contexts that are most similar to that of the statute in question.

B. Pragmatic Meaning

The meaning of an utterance will not always be expressly communicated in its semantic content. Non-verbal (pragmatic) contextual considerations will also be taken into account.[154] Such considerations may encompass the physical or social setting of an utterance, and even an inference about the intent of the speaker.[155]

---

153. *See* Muscarello v. United States, 524 U.S. 125, 143 (1998) (Ginsburg, J., dissenting) (noting that the issue presented "is not 'carries' at large but 'carries a firearm'").

154. *See* CRUSE, *supra* note 125, at 347; Goodwin & Duranti, *supra* note 146, at 6-9. Pragmatics includes concepts like conversational implicature, where the meaning of an utterance is strongly implied but not expressly stated—as where a spouse who says "there sure are a lot of dishes in the sink" is not just making an observation about the state of the universe but is reminding somebody about whose turn it is to do the dishes.

155. Careful scholars have recognized this point, and they have identified it as a basis for concluding that the space between textualism and intentionalism is small. *See* SCALIA, *supra* note 116, at 144 (conceding that "what the text would reasonably be understood to mean" and "what it was intended to mean" are concepts that "chase one another back and forth to some extent, since the import of language depends upon its context, which includes the occasion for, and hence the evident purpose of, its utterance"); Larry Alexander & Saikrishna Prakash, *"Is That English You're Speaking?": Why Intention Free Interpretation Is an Impossibility*, 41 SAN DIEGO L.

Judge Richard Posner's "Keep off the grass" problem is a good illustration. As Posner notes, a sign in a park that says "'Keep off the grass' is not properly interpreted to forbid the grounds crew to cut the grass."[156] Our understanding of the meaning of this sign is informed by more than just its semantic and syntactic content. We understand it in light of its pragmatic context, which includes inferences about the place and manner of the utterance and presumed intentions of the speaker.

Pragmatic considerations are of relevance to any attempt to assess the ordinary meaning of a statutory phrase. An utterance that merely *describes* a person carrying a firearm might be understood to convey one ordinary meaning. But a criminal prohibition—more precisely, a requirement of a mandatory minimum criminal sentence—may be understood differently. At least that is possible, and we may need to take such context into account in assessing ordinary meaning.

Pragmatic context may also inform the utility of frequency information, just as semantic and syntactic considerations do. In searching for the ordinary meaning of a given word, we might focus our search for the most frequent uses of the word that occur in contexts that share similar physical or social features to those represented in the statute. The more frequently a given use of a word occurs in circumstances that reflect a physical and social setting similar to that of the statute, the more confidence we should have that the use in question is the ordinary meaning of the word in that context.

C. Meaning as of When?

Human language is in a constant state of change.[157] But it does not change at a predictable rate.[158] Nor do different linguistic features change at the same

---

REV. 967, 979 (2004) ("[T]he commonplace truth that all understandings of texts are contextual just demonstrates that all texts *qua* texts acquire their meaning from the presumed intentions of their authors.").

156. *See* RICHARD A. POSNER, HOW JUDGES THINK 40 (2010) ("[Pragmatism] refers to basing judgments (legal or otherwise) on consequences, rather than on deduction from premises in the manner of syllogism.").

157. JOHN LYONS, INTRODUCTION TO THEORETICAL LINGUISTICS 43 (1968) ("All languages are subject to constant change. This is an empirical fact . . . . All living languages . . . are of their nature efficient and viable systems of communication serving the different and multifarious social needs of the communities that use them. As these needs change, languages will tend to change to meet the new conditions.").

158. TERRY CROWLEY & CLAIRE BOWERN, AN INTRODUCTION TO HISTORICAL LINGUISTICS 149-51 (4th ed. 2011) (discussing criticisms of attempts to quantify the rate of language change).

time.[159] A compelling theory of ordinary meaning must take account of this variation and allow us to examine the linguistic norms prevailing at different historical periods.

Lawyers and judges are used to thinking about timeframe in constitutional interpretation. There we often acknowledge that *original meaning* may differ from *modern meaning*. But we often ignore the problem in statutory interpretation. Or sometimes we just assume it away. In *Costello*, for example, Judge Posner noted that the *harboring an alien* statute was enacted in 1917, but looked for modern data as to the ordinary sense of the verb *harbor*.[160] Perhaps he did so out of convenience or necessity, given that his Google search framework would not have allowed historical analysis. But his stated reason reflected the reality of much litigation over statutory interpretation: the parties simply didn't bother to consider the possibility that the term *harbor* may have evolved over time—both sides presented dictionary definitions from modern times—so Judge Posner appears to have concluded that this gave him license to do the same thing.

That phenomenon is sometimes reflected in our theory of statutory interpretation. At least a few courts have looked to the ordinary meaning of a statute *as of the time it was enacted*.[161] That approach seems appropriate to the extent we are seeking "intended" meaning. This is the point of the originalists who argue for the vindication of intended original meaning—that the "ratifiers of the Constitution . . . are the persons with authority to make and change constitutional norms," and thus that contemporary interpreters (and citizens) are bound by their views.[162] If we "'interpret' the Constitution as if it had been authored by someone other than its ratifiers," these originalists argue, we are "mak[ing] constitutional 'law' without authority to do so."[163] The same point can be made as to statutes. If intended meaning is the relevant construct, we must be bound by meaning as of the time of the statute's initial enactment. Otherwise, we are vindicating intentions at other times and by other people.

The "public" meaning construct could encompass either contemporary or historical meaning. If we are seeking to protect contemporary reliance interests and fair notice, we should arguably be seeking contemporary (not historical) evidence of ordinary meaning. "Normal" English speakers are guided by their con-

---

159. *Id.*

160. United States v. Costello, 666 F.3d 1040, 1043-44 (7th Cir. 2012).

161. *See, e.g.*, Carcieri v. Salazar, 555 U.S. 379, 388 (2009); Norfolk S. Ry. Co. v. Perez, 778 F.3d 507, 512 (6th Cir. 2015); State Bd. of Nursing v. Ruebke, 913 P.2d 142, 157 (Kan. 1996); State v. Ziska, 334 P.3d 964, 967 (Or. 2014).

162. Alexander, *supra* note 129, at 141.

163. *Id.*

temporary understanding, and they lack a sophisticated understanding of historical usage. So if we are trying to protect those interests we should arguably be seeking contemporary public meaning. But that is not the only way to think about public meaning. The premises of originalism can also be understood to seek to protect *original* public meaning—to preserve the public meaning fixed at the time of adoption or ratification.

Frequency considerations may also be applied to theories of both the "intended meaning" and "public meaning" of historical texts. If we are looking for what the ratifiers of the Constitution intended a particular word to mean, we might search for the most common way that word was used in texts drafted by the ratifiers (taking into account similarities in syntactic, semantic, and pragmatic context). If, on the other hand, we are looking for the original public meaning of a word or phrase in the Constitutional text, we might look to how that word was most commonly used by the public at large in similar contexts.

If we seek to measure historical meaning, how can we do so? One common means of assessing *historical* ordinary meaning is to consult an old dictionary. That is an approach that courts often take in seeking the original meaning of the Constitution. But that practice is fraught with all of the difficulties highlighted above as to contemporary dictionaries: historical dictionaries, just like their contemporary peers, cannot yield reliable information about which of various senses is more ordinary. Indeed, the problems are compounded for historical dictionaries.[164]

This is not to say that historical dictionaries do not have value. Historical dictionaries can be useful for defining unknown terms and attesting contested uses.[165] But we ought to regard them with skepticism when they are offered as evidence of "ordinary" or "original" meaning.

---

164. *See* Rickie Sonpal, *Old Dictionaries and New Textualists*, 71 FORDHAM L. REV. 2177, 2209-10 (2003) ("Supreme Court Justices are sometimes very scrupulous about choosing the dictionary and edition with a publication date close to the date the statute was enacted; yet, this practice is often of deceptively limited value. This practice is of even less value when old dictionaries are used because some popular older dictionaries were not only reprinted but even appeared in new editions without any substantive change to the body of the dictionary. . . . Accordingly, judges who carefully choose the printing or edition of an old dictionary that is most closely contemporary with the statute risk relying on a dictionary the substance of which far antecedes the statute." (footnotes omitted)).

165. The same is true for general-use, unabridged dictionaries. They can be useful for defining unknown terms, showing the range of potential meanings, or attesting contested meanings.

D. Whose Meaning?

Our understanding of meaning is also shaped by our speech community—the group of people with whom we share a set of linguistic norms, conventions, and expectations about linguistic behavior.[166] Meaning may also vary across different linguistic registers—varieties of texts, ranging from spoken communications, to newspapers, academic prose, or even congressional committee reports that tend to share linguistic features. Our theory of ordinary meaning must be able to account for the speech community we are evaluating and address the differences in various linguistic registers.

Limiting our search to a given speech community or register will have an important effect on the usefulness of information about frequency. It would not be unusual to find that a use of a word that is common in one speech community or register is quite rare in another. By limiting a search for ordinary meaning to the relevant speech community and register in question, we can have greater confidence that information about the frequency of use of a given word is telling us something useful about ordinary meaning.

The choice between "public" meaning and "intended" meaning may have implications for our identification of the relevant speech community. The public-meaning construct seems to dictate a speech community consisting of a broad cross-section of the public. The intended-meaning inquiry, on the other hand, could at least arguably point to a more limited community. Members of Congress are generally not common, ordinary people and their usage of certain words may not be colloquial. So, if our search for ordinary meaning is aimed at deriving intended meaning, we may wish to assess the usage or understanding of a more sophisticated group of English speakers. We may also wish to take into account

---

166. *See, e.g.*, Marcyliena H. Morgan, Speech Communities: Key Topics in Linguistic Anthropology 1 (2014) ("Speech communities are groups that share values and attitudes about language use, varieties and practices. These communities develop through prolonged interaction among those who operate within these shared and recognized beliefs and value systems regarding forms and styles of communication."); Reed Dickerson, *Statutory Interpretation: Dipping into Legislative History*, 11 Hofstra L. Rev. 1125, 1154 (1983) (defining *speech community* as the "group of people who share a common language (or sublanguage) and thus a common culture (or subculture), which in turn defines the context that conditions the utterances that occur within it" (footnote omitted)); John Sinclair, *Meaning in the Framework of Corpus Linguistics*, 20 Lexicographica 20, 22 (2004) ("The differences in interpretation between members of a speech community are small and they do not interfere much with normal communication."); Kamal K. Sridhar, *Societal Multilingualism*, *in* Sociolinguistics and Language Teaching 47, 49 (Sandra Lee McKay & Nancy H. Hornberger eds., 1996) ("A conglomeration of individuals who share the[] same norms about communication is referred to as a speech community. A *speech community* is defined as a community sharing a knowledge of the rules for the conduct and interpretation of speech.").

the pragmatic consideration that the more formal nature of legal language can affect human understanding of meaning.[167]

III. OPERATIONALIZING ORDINARY MEANING

The above sets the stage for a more careful formulation of the law's assessment of the ordinary communicative content of the language of the law. A compelling theory of ordinary meaning recognizes that we may choose to measure either public meaning or intended meaning. And however we choose to frame the inquiry, we should account for all of the relevant semantic, pragmatic, temporal, and speech-community considerations.

That leaves the question of measurement or operationalization. We propose the use of tools employed in corpus linguistics. Corpus linguistics is an empirical approach[168] to the study of language that involves large, electronic databases of text known as corpora (the plural of corpus).[169] A corpus is a body or database of naturally occurring language.[170] Corpus linguists draw inferences about language from data gleaned from "real-world" language in its natural habitat—in books, magazines, newspapers, and even transcripts of spoken language.[171] The

---

167. *But see* Doerfler, *supra* note 130, at 983-84 (articulating a "conversation" model of "fictionalist" legislative intent in which "[a]n interpreter occupies the position of conversational participant, hearing statements directed at her and other participants" and credits "information salient both to members of Congress and to citizens").

168. PAUL BAKER ET AL., A GLOSSARY OF CORPUS LINGUISTICS 65 (2006) ("In linguistics, empiricism is the idea that the best way to find out about how language works is by analysing real examples of language as it is actually used. Corpus linguistics is therefore a strongly empirical methodology."); TONY MCENERY & ANDREW HARDIE, CORPUS LINGUISTICS: METHOD, THEORY AND PRACTICE 49 (2012) ("Empiricism lies at the core of corpus linguistics . . . .").

169. *See* MCENERY & HARDIE, *supra* note 168, at 1-3.

170. *See* Douglas Biber, *Corpus-Based and Corpus-Driven Analyses of Language Variation and Use*, *in* THE OXFORD HANDBOOK OF LINGUISTIC ANALYSIS 159, 159 (Bernd Heine & Heiko Narrog eds., 2010) ("Corpus linguistics is a research approach that has developed over the past several decades to support empirical investigations of language variation and use, resulting in research findings that have much greater generalizability and validity than would otherwise be feasible . . . . [I]t utilizes a large and principled collection of natural texts, known as a 'corpus,' as the basis for analysis . . . .").

171. *Id.* at 160-61 ("[Corpus linguistics] depends on both quantitative and qualitative analytical techniques . . . . [T]he major contribution of corpus linguistics is to document the existence of linguistic constructs that are not recognized by current linguistic theories. Research of this type—referred to as a 'corpus-driven' approach—identifies strong tendencies for words and grammatical constructions to pattern together in particular ways, while other theoretically possible combinations rarely occur . . . . [C]orpus-based research investigates the patterns of variation among the full set of spoken and written registers in a language. In speech, these

defining characteristic of corpus linguistics is "the claim that it is possible to actually 'represent' a domain of language use with a corpus of texts, and possible to empirically describe linguistic patterns of use through analysis of that corpus."[172] Through corpus analysis we can test our hypotheses about language through rigorous experimentation with observable and quantifiable data. And the results of a corpus-based conclusion will be replicable and falsifiable.[173]

Corpus data can tell us the relative frequency of different senses of *vehicle* (or of *carrying* a firearm, of *interpreter*, or of *harboring* an alien) in naturally occurring language.[174] And if the search for ordinary meaning entails analyzing the relative frequency of competing senses of a given term, then corpus linguistics seems the most promising tool.[175]

Corpus data can also help us resolve different types of linguistic uncertainty in the interpretation of legal texts.[176] We can use corpus data to address questions of vagueness, where "a word or phrase has borderline cases."[177] The scope

include casual face-to-face conversation, service encounters, lectures, sermons, political debates, etc.; and, in writing, these include e-mail messages, text-messaging, newspaper editorials, academic research articles, etc." (citation omitted)).

172. DOUGLAS BIBER & RANDI REPPEN, THE CAMBRIDGE HANDBOOK OF ENGLISH CORPUS LINGUISTICS 1 (2015).

173. *See* MCENERY & HARDIE, *supra* note 168, at 66 ("As a key goal of corpus linguistics is to aim for replicability of results, data creators have an important duty to discharge in ensuring that the data they produce is made available to analysts in the future.").

174. Assuming, of course, the corpora used are properly constructed such that they enable us to make generalizations about a larger population. *See generally* Douglas Biber, *Representativeness in Corpus Design*, 8 LITERARY & LINGUISTIC COMPUTING 243 (1993) (addressing a number of issues related to achieving "representativeness" in linguistic corpus design).

175. One problem in finding ordinary meaning is the problem of modulation, the idea that "a conventional semantic meaning can be adjusted or modulated to fit the context — essentially, a new meaning is created (sometimes on the spot) so that an old word is used in a new way." Solum, *supra* note 125, at 290. It is possible that corpus data establishing the prevailing use of a given word in a given period could also be used to triangulate instances of modulation. It is similarly possible that historical corpus data could be used to trace circumstances in which modulation resulted in new coinages with important legal implications. *See id.* (arguing that the "Constitution contains a variety of modulations"). We do not address these phenomena here, however.

176. *See id.* at 286 n.60 ("Lawyers sometimes use the words 'ambiguity' and 'vagueness' interchangeably to refer to a lack of clarity."); Lawrence M. Solan, *Pernicious Ambiguity in Contracts and Statutes*, 79 CHI.-KENT L. REV. 859, 860 (2004) ("When discussing indeterminacy in meaning, linguists and philosophers often distinguish between ambiguity and vagueness. . . . Legal writers, and judges in particular, use the word 'ambiguity' to refer to all kinds of indeterminacy, whatever their source. Because this Article focuses heavily on what judges say, I will generally use the word ambiguity in this looser, legal sense.").

177. Solum, *supra* note 125, at 286.

of the term *vehicle* in the *no vehicles* ordinance is perhaps an example of vagueness. We can also use corpus data to address questions of ambiguity, where a word or phrase has more than one potential meaning in a given context.[178] *Muscarello* seems to fit here; the choice between the competing senses of *carry* is largely a question of ambiguity.

Such data can also inform our assessment of linguistic prototype.[179] If the corpus data reveal that most *vehicles* that we speak of are automobiles, or that most instances of *carrying* a firearm involve bearing it on your person, we may infer that those senses are more likely to be prototypical senses of the operative terms.

Below we drill down further on the proposed means of measurement. First we present linguistic tools and means of measuring the components of ordinary meaning identified above. We then illustrate the utility of those tools by applying them to the cases and examples discussed throughout the Article. We conclude this Part with some observations about inferences that can be drawn from the data about the ordinary meaning of *vehicle*, *carry a firearm*, *interpreter*, and *harbor*.

A. Tools

Corpus linguistic tools can be employed to measure ordinary meaning as conceptualized in this Article. Here we explore the range of available corpora and the functionalities they encompass.

1. Varieties of Linguistic Corpora

Linguistic corpora come in a number of varieties, each tailored to suit the needs of a particular set of empirical questions about language use. Corpora may be *general* or *special*. A *general* corpus endeavors to represent the language used by a broad (often national) speech community. *Special* corpora are limited to a

---

178. *See id.* (defining ambiguity as "cases in which a word or phrase has more than one sense.").

179. Intuitively, we might assume that frequency and prototype would map onto one another with some precision, but this is not always the case. *See* John R. Taylor, *Prototype Theory*, *in* 1 SEMANTICS: AN INTERNATIONAL HANDBOOK OF LANGUAGE MEANING, 643, 649-50 (Claudia Maienborn et al. eds., 2011) ("In response to the question 'where does prototypicality come from?', many people are inclined to say that prototypes (or prototypical instances) are encountered more frequently than more marginal examples and that that is what makes them prototypical. Although frequency of occurrence certainly may be a factor (our prototypical vehicles are now somewhat different from those of 100 years ago, in consequence of changing methods of transportation) it cannot be the whole story." (citation omitted)).

particular genre, register, or dialect.[180] There are *monitor* corpora that are continuously updated with new texts in order to track contemporary language use, and there are *historical* or *sample* corpora that reflect the language use of a particular period. We will rely on both monitor and historical corpora in the analysis below.

Corpora may also be *raw*, *tagged*, or *parsed*. A *raw* corpus contains almost no linguistic metadata (e.g., a .txt file containing the complete works of Shakespeare would be a raw corpus). *Tagged* corpora typically contain metadata from a grammatical "tagging" program that automatically marks each word with a part of speech. A tagged corpus can dramatically improve corpus analysis by allowing a researcher to look for all different forms of a single word in a single search (e.g., a search for the verb *carry* would automatically include every verb inflection, including *carries, carrying,* and *carried*) and to limit results to a particular part of speech (e.g., the verb *harbor*, not the noun *harbor*). This type of search is called a *lemmatized* search—a search for the base form of a word that reveals its permutations. *Parsed* corpora contain phrase-, clause-, or sentence-level annotation, revealing the syntactic relationships among the words in the corpus. While automated tagging is highly accurate, automated parsing is not. Thus, parsed corpora tend to require a significant amount of human editing and annotation, which increases the costs of their production dramatically. For that reason, parsed corpora tend to be smaller than tagged corpora. The corpora we rely on in the analysis below are tagged, but not parsed.[181]

2. *Corpus Tools—Frequency, Collocation, and Key Word in Context*

Linguistic corpora can perform a variety of tasks that cannot be performed by human linguistic intuition alone. For example, as noted above, corpora can be used to measure the statistical *frequency* of words and word senses in a given speech community and over a given time period.[182] Whether we regard the ordinary meaning of a given word to be the *possible*, *common*, or the *most common*

---

180. For example, the Linguistic Data Consortium at the University of Pennsylvania produced a corpus of recorded Egyptian Arabic telephone calls. *See* Alexandra Canavan et al., *CALL-HOME Egyptian Arabic Speech*, Linguistic Data Consortium (1997), http://catalog.ldc.upenn.edu/LDC97S45 [http://perma.cc/P6NQ-MUT5].

181. The corpora relied on in this paper were tagged by the Constituent Likelihood Automatic Word-tagging System (CLAWS-7) program. Mark Davies, *The 385+ Million Word* Corpus of Contemporary American English *(1990-2008+): Design, Architecture, and Linguistic Insights*, 14 Int'l J. Corpus Linguistics 159, 164 (2009).

182. Tony McEnery & Andrew Wilson, Corpus Linguistics: An Introduction 82 (2d ed. 2001).

sense of that word in a given context, linguistic corpora allows us to determine empirically where a contested sense of a term falls on that continuum.

Corpora can also show *collocation*, "which is the tendency of words to be biased in the way they co-occur."[183] As we have seen, words are often interpreted according to the semantic environment in which they are found. A collocation program can show the possible range of linguistic contexts in which a word typically appears and can provide useful information about the range of possible meanings and sense divisions.[184]

Corpora also have a *concordance* or *key word in context* ("KWIC") function, which allows their users to review a particular word or phrase in hundreds of contexts, all on the same page of running text. This allows a corpus user to evaluate words in context systematically.

Commonly accepted canons of interpretation like *ejusdem generis* and *noscitur a sociis* already counsel legal interpreters to look for meaning in the surrounding linguistic context of an utterance – to know a word by the company it keeps.[185] The data made available through a linguistic corpus allows one to make such inquiries systematically, and to gain meaningful and quantifiable insight about the range of possible uses of a word and the frequency of its different senses.[186]

3. Representing Speech Community and Register in a Corpus

Linguistic corpora can be built from the ground up using text or speech from any given speech community or register. As Professor Larry Solan has noted:

> When the legal system decides to rely on the ordinary meaning of a word, it must also determine which interpretive community's understanding it wishes to adopt. This choice is made tacitly in legal analysis, but becomes overt when the analysis involves linguistic corpora because the software displays the issue on a screen in front of the researcher.[187]

---

183. SUSAN HUNSTON, CORPORA IN APPLIED LINGUISTICS 68 (2002); *see also* John R. Firth, *A Synopsis of Linguistic Theory, 1930-1955, in* STUDIES IN LINGUISTIC ANALYSIS 1, 14 (1957) ("Collocations are actual words in habitual company.").

184. HUNSTON, *supra* note 183, at 69.

185. ESKRIDGE, *supra* note 5, at 76-78 (discussing the *ejusdem generis* and *noscitur a sociis* canons).

186. Early discussions of collocation analysis used language very similar to the *noscitur a sociis* canon. *See, e.g.*, Firth, *supra* note 183, at 11 ("You shall know a word by the company it keeps!").

187. Solan, *The New Textualists' New Text*, *supra* note 50, at 2059 (footnotes omitted).

In this Article we rely on a pair of corpora of standard written American English (one contemporary and one historical). But a corpus can be constructed to represent the language use of a wide variety of speech communities or registers.[188]

One possibility worth highlighting is that of a distinct *legal* corpus. Some of the language of the law, of course, is written in a distinct legal dialect.[189] Where a given term is thought to be a legal term of art, a legal corpus could be built to analyze its meaning in the legal vernacular. Such a corpus could be employed to compare the ordinary sense of a given term and its legal term-of-art usage.

4. *Representing Historical Language Use*

Finally, a linguistic corpus can be built from texts representing the language use from any period in history. To the extent our understanding of ordinary meaning should be informed by the linguistic norms and conventions prevailing at the time that a given legal text was drafted, corpus linguistics can provide powerful evidence of historic language use.

5. *The BYU Corpora*

Below we will tackle the interpretive problems posed by the *Muscarello*, *Taniguchi*, and *Costello* cases using data from two linguistic corpora: the News on the Web ("NOW") Corpus and the Corpus of Historical American English ("COHA"), both developed at Brigham Young University and referred to here as the BYU Corpora. Here we outline the parameters of each corpus and highlight their differences.

a. *NOW Corpus*

The NOW Corpus is a database of "5.2 billion words of data from web-based newspapers and magazines from 2010 to the present time."[190] It is a monitor corpus that "grows by about 5-6 million words of data each day (from about 10,000

---

188. While corpora vary in size and sophistication, anyone can build a corpus using freely available software like AntCorGen. *See AntCorGen*, LAURENCE ANTHONY'S WEBSITE, http://www .laurenceanthony.net/software/antcorgen [http://perma.cc/NJV9-5JVP].

189. *See generally* McGinnis & Rappaport, *supra* note 4, at 4 (asserting that the Constitution is written in the "language of the law," not ordinary English).

190. *NOW Corpus (News on the Web)*, BYU, http://corpus.byu.edu/now [http://perma.cc/UTD2 -BC86].

new articles), or about 150 million words each month."[191] The NOW Corpus downloads content every night from dozens of websites listed on Google News, using an automated software program.[192] These texts are then automatically tagged and lemmatized (adding part-of-speech metadata to each word) and integrated into the existing corpus.[193] Because of this extraordinary rate of growth, the NOW Corpus is currently the largest tagged corpus of English in the world.

"[T]here is no data like more data,"[194] and the chief virtues of the NOW Corpus are its size and immediacy. With the NOW Corpus, the user is able to examine what is happening in the language at the moment. And because of the size and scope of the corpus, lower frequency linguistic phenomena (words, word senses, syntactic structures, etc.) are more likely to be attested, while the distribution of higher frequency phenomena will be better and more completely represented.

The NOW Corpus has a few limitations. First, even with searches limited to U.S. sources, the NOW Corpus records the language use of a single, large speech community (the United States) in a single linguistic register (newsprint). But if the interpretation of a federal statute requires us to consider the linguistic norms and conventions of the citizens subject to that statute, then U.S. newsprint may be the appropriate speech community and register. Spoken dialects of American English show sharp (and increasing) differences in vocabulary, grammar, and phonology,[195] but the norms and conventions of the written variety of American English (sometimes called standard written American English) tend to be more uniform. Since we are interpreting a written text, evaluating that text through the lens of standard written American English (from newsprint) may be the right approach.

The NOW Corpus is also limited with respect to timeframe. NOW tracks the linguistic norms and conventions over the past decade. So if we want to evaluate interpretive problems against the backdrop of linguistic norms prevailing at the enactment of the Constitution, we will need to turn elsewhere.

---

191. *Id.*

192. *Id.*

193. *Id.*

194. Kenneth Church, *Has Computational Linguistics Become More Applied?*, *in* COMPUTATIONAL LINGUISTICS AND INTELLIGENT TEXT PROCESSING 1, 3 (Alexander Gelbukh ed., 2009) (internal quotation marks omitted) (attributing the statement to Robert Mercer).

195. *See* WILLIAM LABOV, DIALECT DIVERSITY IN AMERICA: THE POLITICS OF LANGUAGE CHANGE 1–2 (2012).

b. *Corpus of Historical American English ("COHA")*

The COHA is "the largest structured corpus of historical English."[196] It contains "more than 400 million words of text from the 1810s-2000s (which makes it 50-100 times as large as other comparable historical corpora of English) and the corpus is balanced by genre decade by decade."[197] Using data from the COHA, we can gather linguistic information from the decade that a statute was enacted, going back approximately 200 years.

Like the NOW Corpus, the COHA is limited in terms of speech community and register. Though it has texts from a wider variety of registers than the NOW Corpus (including fiction, magazines, and non-fiction), these tend to fall within the ambit of standard written American English. In addition, the 400 million words of the COHA are spread out over 200 years. Consequently, the COHA is essentially a collection of twenty separate corpora (one for each decade from 1810 to 2010) averaging just over twenty million words).[198]

There is a lot of linguistic information to be gleaned from a twenty-million-word corpus. But as we have seen, in the specialized setting of statutory interpretation it is important to evaluate words in context. These contexts may be poorly represented (or not represented at all) in the corpus with limited data for a given period. The earliest texts in the COHA date from the period of 1810 to 1820. These texts come in twenty to thirty years shy of the Founding Era, leaving us without a data source for the prevailing linguistic norms during the drafting and ratification of the Constitution.[199]

---

196. *Corpus of Historical American English*, BYU, http://corpus.byu.edu/coha [http://perma.cc/N44U-NQ8T].

197. *Id.*

198. For a breakdown of the total number of words in the COHA for each decade, go to http://corpus.byu.edu/coha and click on "400 million words." Note that any searches performed will normalize frequency measurements in words-per-million, so that measurements of statistical frequency over multiple decades will not be adversely affected by differences in the number of words in the corpus for each decade.

199. There are good reasons for this omission. Prior to the 1806 publication of Noah Webster's influential text, *A Compendious Dictionary of the English Language*, American spelling was very much in disarray, with many common words having as many as a half dozen potential spellings. This makes the construction of a corpus interface and the automated tagging of corpus data very difficult (and expensive). Moreover, because of widely varied orthographic practices, many historical texts are difficult, if not impossible, to subject to optimal character recognition ("OCR"). BYU Law School is seeking to fill this gap. It is currently working on a Corpus of the Founding Era American English ("COFEA"). *See Law & Corpus Linguistics Conference*, BYU LAW: LAW & CORPUS LINGUISTICS, http://lawcorpus.byu.edu [http://perma.cc/S256-N8FQ].

With all of that said, the COHA remains the largest corpus of historical American English and it contains significant linguistic information relevant to the statutes at issue in *Muscarello*, *Taniguchi*, and *Costello*.

B. Applications

1. Vehicles in the Park

The "no vehicles" problem seems a mandatory subject for any serious treatment of statutory interpretation. It was introduced initially by Professor H.L.A. Hart[200] in his famous debate with Professor Lon Fuller,[201] but seemingly everyone has treated the problem since then.[202] There is also no shortage of extensions of the hypothetical. Hart says that "[p]lainly" the rule "forbids an automobile," but asks "about bicycles, roller skates, toy automobiles" and airplanes.[203] The airplane example invokes an actual case—*McBoyle v. United States*,[204] in which the U.S. Supreme Court held that an airplane was not a vehicle under the National Motor Vehicle Theft Act, which prohibited transporting stolen "vehicles" across state or national borders.[205]

The scholars cited throughout this Article have offered their own views on the scope of "vehicle." Justice Scalia and Garner's *Reading Law* says that the Hart prohibition should extend to any "*sizable* wheeled conveyance," and thus to automobiles—including "ambulances, golf carts, mopeds, motorcycles, and (perhaps) Segways"—but not "remote-controlled model cars, baby carriages, tricycles, or perhaps even bicycles."[206] Professor Fallon objects to the extension to ambulances. He says the "reasonable meaning" of vehicle should not be understood to extend to ambulances—at least those responding to emergencies.[207] Professor William Eskridge disagrees with Justice Scalia and Garner's suggestion as to bicycles; he says that "bicycles are commonly considered vehicles," a conclusion he claims to confirm using corpus data.[208]

---

200. Hart, *supra* note 41, at 607.

201. Lon L. Fuller, *Positivism and Fidelity to Law—A Reply to Professor Hart*, 71 HARV. L. REV. 630, 662-69 (1958).

202. *See, e.g.*, ESKRIDGE, *supra* note 5, at 45-46; SCALIA & GARNER, *supra* note 28, at 36-39; Fallon, *supra* note 2, at 1260-62.

203. Hart, *supra* note 41, at 607-08.

204. 283 U.S. 25, 26-27 (1931).

205. 18 U.S.C. § 2312 (2012).

206. SCALIA & GARNER, *supra* note 28, at 37-38.

207. *See* Fallon, *supra* note 2, at 1260-61.

208. ESKRIDGE, *supra* note 5, at 45-46.

Yet there has been very little attempt to assess the ordinary meaning of *vehicle* with any data. We present some relevant data below, concerning the frequency or prototypicality of various senses of this term.

a. Lexical Collocation of Vehicle Through Time

One way to examine the most common context in which a word appears is collocation. The collocation function of the corpus can show us the words that are statistically most likely to appear in the same context as *vehicle* for a given period. We can use collocation to get a snapshot of the semantic environment in which *vehicle* appears and the kinds of vehicles that tend to appear in that environment.

We can view the most common contemporary collocates[209] of *vehicle* in the NOW Corpus.[210] In NOW, the fifty most common collocates of *vehicle* are as follows:

> *electric, motor, plug-in, unmanned, armored, connected, cars, aerial, charging, pure, launch, owners, hybrid, traffic, fuel, driving, gas, autonomous, struck, operating, road, safety, accidents, battery, ownership, emergency, batteries, emissions, seat, advanced, driver, primary, demand, gmv, commandeered, fuel-efficient, uavs, automakers, demonstrators, excluding, lunar, passenger, fleet, gasoline, luxury, drove, parking, retirement, vehicles, infrastructure*[211]

Many of the collocates of *vehicle* in the NOW Corpus strongly indicate *automobile* as a likely candidate for the most common use of the term. The NOW Corpus lists a number of automotive collocates like *motor*, *car*, *traffic*, *fuel*, *driving*,

---

209. The NOW Corpus and other BYU corpora are available without a subscription. To access NOW, go to http://corpus.byu.edu/now. To generate a list of collocates in NOW, take the following steps: (1) Select "Collocates" on the NOW Corpus homepage; (2) Enter "VEHI-CLE_n" in the "Word/phrase" field (capitalization makes the search lemmatized—assuring that we find all inflections of the word; the "_n" is to limit the search to noun forms); (3) Enter an asterisk "*" (a wildcard) in the "Collocates" field; (4) Select "Sections" and select "United States" in column "1" (ignoring column "2"); (5) Select "Sort/Limit" and set the "Minimum" to "MUT INFO"; and (6) Click "Find collocates."

210. The following link will reproduce the search above, except that the user would need to repeat step four, select "Sections," and select "United States," which doesn't repopulate automatically. *See NOW Corpus (News on the Web)*, BYU, http://corpus.byu.edu/now/?c=now&q=54596 680 (last visited Mar. 8, 2017).

211. The search results are saved at the following link. *See NOW Corpus (News on the Web)*, BYU, http://corpus.byu.edu/now/?c=now&q=52902048 (last visited Dec. 21, 2016).

gas, battery, batteries,[212] *emissions, driver, fuel-efficient, automakers, gasoline, drove,* and *parking*. It also includes more recent automotive collocates of *vehicle* like *electric, plug-in, connected, charging,* and *hybrid*. Some of the collocates by themselves have a range of possible uses (*owners, operating, safety, accidents, ownership, emergency, seat, primary, infrastructure*), but when examined in context almost always indicate an automotive meaning.[213] *Airplane* does not appear, though two particular types of aircraft are attested in the collocates—unmanned aerial vehicles (drones) and spacecraft.[214] Similarly, *bicycle* does not appear among the collocates of *vehicle* in contemporary usage.

We can also examine the collocates of *vehicle* during the 1950s, the decade of the Hart/Fuller debate, in the COHA. These collocates are listed below:

> *motor, space, trucks, moving, wheeled, tax, self-propelled, passenger, unit, tracked, orbit, test, b.g., launching, highways, tanks, license, robot, emergency, units, taxes, streets, equipment, manned, armored, vehicles, fees, vehicle, traveling, operate, loaded, fuel, commercial, driver, ride, traffic, designed, weight, speed, cars, carrying, operation, unsafe, horse-drawn, high-powered, amphibious, administrators, tactical, registration, delivery*[215]

We can see from this data that the meaning of *vehicle* has evolved significantly from the 1950s, though the automotive use of *vehicle* predominated then as well. The decade is remarkable as the first in which the spacecraft sense of *vehicle* appears, but also the last in which the *horse-drawn* collocate of *vehicles* appears.[216] *Unmanned* does not appear, but *manned* vehicle does (*spacecraft* in this case). Still, the overwhelmingly most common use of *vehicle* is the *automotive* sense, while a number of context-specific possible senses are attested. Again, none of the top fifty collocates of *vehicle* include the notions of *airplane* or *bicycles*.

---

212. Collocates are not lemmatized in the BYU corpora, so the singular and plural form of a given collocate are counted separately. In this case, that means that *battery* and *batteries* both make the list separately.

213. To the extent that there is any doubt that any of these collocates suggest the automotive meaning of *vehicle*, clicking on any of the listed collocates in the NOW Corpus interface will display the context in which it appears and confirm the automotive meaning is intended. For example, it is possible to speak of *bicycle traffic* or *airplane emissions*, but in the context of the word *vehicle*, the words *traffic* and *emissions* are used in the automotive sense.

214. *NOW Corpus (News on the Web), supra* note 211.

215. *See The Corpus of Historical American English*, BYU, http://corpus.byu.edu/coha/?c=coha&q =52600298 (last visited Dec. 21, 2016).

216. An additional *vehicle* is added to our collection with *amphibious vehicle*, and *tanks* makes an appearance again. Two collocates (the abbreviation *b.g.* for background, and *robot*) are the result of including science fiction screenplays in the corpus. In both cases the vehicles in question are spacecraft.

We can also use the COHA to examine the collocates of *vehicle* from the period relevant to the *McBoyle* case. Because the statute at issue in *McBoyle* was enacted in 1919,[217] and because the COHA only allows us to search in ten-year increments, it may make sense to include data from 1910 through 1930.

Whether or not the use of the word *vehicle* "evoke[s] in the common mind only the picture of vehicles moving on land," as Justice Holmes suggests, may not be a question that can be addressed with a corpus.[218] But the collocate data from this period (consistent with the collocate data above) allow us to draw a similar inference that the automotive use is the most common use of *vehicle*, and that the *airplane* sense remains unattested:

> *motor, horse-drawn, wheeled, horses, pedestrians, kinds, expression, driver, passing, moving, various, horse, automobiles, tax, heavy, drawn, carry, roadless, rickety, trucks, communication, approaching, traffic, electric, mental, physical, 3,500,000, astral, belonging, steam, transportation, commissioner, rear, total, carrying, propulsion, propelled, oncoming, carriages, registration, ego, conceivable, tires, drivers, vehicle, carriers, 45, loaded, halted, manufacturers*[219]

The collocates from this period add a few interesting *vehicles* to our growing list, including *astral vehicle* (a reference to the theosophical notion of an "intermediate between the intelligent soul and the mental body, composed of a subtle material").[220] It should also be noted that only a few of the collocates in this period occur more than once, and only four—*motor, horse-drawn, wheeled*, and *horses*—occur ten times or more, with *motor* occurring twice the number of times as the other three combined.[221]

From the collocates of *vehicle* displayed by the NOW Corpus and the COHA, we can make the following preliminary observations (observations that we can later confirm by reviewing KWIC data). First, the collocates of *vehicle* strongly suggest that the most common use of *vehicle* is with reference to automobiles. Second, the absence of *airplane* and *bicycle* in the top fifty collocates of *vehicle*

---

217. *See* McBoyle v. United States, 283 U.S. 25, 25-26 (1931) (citing the National Motor Vehicle Theft Act, 18 U.S.C. § 408 (1928)).

218. *Id.* at 27.

219. *See The Corpus of Historical American English*, BYU, http://corpus.byu.edu/coha/?c=coha&q =53847214 (last visited Feb. 8, 2017).

220. *See Astral Body*, WIKIPEDIA, http://en.wikipedia.org/wiki/astral_body [http://perma.cc /R98L-A57F]. This notion also explains the presence of *ego* and *mental* in the collocates of *vehicle*.

221. *See supra* note 219.

raises an important question for our frequency continuum.[222] If we accept that the necessary and sufficient conditions of *vehicle* are "[a]ny means of carriage, conveyance, or transport"[223] or "a means of carrying or transporting something,"[224] then there seems little question that both an *airplane* and a *bicycle* are *possible* readings of *vehicle*. But if *vehicle* is never used to refer to *bicycle* or *airplane* in the corpus data, then we may end up with an even further extension of our frequency continuum from *possible but rare* to *possible but unattested*. Before jumping to the conclusion that the *airplane* and *bicycle* uses of *vehicle* are entirely unattested in the corpora or the language at large, however, we should evaluate the use of *vehicle* in the concordance data.

b. Vehicle as a KWIC

We can extract concordance data from the NOW Corpus.[225] A NOW search for concordance lines of *vehicle* will yield an output along these lines:

---

222. It is worth noting that while the words *airplane* and *plane* do not appear among the collocates of *vehicle* during any of the timeframes examined above, there are, at least, a number of aircraft terms that do appear, including reference to two specific types of aircraft: *spacecraft* and *unmanned aerial drones*. It is not clear why the more general terms *airplane* and *plane* do not appear in the collocate data for *vehicle*, while terms like *spacecraft* and *unmanned aerial drone* do appear.

223. 19 OXFORD ENGLISH DICTIONARY 480 (2d ed. 1989).

224. WEBSTER'S THIRD NEW INTERNATIONAL DICTIONARY 2538 (1961).

225. The concordance line search in NOW is executed as follows: (1) Select "KWIC" on the NOW Corpus homepage; (2) Enter "VEHICLE_n" in the "Word/phrase" field; (3) Click on "Sections" and select "United States"; (4) Click "Keyword in Context (KWIC)." *See NOW Corpus (News on the Web)*, BYU, http://corpus.byu.edu/now/?c=now&q=54499369 (last visited Mar. 5, 2017). By selecting "Options" and "# KWIC," the corpus user can select the number of randomized concordance lines to be reviewed. While the search parameters can be saved in a link, the corpus randomizes the results, and, in the case of the NOW Corpus, the corpus updates with millions of new words on a nightly basis. Therefore, until the BYU corpora develop the ability to save the exact content of a particular randomized search, it is useful to copy the results of the search into a spreadsheet.

TABLE 1.
KWIC OF *VEHICLE*[226]

the driver, Bhaskar Jha, apparently lost control of the	vehicle	because he was traveling too fast for the wet road conditions.
of the troopers. Parrott says the suspects in the	vehicle	began showing aggression and shots rang out. Corporal Shane
injury and leaving a child under 12 unsupervised in a motor	vehicle	but released on a written promise to appear.) Risk
Hybrid electric vehicles use regenerative braking (when the	vehicle	captures energy that would be otherwise lost from braking) and
pushed onto the property because of the speed of which these	vehicles	collide," said Dr. Tom Lawrence, of Clinical Nutrition
, 2009. That day the two officers saw a	vehicle	connected to a domestic violence case in which shots had been
say automakers would be better. Wakefield says autonomous	vehicles	could erode the image of certain brands more than others. Brands
biogas, and Daimler, which supplies a number of experimental	vehicles	designed to run on natural gas. The German Federal Ministry of
is that they aren't kept on file with the Motor	Vehicle	Division or any other entity. By contrast, beneficiary

The KWIC output in the NOW Corpus allows us to select anywhere from one hundred to one thousand randomized sample uses of *vehicle(s)* and display them in their semantic environment. To the extent that the snippet view above fails to provide sufficient evidence of usage, the corpus interface allows us to click through to an expanded passage from the article referenced in a given concordance line.

In order to examine the sense distribution of *vehicle*, we reviewed one hundred randomized concordance lines of *vehicle* in the NOW Corpus. Of those, ninety-one were *automobiles*. There was a single reference to a bus, and one reference to an ambulance, but in every other instance, a passenger car was referenced. Of the remaining *vehicles,* there was one cargo ship, one jet ski, and an ambiguous reference to a military ground vehicle of an unknown type. There were three metaphorical uses of *vehicle* (e.g., the role of the city as a vehicle for

---

226. Note that these concordance lines have been shortened to fit this page. Moreover, because the content of the corpus is constantly updated, and because the results of the search are typically randomized, a search performed on any given day will provide a different data set.

development). The results also contained a reference to the military's efforts to create a flying Humvee/helicopter hybrid.[227]

The NOW Corpus data included no *airplanes*, *bicycles*, *tricycles*, *skateboards*, *roller-skates*, *toy cars*, or any of what Hart and others have characterized as penumbral, disputed cases. To the extent that our notion of ordinary meaning has a frequency component, this data suggests that *automobile* is overwhelmingly the most common use of the word *vehicle* in the modern written American English represented in the NOW Corpus. The corpus data also suggest that there are numerous possible (if much less common) uses of *vehicle*, and that some seemingly *possible* meanings are unattested and may not be current.

A similar review of data from the COHA for the 1950s showed a wider range of *vehicles*. Still, approximately sixty-five percent of the usages of *vehicles* during this timeframe referred to *automobiles*. Another thirty percent referred to the space program or missile defense, while the remaining five percent referred to metaphorical uses of *vehicle* (e.g., a film as a starring vehicle for an actor). For the period spanning the 1910s and 1920s, *automobiles* made up approximately sixty percent of the instances of *vehicle*. References to carriages or horse-drawn *vehicles* were more common, and there were a significant number of cases where the choice between automobile and horse-drawn vehicle was not clear. (When a text from 1915 says that Fifth Avenue was crowded with vehicles, it is not clear from context whether automobiles, carriages, or both were intended.) Finally, there were a number of references to theosophy and the notion of an astral vehicle. The COHA data also included no references to *airplanes*, *bicycles*, *tricycles*, *skateboards*, *roller-skates*, or *toy cars* for either period.

c. Searching for Vehicles in the Context of a Park

Hart's interpretive puzzle is not simply about *vehicles* at large, but *vehicles in the park*. As we have discussed, with the corpus we can examine the question of ordinary meaning in the relevant semantic and syntactic context. We can search for *vehicles* that collocate with the term *park*.[228]

---

227. We are not making this up. *See Aerial Reconfigurable Embedded System*, WIKIPEDIA, http://en.wikipedia.org/wiki/Aerial_Reconfigurable_Embedded_System [http://perma.cc/9HVS-82R9].

228. (1) Select "Collocates" on the NOW Corpus homepage; (2) Enter "VEHICLE_n" in the "Word/phrase" field; (3) Enter "PARK_n" in the "Collocates" field; (4) Select "Sections" and select "United States"; (5) Select "Sort/Limit" and set the "Minimum" to "FREQUENCY" and "15"; (6) Click "Find collocates"; and (7) Click "PARK" or "PARKS." *See NOW Corpus (News on the Web)*, BYU, http://corpus.byu.edu/now/?c=now&q=63434268 (last visited Mar. 5, 2017).

A review of the concordance data from this search reveals at least one limitation of the corpus. We can search for specific parts of speech (e.g., nouns, verbs), but not specific senses. That means that our search for *vehicles in the park* must begin by eliminating the approximately forty percent of the concordance lines that refer to *vehicles* that are *in park* as opposed to *in reverse*, *neutral*, or *drive*. Approximately five percent of the concordance lines refer to recreational vehicles in recreational vehicle parks. Of the remaining instances of vehicle, more than fifty percent refer specifically to *automobiles*. *Bicycles* are not attested in this context, nor are *airplanes, skateboards,* or *roller skates*.

Our understanding of a prohibition on *vehicles in the park* may depend largely on the physical and spatial characteristics of the park itself. If a municipal park has no means of ingress or egress for automobiles, then we might assume that cutting across the grass in a car would be prohibited. It is not surprising then that where municipal parks are concerned, the *vehicle* most likely to show up in the context of *park* in the corpus data (i.e., automobiles) is often not *in the park*, as in (1) and (2) below:

> (1) *juvenile[s] were taken into custody Wednesday, accused of discharging a BB gun at passing vehicles <u>near Sunset Park</u>.*
> (2) *two males in another <u>vehicle near a park</u> on Toledo's west side when one of those males opened fire*

In the very rare circumstance in which there is any actual debate about vehicles in municipal parks, such debates tend to center around closing off existing roads through the park, as in (3) below:

> (3) *A revived plan to remove vehicle traffic from the center of San Diego's Balboa Park was moved forward Monday by the City Council, which agreed to spend $1 million to complete planning and documentation.*

Yet even in the specific *park* context, where the physical and spatial features of a park might seem to preclude the entrance of an automobile, it is the *automobile* usage of *vehicle* that predominates.

d. Is Bicycle a Vehicle? Is Airplane a Vehicle?

We can use the KWIC function of the corpus to perform targeted searches for concordance lines featuring two key terms raised in the Hart/Fuller debate — *bicycle* and *airplane*.[229]

Professor Eskridge has asserted that "[a] corpus search reveals that bicycles are commonly considered vehicles — a quantitative result in striking contrast to the understanding advanced by linguist Bryan Garner, who joined Justice Scalia in opining that the ordinary meaning of 'vehicles' excludes bicycles."[230] Professor Eskridge is certainly correct that there are numerous instances of the co-occurrence of *bicycle* with *vehicle*. Some of these instances establish that the *bicycle* sense of *vehicle* is, at the very least, attested, as in (1) and (2) below:

> (1) *There are a lot of potholes. It is hard to ride bicycles <u>and other vehicles</u>.*
> (2) *In New Jersey, <u>bicycles are considered vehicles</u> and must follow the same laws as motorists.*

Yet other instances show that *bicycle* is often used in contrast to the word *vehicle*, as in (3) and (4) below:

> (3) *there were 68 collisions between bicycles, pedestrians and vehicles*
> (4) *side mirrors to detect hazards (bicycles, humans, vehicles, pets, etc.)*

Based on the corpus data reviewed above, *bicycle* is certainly a possible sense of *vehicle*, but from the standpoint of statistical frequency, it is not a common meaning and certainly not the most common.

With respect to the use of *vehicle* to reference *airplane*, the answer is simpler. In both the contemporary NOW Corpus and the COHA (for the relevant periods of the 1910s, 1920s, and 1950s), we were unable to find a single collocation or concordance line that reflected the use of *vehicle* to mean *airplane*. *Vehicle* is neither most commonly used nor even commonly used when discussing airplanes, and based on its absence from any of our corpus data, we might ask if *airplane* is even a possible sense of *vehicle*. To the extent that *airplane* fits what some lexicographers have regarded as the necessary and sufficient conditions for

---

229. (1) Select "Collocates" on the NOW Corpus homepage; (2) Enter "VEHICLE_n" in the "Word/phrase" field; (3) Enter "BICYCLE_n" in the "Collocates" field; (4) Select "Sections" and select "United States"; (5) Select "Sort/Limit" and set the "Minimum" to "MUT INFO" and "3"; (6) Click "Find collocates"; and (7) Click "BICYCLES." *See NOW Corpus (News on the Web)*, BYU, http://corpus.byu.edu/now/?c=now&q=54497865 (last visited Mar. 5, 2017).

230. ESKRIDGE, *supra* note 5, at 45-46.

inclusion in the class of *vehicles* (i.e., anything that is a "means of carriage, conveyance, or transport"), all that can be said of *airplane* is that it may be a possible meaning of *vehicle*, but it is unattested in the corpus data.

Thus, corpus linguistics can advance the theory of ordinary meaning by, on the one hand, allowing judges and lawyers to more specifically define what they mean by *ordinary* and, on the other hand, providing objective data illustrating the way in which words are used in particular contexts.

2. Muscarello *and Carries a Firearm*

The *Muscarello* question—of the meaning of *carry*—is likewise susceptible to measurement. We can assess the relative frequency of the *personally bear* sense and the *transport* sense using corpus analysis.

a. The Collocates of Carry

We can view collocation data for *carry* in the NOW Corpus. The fifty most common collocates of *carry* in the NOW Corpus are listed as follows:

> *out, yards, concealed, weight, gun, attacks, weapons, guns, sentence, weapon, exchange, maximum, margin, passengers, heavy, penalty, bag, signs, opinions, firearm, express, burden, permit, thoughtful, load, bags, plane, firearms, virus, tradition, flag, capable, torch, handgun, cargo, openly, permits, duties, pipeline, mosquitoes, touchdowns, ships, executions, loads, trucks, felony, tasks, handguns, experiments, knife*[231]

These collocates suggest that a number of uses of *carry* do not fit neatly into the syntactic structure and the semantic relationships we have previously identified. There are instances in which an inanimate object serves as *carry*'s subject (*planes carrying passengers*, *trucks carrying loads*, *ships carrying cargo*). There are also a number metaphorical uses of *carry* (*felonies carrying certain penalties*, *people carrying opinions*). There are also references to *carrying out of attacks* and *executions*, and sporting references (*carrying the ball for so many yards* or *for so many touchdowns*).[232]

---

231. Follow the same steps set forth in note 209, substituting "CARRY_v" for "VEHICLE_n." *See NOW Corpus (News on the Web)*, BYU, http://corpus.byu.edu/now/?c=now&q=54015027 (last visited Feb. 15, 2017).

232. We can see similar results in the COHA using the same instructions in note 231, except that when we click on "Sections" we select "1960." The results of this search in the COHA can be viewed at the link below. *See Corpus of Historical American English*, BYU, http://corpus.byu

Yet "[a]t issue here is not 'carries' at large, but 'carries a firearm.'"[233] And a list of collocates simply tending to show that there are a variety of small, inanimate, concrete objects (including weapons) that can be carried on your person or in your car does not get us much closer to determining which of these senses of *carry* is the most frequent.[234] But as we will see, this search reveals common collocates of *carry* that have similar semantic features to *firearm* (i.e., *pistol, handgun, rifle, gun*) that will help us better evaluate the contexts in which *carry a firearm* occurs.

b. Carry as a KWIC

The KWIC data give us a clearer picture of the use of *carry*.[235] The NOW Corpus gives us a randomized sample of concordance lines featuring *carry*,[236] and we can review these concordance lines to determine both the range of possible meanings of *carry* and the comparative frequency of those meanings. We can also locate (and determine the comparative frequency) of instances of *carry* with the same syntactic and semantic features as § 924(c)(1).

Yet we might be able to eliminate a lot of irrelevant uses of *carry* by searching instead for *carry* within a few words of *firearm*.[237] A search for concordance lines containing these terms will require coding. Because every interpretative question is different, the process of coding concordance lines will vary with each task.

.edu/coha/?c=coha&q=54015512 (last visited Feb. 15, 2017). The COHA, when divided by decade, results in a functionally smaller corpus for that decade. As a consequence, it is more susceptible to being offset by unusual collocations.

233. Muscarello v. United States, 524 U.S. 125, 143 (1998) (Ginsburg, J., dissenting).

234. As we will see, in the case of *carry*, the collocates do help us in identifying words with similar semantic features as *firearm* — g*un(s)*, *weapon(s)*, *handgun(s)*, *rifle(s)*, *pistol(s)* — and that would serve similar functional roles in a sentence. This will help us locate relevant concordance data, but does not answer the question of which sense of *carry* is most common.

235. One way to examine carry in context is simply to enter a search similar to that in note 225, but substituting "CARRY_v" for "VEHICLE_n."

236. As we have already seen, *carry* has a transitive argument structure and, in the relevant context of § 924(c)(1), *carry* has a human subject and a non-human, inanimate, weapon object. *See supra* Section II.A.2. If we are going to take context into consideration, we should be looking for uses of *carry* that reflect the same or similar syntactic structure and semantic relationship.

237. Such a search can be executed as follows: (1) Select "Collocates" on the NOW Corpus homepage; (2) Enter "CARRY_v" in the "Word/phrase" field; (3) Enter "FIREARM_n" in the "Collocates" field; (4) Click on "Sections" and select "United States"; (5) Select "Sort/Limit" and set the "Minimum" to "FREQUENCY" and "3"; and (6) Click "Find collocates." *See NOW Corpus (News on the Web)*, BYU, http://corpus.byu.edu/now/?c=now&q =63434628 (last visited Feb. 15, 2017).

Here, after examining only a few concordance lines, a problem emerges: a significant majority of the instances of *carry* in the context of *firearm* in the NOW Corpus refer back to the statutory prohibition in § 924(c)(1) or similar statutes. In order to ensure that we have sufficient data from outside of a legal context, we also examined instances of *carry* in the context of a number of common synonyms of *firearm* listed among the most common collocates of *carry* — *gun(s)*, *pistol(s)*, *handgun(s)*, and *rifle(s)*. These synonyms share the same semantic features with *firearm*, but less commonly appear in statutory prohibitions against carrying a firearm.

Our search parameters eliminate a number of irrelevant uses of *carry*.[238] All that is left is to review the concordance lines and determine in how many instances *carry a firearm* refers to *carry on one's person* or *carrying in a car*. Here, the physical and spatial context can be helpful, as with the physical locations in (1), (2), and (3) below:

> (1) *Dressed in body armor and carrying two handguns, [the suspect] tried to flee out a back door . . . .*
> (2) *adults with the proper permits no longer need to hide the handguns they carry in their shoulder or belt holsters . . . .*
> (3) *the crowd was sedate and well-behaved with those carrying guns checking their ammunition at the door.*

A number of concordance lines were unclear, and a number of them, as noted, refer to statutory provisions similar to § 924(c)(1).

All told, we reviewed 271 concordance lines from the NOW Corpus in which *carry* co-occurred with *firearm(s)*, *gun(s)*, *pistol(s)*, *handgun(s)*, and *rifle(s)*. Of these instances of *carry*, we found that 104 instances indicated a sense of *carry a firearm on one's person*, while only five instances suggested a *carry a firearm in a car* sense. The remaining senses either were unclear (i.e., the appropriate sense could not be determined by context) or were senses of *carry* unrelated to the question at hand. As would be expected, much less data was available for *carry* in the COHA. We found twenty-eight concordance lines from the COHA, in which *carry* co-occurred with *firearm(s)*, *gun(s)*, *pistol(s)*, *handgun(s)*, and *rifle(s)*. Of these instances of *carry*, we found that eighteen were instances of *carry on one's person*, and two were instances of *carry in a car*. The remaining instances were either unclear or reflected a different sense of *carry*.

---

238. For example, the metaphorical sense (*carry a tune*), senses where the subject or agent is an inanimate object (*the ship carries cargo*), and the sporting sense (*carried the football nine yards*) were all eliminated.

To the extent that we view the question of ordinary meaning as involving statistical frequency, the analysis above tells us that carry on one's person is overwhelmingly the most common use, while carry in a car is a possible but far less common use.

3. Taniguchi *and the Meaning of Interpreter*

We can also measure the relative frequency of the *written translator* and *oral translator* senses of *interpreter*. We can do so using collocation and concordance analysis.

a. The Collocates of Interpreter

The fifty most common collocates of *interpreter* in the NOW Corpus are as follows:

> *an, through, language, sign, spanish, via, speaking, afghan, translators, iraqi, certified, served, english, qualified, translator, army, basic, deaf, spoke, moderator, sign-language, asl, costumed, interpreter, translate, full-time, dream, trained, soldiers, yun, interpreters, arabic, translated, translation, freelance, certification, courts, maladies, requests, spanish-language, communicate, cespedes, languages, troops, carlotto, simultaneous, somali, listened, proceedings, employed*[239]

A number of the collocates tend to support the *Taniguchi* majority's position that *interpreter* most commonly refers to an *interpreter of spoken language*. These include *speaking*, *spoke*, and *listen*. A number of the collocates refer to battlefield *interpreters* (such as *Afghan* or *Iraqi*),[240] where context would suggest their role is primarily as spoken interpreters. The collocates *an* and *through* both come from the very common phrase that a public figure is *speaking through an interpreter*. These collocates stand in contrast to the collocates of *translator* in the NOW Corpus, which make a number of references to the writing and publishing contexts, including *bible*, *writer*, *poet*, *editor*, *literary*, *publisher*, *journalist*, *Borders*,

---

239. *See NOW Corpus (News on the Web)*, BYU, http://corpus.byu.edu/now/?c=now&q =54018483.

240. We do not need to assume that the Afghan or Iraqi interpreters listed in the collocate display are battlefield interpreters. By clicking on each individual collocate in the display, we can view concordance lines—lines of running text showing the word in context. This expanded context feature shows a battlefield context for these interpreters in numerous instances.

and even *Wycliffe*.[241] The collocates from the 1970s, when the Court Interpreters Act was passed, suggest a similar conclusion.[242]

b. *Interpreter as a KWIC*

With respect to Key Words in Context, we reviewed 188 concordance lines from the NOW Corpus in which *interpreter* occurred.[243] In a number of instances, *interpreter* referenced an artistic expression or the interpretation of works of art (*a noted interpreter of modern music*). Another common sense refers to the interpretation of documents written in a primary language (*interpreters of the Constitution*). There were numerous instances in both corpora of cases of the spoken language conversion from a primary language to a second language notion of *interpreter*, as in (1) and (2) below:

> (1) *civil rights violations for not providing professional interpreters for patients who do not feel comfortable speaking English*
> (2) *Motto was speaking in French, through a volunteer interpreter*

In addition, there were numerous transcripts of spoken interviews from news sites with the annotation "through interpreter," referencing a spoken interview facilitated by an interpreter. There was one instance of an interpreter translating a foreign language document into spoken English, included below:

> (3) *In 1992, during a top-level meeting in Moscow, Russia finally released the cockpit voice recorder transcript. It was 10 p.m. in a dimly lit meeting room of the Presidential Hotel when an interpreter for the U.S. ambassador translated the Russian transcript into English for Ephraimson-Abt and other delegates.*

---

241. *See NOW Corpus (News on the Web)*, BYU, http://corpus.byu.edu/now/?c=now&q=546 09539 (last visited Mar. 9, 2017).

242. *Interpreter* has very few frequent collocates during the 1970s. The two most common collocates of *interpreter* from this period are *an* and *through*, function words that mutual information scoring typically eliminates if other options are available. A review of the concordance lines associated with these collocates reveals their origin in the extremely common phrase *speaking through an interpreter*, or related phrases. *See Corpus of Historical American English*, BYU, http://corpus.byu.edu/coha/?c=coha&q=54495283 (last visited Mar. 5, 2017).

243. The statute at issue in *Taniguchi* states: "A judge or clerk of any court of the United States may tax as costs the following: . . . (6) Compensation of interpreters . . . ." 28 U.S.C. § 1920 (2012). The noun phrase *compensation of interpreters* is part of a standalone enumeration that has an attenuated relationship to the argument structure of the verb *to tax*. What we can say about the relevant context for *interpreter* is that we are looking for individuals who are capable of decoding a foreign language into a native one. The operative variable is whether the language at issue is spoken or written.

Absent from all of these concordance lines was a single instance of anyone referred to as an *interpreter* performing a text-to-text translation from a foreign language into a primary language like English. To the extent that our notion of ordinary meaning has a frequency component, we can say from this data that the *text-to-text translation* sense of interpreter is neither the most common nor even a common use of *interpreter*. We might question whether it is even a possible sense of *interpreter* as the *text-to-text translator* sense of *interpreter* is entirely unattested in our data.

4. Costello *and Harboring an Alien*

The interpretive issue in *Costello* bears some similarity to the question at issue in *Muscarello*. In both cases the question turns on the meaning of a transitive verb and its relation to its object, though in the case of *harbor* our object has the semantic features of *human*, *animate*, etc. We would therefore look to the corpus data to tell us which senses of *harbor* are the most frequent, common, or possible senses of *harbor*, and to help us make informed decisions about sense division. We will look at the use of *harbor* in contemporary English, using the NOW Corpus, and in the decade 1910-1919, the period during which the relevant statute was enacted.[244]

a. Collocation of Harbor

With respect to the collocation data, it is immediately apparent from a review of the collocates of *harbor* that the overwhelmingly most common use of the term *harbor* refers to *harboring feelings*:

> *bacteria, feelings, resentment, doubts, terrorists, species, secret, mariners, views, ambitions, immigrants, fugitive, planets, illusions, hatred, dreams, cells, mutations, ocean, hopes, animosity, virus, secrets, anger, grudge, suspicions, fantasies, planet, fears, sentiments, desire, pathogens, galaxy, viruses, suspicion,*

---

244. *See NOW Corpus (News on the Web)*, BYU, http://corpus.byu.edu/now/?c=now&q=544 96834 (last visited Mar. 5, 2017). This search examines only the nominal (noun) collocates of *harbor*. *Harbor* is a low frequency verb and as such instances of *harbor* are rare in the COHA for the period of 1910-1919. Even expanding the search through the 1920s reveals only a sparse number of collocates. While some of these are relevant to our present inquiry (such as *alien* and *refugee*), no other relevant collocate appears more than once in the COHA. *See Corpus of Historical American English*, BYU, http://corpus.byu.edu/coha/?c=coha&q=54496926 (last visited Mar. 5, 2017).

persons, thoughts, fugitives, germs, mutation, tumors, aliens, moon, bias, genes, gene, hole, diversity, grudges, resentments[245]

This use of *harbor* does not match the semantic features in the relevant statute. We are looking for objects of *harbor* that are human, animate, concrete, etc. With that in mind, we tailored our searches to those nominal objects of *harbor* reflected in the collocates listed above that had these same semantic features — *fugitives*, *terrorists*, *criminals*, *aliens*, and *refugees*.

b. Harbor as a KWIC

In the NOW Corpus, we examined 140 concordance lines in which *harbor* occurred in the same environment as *fugitives*, *terrorists*, *criminals*, *aliens*, and *refugees*. Of these, twenty-three instances of *harbor* referred to *concealment* while thirty-two referred to *shelter*. In an additional eighty-three instances, the distinction could not be determined by context. There were also three instances of unrelated senses of *harbor*. In the COHA, there were only three clear-cut cases of the *shelter* sense. The remaining five instances of *harbor* could not be determined by context.

This data raises more questions than it answers. With respect to frequency, we would be hard-pressed to say that either the *shelter* meaning or the *conceal* meaning of *harbor* are the most common. We might say that both are common meanings, and they are both certainly possible and attested meanings. But where more than half of the instances of *harbor* are unclear as to whether they include *shelter* or *concealment* or both, it is hard to state from the standpoint of frequency what the ordinary meaning actually is.

C. Caveats and Conclusions

Such are the data. But what to make of them? Do corpus data yield means of measuring ordinary meaning? We think the answer is a resounding yes — with a few caveats. Certainly, the answer is yes by comparison with existing means of measurement. If ordinary meaning is an empirical construct — and we think it is — then corpus analysis is superior to an intuitive guess (or, worse, crediting a dictionary or a word's etymology).

We also think that corpus data are well suited to give reliable answers to the question of ordinary meaning. To support this conclusion (as applied to the problems analyzed throughout the Article), here we provide a more careful synthesis of the theory of ordinary meaning discussed above. We then offer some

---

245. *See id.*

conclusions about what the corpus data tell us about the ordinary meaning of *vehicle*, *carry a firearm*, *interpreter*, and *harbor*.

1. Caveats

Corpus analysis may be applied to the range of issues bearing on ordinary meaning identified above — to semantic context, pragmatic context, the temporal aspects of meaning, and speech community and register. Through data from the COHA, the NOW Corpus, or other corpora, we can assess the relative frequency of competing senses of a statutory term or phrase. From frequency and collocation data we can draw inferences about the semantic meaning of the language of the law and even about intended or public meaning. Yet we see some possible limitations on the strength of the inferences to be drawn from this sort of data.

a. Semantic Meaning

One possible limitation stems from the vagaries of word sense division. Sense division is subjective.[246] Linguists, as noted above, have no agreed-upon formula for distinguishing senses of a word.[247] They concede that distinctions among senses may be "more of a descriptive device rather than a claim about psycholinguistic reality."[248] This seems particularly true as regards closely related or fine-grained sense distinctions. The space between some senses will be sufficient to justify a strong inference from clear corpus data. Consider the above-cited example of the use of the term *nail* in *Reading Law*: "*Nail* in a regulation

---

246. Nikola Dobrić, *Word Sense Disambiguation Using ID Tags — Identifying Meaning in Polysemous Words in English*, *in* Proceedings of the 29th International Conference on Lexis and Grammar/LGC 97, 97 (Dusko Vitas & Cvetana Krstev eds., 2010) (explaining that polysemy — multiple word meaning — is "[o]ne of the persisting issues in modern lexicography").

247. No one is quite sure where to draw the line — research "show[s] that different polysemy criteria (i.e., criteria that may be invoked to establish that a particular interpretation of a lexical item constitutes a separate sense rather than just being a case of vagueness or generality) may be mutually contradictory, or may each yield different results in different contexts." Dirk Geeraerts, Theories of Lexical Semantics 196 (2009). And there is no agreed-upon taxonomy of polysemy. While some linguists speak of senses and subsenses, *see, e.g.*, Glynn, *supra* note 136, at 17, others speak of more or less prototypical exemplars of senses, *see, e.g.*, Dagmar Divjak & Antti Arppe, *Extracting Prototypes from Exemplars: What Can Corpus Data Tell Us About Concept Representation?*, 24 Cognitive Linguistics 221, 222-30 (2013).

248. Gries, *supra note* 136, at 482. The "problem of an apparent lack of decisive criteria for defining word senses and clearly discriminating between them has always been a burning issue of lexical semantics to the point that it fundamentally questions the possibility to provide a clear account of polysemy." Dobrić, *supra* note 135, at 78.

governing a beauty salon has a different meaning from *nail* in a municipal build-ing code."[249] Surely we could confirm that using corpus data. We could show that the term *nail* as used in the context of a beauty salon is almost always with reference to a fingernail or toenail. We would likely feel confident concluding that such data supports the conclusion that the ordinary understanding of *nail* in this semantic setting is not a piece of metal used to attach pieces of wood.

But what about more closely related senses? The two competing notions of *carry* in *Muscarello* are closely related. Both get at the idea of *transport*; the differ-ence concerns the mechanism—on one's person or in a vehicle. Accordingly, it seems hard to know whether this difference is reflected in the way that human beings perceive the different uses of *carry*. The *bear personally* sense seems to be the notion of *carry* that we speak of almost always, and for that reason it may also be the sense we think of most often. But if pressed, we might well concede that the *transport by vehicle* sense may be encompassed within the way in which we perceive the notion of carrying a firearm. It could be that most ordinary peo-ple first think of the *bear personally* sense but on reflection agree that the *transport* sense is included.

b. Pragmatic Meaning

Even with a very large corpus, some pragmatic information may be elusive—because the relevant physical or social setting is rare, for example, or the prag-matic information needed is not of the type that would appear in a corpus of written texts.

Consider the two illustrations referenced above: Posner's "Keep off the grass" sign at a park and Fallon's extension of the "no vehicles" rule. Posner rightly says that the park sign would "not properly [be] interpreted to forbid the grounds crew to cut the grass."[250] And Fallon understandably asserts that the lawmaker adopting the "no vehicles in the park" rule would "reasonably" be un-derstood to intend for the "gatekeeper" at the park to allow an ambulance to enter in the event of an emergency.[251]

We may be able to examine these questions from a corpus-based perspective. If we had a large enough database, that contained a sufficient number of park prohibitions (together with references to groundskeepers, ambulances, etc.), we might be able to draw conclusions about the pragmatic circumstances in which such prohibitions are most commonly invoked and how they are most com-monly interpreted. To find any ordinary exceptions to the "Keep off the grass"

---

249. SCALIA & GARNER, *supra* note 28, at 20.

250. RICHARD A. POSNER, REFLECTIONS ON JUDGING 180 (2013).

251. Fallon, *supra* note 2, at 1260-61.

853

or the "no vehicles" rules we might look for park owners who have these rules in place. If park owners and municipalities routinely allow ambulances into their parks or routinely allow groundskeepers access, we can infer something about how these prohibitions are ordinarily used or understood. The point is that corpus analysis often contains at least some pragmatic data and is at least theoretically capable of providing information about the pragmatic context. But there is no guarantee that even a very large and targeted corpus would contain sufficient examples of circumstances with similar pragmatic content. And the question for corpus linguistics is how much of the relevant pragmatic context is reflected in the formal record found in the corpus.[252]

In many cases such meaning may be beyond the reach of most corpora that are currently available. As to *Muscarello*, for example, it might be impossible to find a corpus sufficient to identify the pragmatic components of the intended meaning of a sentencing enhancement for carrying a firearm in connection with a drug crime. If we are looking at the question of the intended meaning of Congress, the right corpus may be one that would reflect dialogue among the 535 members that voted on the sentencing enhancement in § 924(c)(1). If we had such a corpus, and if it recorded extensive discussion among them about the kind of gun carrying they were talking about when they enacted this statute, we might be able to get data of relevance to the intended meaning of this provision. Perhaps it would reveal only examples of personal *bearing* of firearms and never of *transporting* in a vehicle. If so, that might tell us that the intended meaning is limited to the former.

Even then, however, the *might* qualifier is necessary. The limitation here is whether a preponderance of examples of uses of one sense of *carry* may indicate only that this is the first sense to come to mind, and whether a broader sense that might occur to a lawmaker on reflection should count as ordinary. Moreover, data from a general, balanced corpus could tell us something about the way the human mind conceptualizes the notion of carrying a firearm. But that might not be the right question to ask. We might be missing an important element of pragmatic context if we ask only about *carrying a firearm* in the abstract. Another relevant element of such context may be the legal nature of the language of this law. The human mind may react differently to a criminal prohibition—a law imposing harsh consequences like a sentencing enhancement—than to a mere statement of description. Thus, we may form one understanding when listening to a

---

252. Not all corpora are collections of written texts. Recent work in corpus-based pragmatics includes "'multi-modal' corpora" with audio and visual components that allow researchers to study "feedback in the form of gesture, body posture and gaze as well as their integration with discourse." Christoph Rühlemann & Karin Aijmer, *Introduction: Corpus Pragmatics: Laying the Foundations*, *in* CORPUS PRAGMATICS: A HANDBOOK 1, 4-5 (Karin Aijmer & Christoph Rühlemann eds., 2015).

descriptive narrative of a person *carrying a firearm* in connection with a drug crime, and another when warned that the punishment for a drug crime could be significantly enhanced if we *carry a firearm* in that circumstance. That sort of context may be impossible to suss out with corpus analysis alone. We may have to turn to other empirical approaches to language meaning and perception.

How might a judge answer this question? Some such questions may be framed within the standard picture. Where the question is presented as one of the likely intended meanings of rules like the "Keep off the grass" sign or the prohibition on "vehicles," we think judges are in a good position to assess likely intended meaning (even absent hard data about actual usage). We say that because we think the relevant pragmatic context of these rules is likely to be apparent in the cited circumstances. It seems difficult to think of a legislative "compromise" that would call into question the inference of uniform legislative intent to allow groundskeepers on the grass or ambulances in the park.[253] If so, it seems safe to conclude that the intended communicative content of these rules would sustain exceptions for groundskeepers and ambulances.

That will not always be so, however. *Muscarello* may be a good example. If we lack confidence in the corpus data on *carry*, we may be left to make an inference about likely legislative intent. Here that seems hard. As the majority and dissenting opinions in that case demonstrate, it is easy to contemplate legislative intent running in either of two directions—to call for a sentencing enhancement (a) whenever a gun is available to the defendant in a drug deal, since a gun may

---

253. *See* MANNING & STEPHENSON, *supra* note 11, at 54 (stating that "laws will be messy, uneven, and ill-fitting with their apparent purposes not because Congress is short-sighted or imprecise, but rather because legislation entails compromise, and compromise is untidy by nature"); SCALIA & GARNER, *supra* note 28, at 39 ("Not only is legal drafting sometimes imperfect, but often the imperfection is the consequence of a compromise that it is not the function of the courts to upset—or to make impossible for the future by disregarding the words adopted."); Frank H. Easterbrook, *Text, History, and Structure in Statutory Interpretation*, 17 HARV. J.L. & PUB. POL'Y 61, 68 (1994) (noting that if a particular outcome under a statute seems "unprincipled," it may be the "way of compromise" in the legislative process and that "[l]aw is a vector rather than an arrow," "[e]specially when you see the hand of interest groups").

always be used in a harmful way if it is available,[254] or (b) only if the gun is being carried on the defendant's person, since that kind of availability is even riskier.[255]

Muscarello is thus an example of a case in which pragmatic judgments about legislative intention are likely to be difficult. If we cannot decide the case on the basis of usage, informed by semantic context, and if we lack reliable evidence from similar pragmatic contexts, then we may be left to resolve it on other grounds. Here we could simply turn to the law of interpretation — giving the law legal content that does not pretend to be based on communicative content (because we have not been able to find it).[256] We could do so, for example, on the basis of a substantive canon like the rule of lenity. Reliance on that canon may make sense doctrinally, as lenity appears appropriate given genuine ambiguity about statutory meaning. Such a move, moreover, would be more open and transparent than a false assertion about communicative content. For that reason, we would favor it, even though it might not obviously vindicate the principles motivating the law's baseline devotion to ordinary meaning.

---

254. Muscarello v. United States, 524 U.S. 125, 133 (1998) ("How persuasive is a punishment that is without effect until a drug dealer who has brought his gun to a sale (indeed has it available for use) actually takes it from the trunk (or unlocks the glove compartment) of his car? It is difficult to say that, considered as a class, those who prepare, say, to sell drugs by placing guns in their cars are less dangerous, or less deserving of punishment, than those who carry handguns on their person.").

255. *Id.* at 145 (Ginsburg, J., dissenting) ("It is reasonable to comprehend Congress as having provided mandatory minimums for the most life-jeopardizing gun-connection cases (guns in or at the defendant's hand when committing an offense), leaving other, less imminently threatening, situations for the more flexible Guidelines regime.").

256. An alternative formulation would follow under the interpretive premises of the original methods originalists. *See* John O. McGinnis & Michael B. Rappaport, *The Constitution and the Language of the Law* (San Diego Legal Studies Paper No. 17-262, 2017), http://papers.ssrn.com /abstract_id=2928936 [http://perma.cc/TD97-EJ6Q] (articulating premises of original methods originalism, including the idea that the Constitution is written in the "language of the law," not ordinary English, and thus that it should be interpreted in accordance with the canons and legal conventions that would have been accepted by the legal community at the time of the founding). To the extent the law is written in a specialized legal dialect, we can think of the "law of interpretation" as a mere component of the "communicative content" of the law. We can do so by treating canons of construction not as departing from communicative content but as informing our understanding of the peculiar dialect of the law.

   We have no problem with the framing proposed by McGinnis and Rappaport. But we mostly speak here of a distinction between ordinary communicative content and the law of interpretation — because we think the distinction helps highlight a component of the inquiry that corpus linguistics can help us improve.

c. *Meaning as of When?*

Judges sometimes make reference to the temporal aspects of interpretation and insist that they are seeking the meaning of the text at the time it was drafted. Yet in practice judges often ignore the temporal aspect of interpretation or attempt to address it using tools of questionable utility, like historical dictionaries. Our linguistic intuitions about usage and meaning in our own time and our own speech community can be highly unreliable. But this problem is amplified when we are interpreting a text that dates from a period of which we have no linguistic memory or experience. To the extent that the law wishes to take into account the meaning of a text at the time of its enactment, some empirical measure of historical usage is necessary and corpus linguistics presents itself as an attractive option.

Of course, historical data from linguistic corpora face the same challenges that contemporary data face. If it is not clear whether carrying a firearm on one's person or carrying a firearm in an automobile would be perceived as two distinct senses in contemporary usage, it may not be clear from historical data either. Moreover, while we may be able to address the sense division problem using other linguistic empirical methods (discussed below), these methods generally involve attempting to measure the perceptions of living, human study participants. In many cases of historical interpretation, test subjects from the relevant speech community will not be available. In a historical context, corpus data may not just be a type of linguistic evidence; it may be the only type available.

d. *Whose Meaning?*

Corpus linguistics allows us to take account of variations in usage among different speech communities and linguistic registers. Because the interpretive problems addressed above have to do with the interpretation of federal statutes — written texts meant to be applied broadly to the population of the United States — we have relied on linguistic corpora that present evidence of usage from standard written American English.

Linguistic corpora are not limited to broad-based, standardized dialects or speech communities. We can well imagine interpretive problems that require appeal to language data from more narrowly drawn speech communities (including language use from different geographical regions) and registers (including language use from different professions or industries). Corpora can be created and corpus data made available to address questions of interpretation from these narrowly drawn speech communities and registers.

But sometimes the public will interpret statutory language in pragmatic context differently than a legislative body would interpret it. And that implicates the "whose meaning" and "speech community" questions.

The speech community question, as we have noted, has implications for the selection of a relevant corpus. If we are trying to measure intended meaning, we might want to gather data from a corpus of a community of speakers who look demographically like Congress. Yet if we are interested in public meaning, we would want to turn to a broader corpus.

What if our sense of public meaning differs from our sense of intended meaning? If that happens we would need to decide which data set to rely on. That is a problem for legal theory — and essentially a choice of which of two sets of justifications for the "standard picture" we seek to vindicate. In the *Muscarello* setting, the answer may well be the fair notice rationale. The law of interpretation may already have given that answer in the rule of lenity. In criminal cases the rule of lenity suggests that the notice rationale predominates. It indicates that a criminal defendant is entitled to the benefit of the doubt in cases of ambiguity as to the law's communicative content.

The question may be harder to answer in civil cases. But again, that is a problem for legal theory. As above, we will simply say for now that transparent answers are better than opaque ones. Further thinking on this problem is needed. Yet surely we will be better off with an open, transparent discussion about whether (and when) to give primacy to intended meaning and when to credit public meaning. Once we speak more carefully about the meaning we are looking for and proceed more reliably in trying to measure it, we can have a better dialogue about these difficult questions of legal theory.

2. Conclusions

Here we offer some data-backed conclusions about the ordinary sense of *vehicles* in the park, *carrying* a firearm, *interpreter*, and *harboring* an alien. In so doing, we highlight strengths of the corpus analysis while also acknowledging some drawbacks and unresolved questions.

In each of the test cases, we start with a premise of ordinary meaning that is susceptible to both definition and measurement. The premise is that the ordinary sense of a term is that which occurs most frequently in a properly controlled linguistic context—namely, a context that controls for relevant syntactic and semantic considerations, that is aimed at the relevant speech community, and that is limited to the appropriate time frame. In other words, the sense of a word that is most frequent (after taking semantic factors into account) is prima facie also the sense most likely to avoid unfair notice (public meaning) and to vindicate the will of the legislature (intended meaning).

Yet we also recognize some possible grounds for questioning this prima facie showing. One possible response would be to question the viability of the relevant sense division—to suggest that the less frequent sense is just the dodo bird (an unusual example of a bird, but no less a bird). Another would be to identify pragmatic considerations that are not adequately assessed through a corpus search. In circumstances in which either of these concerns is present, we think the conclusion that the most frequent sense of a term is the ordinary one may be in doubt. In that event, we may turn to alternative means of empirical analysis (discussed below) or, ultimately, considerations that go to legal content rather than communicative content.

a. Vehicles

Based on the common collocates of *vehicle* and our analysis of its use in concordance lines, we can conclude that the most common sense of this term is in reference to automobiles. Airplanes and bicycles appear on our frequency continuum: they are attested in the data as possible examples of *vehicle*. But they are unusual—not the most frequent and not even common. If we accept the most common use of the word as the ordinary meaning, we can conclude that the ordinary meaning of *vehicle* is *automobile*.

We can also make a strong case for crediting the most common meaning as the ordinary one, in that it will best avoid unfair surprise (public meaning) and vindicate the presumed intent of the lawmaker (intended meaning). A decision to extend the law to bicycles or airplanes could upset reliance interests of those who—according to the data—are likely to think of automobiles when they read the law prohibiting vehicles. And the data give us no reason to think that those who enacted this prohibition were thinking of airplanes, bicycles, or toy cars. In our view, this weighs against treating these examples as falling under the ordinary sense of *vehicle*. But, as discussed above, that is a question for legal theory.

A similar question for legal theory concerns the ambulance question. Again, *ambulance* is attested as a *vehicle* in the corpus data. Ambulance also easily fits within the ordinary (automobile) sense of *vehicle*. So the question here is one of intended meaning or pragmatic public meaning—another question for legal theory.

What about golf carts? We found no examples of golf carts as vehicles in the corpus. But does that mean they do not qualify under the ordinary meaning of *vehicle*? Like the ambulance, a golf cart shares a number of features with the most common vehicles: automobiles. On the other hand, we would not expect to see a lot of golf carts on the Autobahn. The question whether a golf cart fits into the ordinary meaning of *vehicle* (an ordinary meaning that the corpus data tells us is the *automotive* use of *vehicle*) is accordingly a difficult one. It turns on the viability

859

of the sense divisions at work—on whether the golf cart is an unusual example or perceived as a distinct linguistic construct. That is not an easy question to answer. It depends, as noted in Part II above, on the sufficient conditions for the *automobile* sense of *vehicle*.

There is more than one way to answer questions like this one. One way would be through further corpus analysis. With sufficient corpus data, we could assemble a list of criteria for things we speak of as an automobile, and then ask whether a golf cart has those criteria.[257]

In addition to corpus analysis, there are other empirical linguistic techniques that could be employed. One alternative may be empirical methods employed in the field of psycholinguistics. Psycholinguists use a variety of experimental techniques in order to measure how we perceive and interpret language, including cross-modal priming,[258] visual world paradigm analysis,[259] and eye tracking during reading.[260] Yet it is costly to design and implement psycholinguistic experiments of this sort; both specialized equipment and a high degree of expertise are required. While psycholinguistic approaches to ordinary meaning are promising, the current hurdles are significant.

---

257. Possible criteria, for example, would likely include a steering wheel, motor, wheels for passage on land, and seats for passengers. If those are the criteria, then a golf cart might count. But we can also imagine other criteria, like usual usage on paved roads or highways, or licensure by the state motor vehicle division. And if those are the criteria, then a golf cart might not count.

258. Simon Garrod, *Psycholinguistic Research Methods*, *in* 10 ENCYCLOPEDIA OF LANGUAGE & LINGUISTICS 251, 252 (Keith Brown et al. eds., 2d ed. 2006) ("Cross-modal priming can indicate the immediate interpretation of an ambiguous word, such as *bug*, in contexts that promote either one or other meaning of the word (e.g., 'insect' or 'listening device'). As participants listen to *bug* in the different contexts, they are presented with a written word (ANT or SPY) or a nonword (AST) and have to decide as quickly as possible whether the target is a word or not (this is called lexical decision).").

259. *Id.* at 253 (explaining that the visual world paradigm uses eye-tracking technology to measure "the focus of attention correspond[ing] to the words being looked at [at] any time or it can be used to measure which part of a scene a participant attends to as they interpret spoken utterances about that scene").

260. *Id.* at 254 ("Eyetracking has been used to study a wide range of linguistic processes, including lexical access, resolving lexical ambiguities, syntactic analysis, and various discourse processing phenomena, such as anaphora resolution. It is particularly effective in determining precisely when the reader makes a decision about some aspect of the linguistic input during sentence or discourse processing.").

Cognitive linguists[261] and sociolinguists[262] assemble language data through surveys or interviews with test subjects. Analysts in these fields may view the mind's conception of words as "represented in cognition not as a set of criterial features with clear-cut boundaries but rather in terms of prototype (the clearest cases, best examples) of the category."[263] In an important study, participants ranked words as "good examples" of particular categories, including *toys, fruits, birds*, *weapon*, and *vehicle*,[264] and demonstrated "high agreement" on these rankings.[265] *Chair* was found to be a more prototypical example of *furniture* than *stool*,[266] *automobile* was found to be a more prototypical *vehicle* than *yacht*,[267] and *robin* was found to be a more prototypical *bird* than *ostrich*.[268]

Similar survey methodologies have been employed to address questions of ordinary meaning, both in statutory interpretation[269] and the interpretation of contracts.[270] Yet there are significant barriers to using survey data to address questions of ordinary meaning. If we want to find meaning as of the date of a statute's enactment, we will never be able to measure it through survey data. And survey data is notoriously susceptible to context effects and response bias.[271]

---

261. Dirk Geeraerts & Hubert Cuyckens, *Introducing Cognitive Linguistics*, *in* THE OXFORD HAND-BOOK OF COGNITIVE LINGUISTICS 3, 3 (Dirk Geeraerts & Hubert Cuyckens eds., 2010) (explaining that cognitive linguistics is a linguistic discipline that "focuses on language as an instrument for organizing, processing, and conveying information" and as "the analysis of the conceptual and experiential basis of linguistic categories").

262. BERNARD SPOLSKY, SOCIOLINGUISTICS 3 (1998) ("Sociolinguistics is the field that studies the relation between language and society, between the uses of language and the social structures in which the users of language live.").

263. Rosch, *supra* note 46, at 193.

264. *Id.* at 197-98.

265. *Id.* at 198.

266. *Id.* at 229.

267. *Id.* at 230.

268. *Id.* at 232.

269. Clark D. Cunningham et al., *Plain Meaning and Hard Cases*, 103 YALE L.J. 1561, 1599-1601 (1994); J.P. Sevilla, Measuring Ordinary Meaning Using Surveys (Sept. 28, 2014) (unpublished manuscript), http://papers.ssrn.com/abstract_id=2466667 [http://perma.cc/BK82-GM2C].

270. Omri Ben-Shahar & Lior Strahilevitz, *Interpreting Contracts via Surveys and Experiments* (Coase-Sandor Inst. for Law & Econ., Working Paper No. 791, 2017), http://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?article=2464&context=law_and_economics [http://perma.cc/Z8JY-CTG8].

271. *See* 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, 202 (2011) (discussing the limitations of survey data when applied to questions of ordinary meaning).

Nevertheless, these alternative empirical linguistic methods provide possible approaches to addressing questions of ordinary meaning beyond the use of corpus linguistics.

The limitations of the empirical methods discussed here may be prohibitive. They may lead us to conclude that we cannot give a conclusive answer to the question of whether the ordinary understanding of *vehicle* extends to the *golf cart* — or to related questions about go-karts or four-wheelers. At that point it may be time to abandon the standard picture — to fall back on "fake" answers giving legal content to the law that is not necessarily in line with its communicative content. That seems fine, but as a fallback. As our sense of the law's communicative content becomes less clear, the reasons for crediting it are much weaker. Our point is just that this should not be the law's first instinct.

b. *Carrying a Firearm*

The corpus data tend to support the dissenting position in *Muscarello*. In both the NOW Corpus and the COHA, the vast majority of concordance lines involved the *bearing on your person* sense of *carry*. That gives us some meaningful empirical data about language usage. It tells us that when people speak of carrying a firearm they are almost always talking about carrying it on their person. That provides a prima facie basis for concluding that the ordinary communicative content of the mandatory minimum sentencing provision in § 924(c)(1) is limited to the personally *bearing* notion of *carry*.

Solan and Tammy Gales might observe that the data may merely be an artifact of the greater commonality of the personally *bearing* notion of *carrying* in the real world.[272] That is probably correct, but not necessarily a reason to distrust the data. If most every time we speak of carrying a firearm we are talking about personally *bearing* it, then the first sense of *carrying* to come to mind is likely to be that sense. Extending the statute to the *transporting* in a car sense may therefore jeopardize significant reliance interests.

That leaves, as above, the question of whether *bearing* and *transporting* are two distinct linguistic constructs or just alternative examples within the same construct. Again, we could test this by further empirical analysis — by finding (through corpus or other empirical linguistic study) the sufficient conditions of *carrying*, and asking whether *bearing* and *transporting* both qualify.

Perhaps we will not ultimately find a satisfactory answer to this question in any empirical data. But even then the data will have been helpful. They will allow us to avoid the smokescreen grounds for assessments of ordinariness articulated

---

272. *See* Solan & Gales, *supra* note 44.

by the competing opinions in *Muscarello*, and provide a sufficient basis for turning to other means of assessment.

One such means could be an attempt to assess intended meaning. This inquiry may be a difficult one, as noted above. But again, at least a decision on this basis will be a transparent one—rooted in a disagreement about whether Congress was likely concerned only about firearms on a drug dealer's person, or might also have been concerned about guns within relative reach in the dealer's vehicle. That sort of debate may seem an empty one to a judge seeking determinacy in the ordinary meaning of the text; but where such meaning is indeterminate, this debate seems preferable to a completely fabricated answer—like one rooted in a dictionary or etymology.

c. Interpreter

The data seem to provide support for Justice Alito's majority view in *Taniguchi*. We did not find a single instance of *interpreter* in the context of text-to-text written translation in the concordance lines we reviewed in the NOW Corpus. That strongly indicates that this is not the kind of interpreter that first comes to mind when we use this term.

That leaves the same question highlighted in the other examples: whether the *written translator* sense would be perceived as separate from the *oral translator* notion.[273] Here we see reason to suspect that these are just alternative examples of a single linguistic construct. There is at least some indication of that in the fact that some lexicographers treat these as just alternative examples of a single sense.[274] And, again, that is likely a question that could be tested empirically.

We have not sought to study intended meaning in our corpus analysis. But as noted above we think such a study is possible. One approach would be to think of *interpreter* as a term used by lawmakers, and to look for evidence of usage in this speech community. If we assembled such evidence, then we could have

---

273. *See* Taniguchi v. Kan Pac. Saipan, Ltd., 566 U.S. 560, 580 (2012) (Ginsburg, J., dissenting) (asserting that "[d]istinguishing written from oral translation" is a "dubious" endeavor, noting that "some translation tasks do not fall neatly into one category or another," and asserting that an oral interpreter "may be called upon to 'sight translate' a written document").

274. *See id.* ("[A]s the Court acknowledges, *ante*, at [567-568], and n. 2, 'interpreters' is more than occasionally used to encompass those who translate written speech as well. See Webster's Third New International Dictionary of the English Language 1182 (1976) (hereinafter Webster's) (defining 'interpreter' as 'one that translates; *esp*: a person who translates orally for parties conversing in different tongues'); Black's Law Dictionary 895 (9th ed. 2009) (defining 'interpreter' as a 'person who translates, esp. orally, from one language to another'); Ballentine's Law Dictionary 655 (3d ed. 1969) (defining 'interpreter' as '[o]ne who interprets, particularly one who interprets words written or spoken in a foreign language')." (alteration in original)).

the debate flagged above — as to whether intended meaning should win out over public meaning, or whether they ought to collapse together as a matter of theory.

d. Harbor

Our *harbor* data seem inconclusive. We found a significant number of instances of both the *conceal* sense and the *shelter* sense of this term. That suggests that both senses are common and attested. To the extent we regard the ordinary meaning as a common or attested sense, then the data indicate that both are "ordinary." To the extent we regard the ordinary meaning as the most common sense of a word, however, the data appear to indicate that neither sense is "ordinary."

It is hard to know what conclusion to draw from these inferences (even accepting that we have a statistical basis for doing so). One possibility is to say that both senses are ordinary in that they are both commonly attested. This is presumably the dissenting view in *Costello*, and in line with the approach at least sometimes taken on the "carry" question in *Muscarello* (that both personal carrying and car carrying count as ordinary).

Another alternative is simply to abandon our search for the "standard picture." If we lack probative data on the most frequent sense of a given term, we may conclude that we cannot determine the ordinary communicative content of the law — and thus that we need a "fake" answer, like that provided by the rule of lenity. Yet even here the data will have proven useful. The application of the rule of lenity requires an antecedent finding that the criminal statute is ambiguous — that the words of the statute are susceptible of two competing interpretations.[275] Often, such determinations are highly impressionistic.[276] Here, the antecedent finding of ambiguity necessary to the application of the rule of lenity is based not on intuitions or dictionaries, but on quantifiable data about real-world usage — data that establishes that both senses of *harbor* are attested and comparatively common. The standard picture here yields to the law of interpretation, but only after the necessary work has been done.

---

275. Barber v. Thomas, 560 U.S. 474, 500-01 (2010) (Kennedy, J., dissenting) ("When a penal statute is susceptible of two interpretations, the one more favorable to the defendant must be chosen unless 'text, structure, and history . . . establish that the [harsher] position is unambiguously correct.'" (alteration in original)).

276. Ward Farnsworth, Dustin F. Guzior & Anup Malani, *Ambiguity About Ambiguity: An Empirical Inquiry into Legal Interpretation*, 2 J. LEGAL ANALYSIS 257, 276 (2010) ("[T]here is no avoiding the fact that impressionistic judgments are doing important work. Some judges read the text and say that it just seems clear. Other judges read the same text and say that it just doesn't.").

IV. OBJECTIONS AND RESPONSES

We have little doubt of the need and basis for corpus linguistic analysis of ordinary meaning. But we anticipate — and already have seen — significant objections to the use of these new tools of interpretation. In a few recent cases, judges have proffered corpus linguistic analysis in support of their assessment of the ordinary meaning of statutory terms.[277] Some of these attempts have prompted doubt and criticism from fellow judges.[278] And even the judges who have advocated for this approach (present company included) have acknowledged cause for concern and care in this endeavor.[279]

The criticisms that we have considered fall into three categories: proficiency, propriety, and practicality. Each concern has an element of viability but crumbles under careful scrutiny.

After addressing these concerns, we close with some observations about the potential role for corpus linguistic analysis going forward — about the extent to which corpus data can address the deficiencies in the ordinary meaning analysis highlighted herein. We also highlight the ideological or theoretical neutrality of this methodology, explaining why corpus analysis is not just for textualists (or originalists), but for anyone who takes language seriously.

A. Proficiency: Judges (and Lawyers) Can't Do Corpus Linguistics

Judges and lawyers are not linguists. Most all of us, at least, are not professionally trained ones. From that premise it is easy to jump to the conclusion that judges and lawyers should leave the linguistic analysis to professional linguists — meaning, in practice, to expert witness reports or testimony. A majority of the Utah Supreme Court has so concluded in recent cases.[280]

---

277. *See, e.g.*, People v. Harris, 885 N.W.2d 832 (Mich. 2016); State v. Rasabout, 2015 UT 72, 356 P.3d 1258 (Lee, Associate C.J., concurring in part and concurring in the judgment); J.M.W. v. T.I.Z. (*In re* Adoption of Baby E.Z.), 2011 UT 38, 266 P.3d 702 (Lee, J., concurring in part and concurring in the judgment).

278. *See, e.g.*, *Rasabout*, 2015 UT 72, 356 P.3d 1258; *In re Adoption of Baby E.Z.*, 2011 UT 38, 266 P.3d 702.

279. *See, e.g.*, *Rasabout*, 2015 UT ¶ 97, 356 P.3d at 1282-83 (Lee, Associate C.J., concurring in part and concurring in the judgment).

280. *Id.* ¶ 18, 356 P.3d at 1265 (majority opinion) ("The knowledge and expertise required to conduct scientific research are 'usually not within the common knowledge' of judges, so 'testimony from relevant experts is generally required in order to ensure that [judges] have adequate knowledge upon which to base their decisions.'" (quoting Bowman v. Kalm, 2008 UT 9 ¶ 7, 179 P.3d 754, 755-56)); *In re Adoption of Baby E.Z.*, 2011 UT ¶ 19 n.2, 266 P.3d at 708 n.2

The "proficiency" critique has some bite. For reasons noted above, we concede that corpus linguistics is not "plug and play" analysis. Corpus data can be gathered and analyzed properly only with care and a little background and training in the underlying methodology. A judge who proceeds willy-nilly may, either consciously or unwittingly, proffer data that has only the appearance of careful empiricism.[281] For these and other reasons we wholeheartedly agree that the judicial analysis of ordinary meaning will be improved in cases in which the parties or their experts proffer corpus analysis that can be tested by the adversary system.[282]

So we take the "proficiency" critique as an appropriate word of warning. Judges should acknowledge the pitfalls and limitations of the tool of corpus linguistics. They should not overstate its utility, ignore the care required to use it properly, or overlook the potential for subjectivity or even strategic manipulation.

But that is as far as this critique can take us. The fact of the matter is that judges and lawyers *are* linguists. We may not be trained in linguistic methodology, but our work puts us consistently and inevitably in the position of resolving ambiguities in legal language. Judges and lawyers are experts, in other words, in interpreting the law.[283] So the question, ultimately, is not whether we trust judges to engage in linguistic analysis. It is whether we want them to "do so with

---

("Unless this linguistic 'context' is placed in its proper context, it is of little analytical or persuasive value.").

281. *See Rasabout*, 2015 UT ¶ 21, 356 P.3d at 1266 (noting that a potentially significant portion of corpus data "require[s] an interpretive assumption" or retains some level of ambiguity); Michael Stubbs, *Corpus Semantics*, *in* THE ROUTLEDGE HANDBOOK OF SEMANTICS 106, 107 (Nick Riemer ed., 2016) (noting that "a constant background question is whether a corpus can ever, strictly speaking, provide semantic data, since intuition is always required to interpret the data," but concluding that "corpora allow us to study language 'with a degree of objectivity [. . .] where before we could only speculate'" (quoting Adam Kilgarriff, *Putting Frequencies in the Dictionary*, 10 INT'L J. LEXICOGRAPHY 135, 137 (1997))); Ben Zimmer, *The Corpus in the Court: 'Like Lexis on Steroids,'* ATLANTIC (Mar. 4, 2011), http://www.theatlantic.com/national /archive/2011/03/the-corpus-in-the-court-like-lexis-on-steroids/72054 [http://perma.cc /KBQ5-BSQ4] ("While the corpus revolution promises to put judicial inquiries into language patterns on a firmer, more systematic footing, the results are still prey to all manner of human interpretation.").

282. *Rasabout*, 2015 UT ¶ 97, 356 P.3d at 1287 (Lee, Associate C.J., concurring in part and concurring in the judgment) (agreeing that judicial analysis of any kind is "better when adversary briefing is complete and in-depth").

283. *Id.* ¶ 107, 356 P.3d at 1285 (explaining that although they do not have "the kind of training possessed by 'linguistics experts' . . . judges *are* experts on one thing—interpreting the law").

the aid of — instead of in open ignorance of or rebellion to — modern tools developed to facilitate that analysis."[284]

Judges are likewise not historians. And it may rightly be said that many lawyers and judges are even "bad historians" that tend to "make up an imaginary history and use curiously unhistorical methods."[285] As one of us recently noted, "[J]udges of all stripes engage in historical analysis, particularly in their interpretation of the [C]onstitution."[286] "So the response to our lack of historical training is not to back away from the enterprise; it is to arm ourselves with the tools necessary to do the best history we can."[287]

The same goes for linguistic analysis. "We could continue to judge the ordinary meaning of words based on intuition, aided by the dictionary. But those tools are problematic."[288] Thus "it is our *current* methodology and tools that involve bad linguistics produced by unscientific methods."[289] Therefore, "[i]f the concern is reliability, the proper response is to embrace — and not abandon — corpus-based analysis."[290]

The potential for subjectivity and arbitrariness is not heightened but reduced by the use of corpus linguistics.[291] Without this tool, judges will tap into their linguistic memory to make assessments about the frequency or prototypicality of a given sense of a statutory term. Such recourse to memory and judicial intuition is neither transparent nor replicable. Nothing is statistically worse than one data point — especially a biased one. The potential for motivated reasoning is evident.[292]

---

284. *Id.*

285. MAX RADIN, LAW AS LOGIC AND EXPERIENCE 138 (1940).

286. *Rasabout*, 2015 UT ¶ 109, 356 P.3d at 1285 (Lee, Associate C.J., concurring in part and concurring in the judgment).

287. *Id.*

288. *Id.* ¶ 112, 356 P.3d at 1285.

289. *Id.*

290. *Id.*

291. As one of us has noted, while a judge may "go looking for supporting evidence in a corpus," it is possible that "after reviewing a few hundred concordance lines, a salient meaning contrary to the judge's initial conclusion becomes harder to ignore." Mouritsen, *supra* note 271, at 202.

292. A common critique of the use of legislative history comes to mind: "It sometimes seems that citing legislative history is still, as [Judge] Harold Leventhal once observed, akin to 'looking over a crowd and picking out your friends.'" Patricia M. Wald, *Some Observations on the Use of Legislative History in the 1981 Supreme Court Term*, 68 IOWA L. REV. 195, 214 (1983). A parallel problem appears in cases in which judges summon examples of word usage in literary works. *See, e.g.*, Whitfield v. United States, 135 S. Ct. 785, 788 (2015) (interpreting the ordinary meaning of "to accompany" using a host of sources, including quotes from a Jane Austen as well as

Corpus linguistics, by contrast, facilitates transparency and scrutiny.[293] It is "an empirical check on our (imperfect) linguistic intuition."[294] And it is not, ultimately, a terribly complex or difficult endeavor. "Corpus analysis is like math"—everyone can do it at some basic level; at more advanced levels it becomes too complicated for all but the experts.[295] We're advocating rudimentary linguistic analysis that most anyone can do.[296] We "just think we should be using a calculator instead of doing it in our heads."[297]

The path forward is for judges and lawyers to identify the corpus analysis that we can perform sufficiently and reliably to supplement the tools we are now using (and the sort of analysis we must leave to linguists). In time, the law and corpus linguistics movement will develop standards and best practices for this field. Until then we should proceed cautiously and carefully in a direction that will allow us to be the best linguists we can. Paraphrasing an observation made by Justice Scalia and his co-author Garner regarding judges performing historical analysis, we may or may not be able to do corpus linguistics with the precision of an expert, but "[o]ur charge is to try."[298]

B. Propriety: Judges Shouldn't Do Corpus Linguistics

The law puts limits on judicial analysis of matters that exceed the bounds of the briefing and record in a particular case. Our rules of judicial ethics say that "[a] judge shall not investigate facts in a matter independently," but shall "consider only the evidence presented and any facts that may properly be judicially

---

a Charles Dickens novel). That kind of data cherry-picking is fraught with risk of hindsight bias or motivated reasoning.

293. *See* Mouritsen, *supra* note 271, at 203 ("[C]orpus analysis brings these subconscious assumptions about language and meaning out in the open."); Zimmer, *supra* note 281 ("[A]t least these ideological arguments can proceed on a basis of concrete facts about how we use language, rather than on a welter of idiosyncratic assumptions, as has too often been the case.").

294. *Rasabout*, 2015 UT ¶ 66, 356 P.3d at 1277 (Lee, Associate C.J., concurring in part and concurring in the judgment).

295. *Id.* ¶ 115, 356 P.3d at 1286.

296. While the COHA and similar linguist-designed corpora are more foreign than Google or Westlaw, they are being employed in the same way. In short, we are advocating using a corpus as a search tool or database to find uses of language that are as easy to read as a Google search result. The results are just more reliable.

297. *Rasabout*, 2015 UT ¶ 115, 356 P.3d at 1286 (Lee, Associate C.J., concurring in part and concurring in the judgment).

298. SCALIA & GARNER, *supra* note 28, at 400.

noticed."[299] With this in mind, a majority of the Utah Supreme Court has challenged the "sua sponte" use of corpus linguistics as falling beyond the proper domain of the judge.[300]

The analogy here may arguably be to cases in which judges perform their own experiments to assess the factual assertions of the parties in a particular case. A prominent example appears in Judge Posner's opinion in *Mitchell v. JCG Industries*.[301] A question in that case was how long it took poultry processing workers to change in and out of the safety clothing they used to do their jobs. Judge Posner's opinion included a reference to an experiment he conducted on that question in chambers—in which he ordered the clothing in question and asked "three members of the court's staff" to change in and out of it "as they would do if they were workers at the plant."[302] "Their endeavors were videotaped,"[303] and "[t]he videotape automatically recorded the time consumed in donning and doffing and also enabled verification that the 'workers' were neither rushing nor dawdling."[304] Posner referred to the results of this experiment in support of "the common sense intuition that donning and doffing a few simple pieces of clothing and equipment do not eat up half the lunch break."[305]

Chief Judge Wood, in dissent, asserted that the Posner majority went "beyond the proper appellate role" in conducting an experiment of relevance to a factual question in the case.[306] She complained that the results of Judge Posner's experiment "cannot be considered as evidence in the case," which is limited to matters placed in the record pursuant to applicable rules of civil procedure.[307]

This may be the paradigm that critics of corpus linguistics have in mind when they question the viability of sua sponte use of this tool. The governing rules of judicial ethics prohibit judges from "investigat[ing] facts in a matter independently" and only allow the courts to consider "facts that may properly be judicially noticed."[308] But the analogy is inapt. Judges who consider corpus data

---

299. MODEL CODE OF JUDICIAL CONDUCT r. 2.9(C) (AM. BAR ASS'N 2011).

300. *Rasabout*, 2015 UT ¶ 17, 356 P.3d at 1264-65 (majority opinion).

301. 745 F.3d 837 (7th Cir. 2014).

302. *Id.* at 842.

303. *Id.*

304. *Id.*

305. *Id.*

306. *Id.* at 847 (Wood, C.J., dissenting).

307. *Id.* at 849.

308. UTAH CODE OF JUDICIAL CONDUCT r. 2.9(C) (2017); MODEL CODE OF JUDICIAL CONDUCT r. 2.9(C) (AM. BAR ASS'N 2011).

in assessing the ordinary meaning of a statute are not investigating the *adjudicative facts* of a case; they are considering facts of relevance to the proper interpretation of the law. These are known as *legislative facts*, and their investigation is the inevitable — and quite proper — domain of the judge's sua sponte analysis.

The point is supported by the law of evidence. Governing rules of evidence typically state that limitations on the judge's judicial notice power are addressed to "an adjudicative fact only, not a legislative fact."[309] The distinction is this: "[L]egislative facts are matters that go to the policy of a rule of law as distinct from the true facts that are used in the adjudication of a controversy."[310] Such facts "are not appropriate for a rule of evidence."[311] They are "best left to the law-making considerations by appellate and trial courts."[312] And that is precisely what is involved in the corpus linguistic analysis of the meaning of statutory text. Corpus analysis has nothing to do with adjudicative facts — with the who, what, when, or where of an underlying controversy. It has only to do with the proper construction of the applicable law. For that reason, there is no ethical or evidentiary prohibition on sua sponte corpus analysis by a judge.[313]

"A contrary conclusion would call into question a wide range of" inquiries routinely conducted by our courts, including the use of dictionaries:[314]

> If we were foreclosed from considering outside material that informs our resolution of open questions of law, we would be barred from engaging in historical analysis relevant to a question of original meaning of a provision of the [C]onstitution, or from considering social science literature

---

309. FED. R. EVID. 201(a) ("This rule governs judicial notice of an adjudicative fact only, not a legislative fact."); UTAH R. EVID. 201(a) (same).

310. UTAH R. EVID. 201(a) advisory committee note; *see also* Kenneth Culp Davis, *An Approach to Problems of Evidence in the Administrative Process*, 55 HARV. L. REV. 364, 402–03 (1942) ("When an agency wrestles with a question of law or policy, it is acting legislatively, just as judges have created the common law through judicial legislation, and the facts which inform its legislative judgment may conveniently be denominated legislative facts. The distinction is important; the traditional rules of evidence are designed for adjudicative facts, and unnecessary confusion results from attempting to apply the traditional rules to legislative facts.").

311. UTAH R. EVID. 201(a) advisory committee note.

312. *Id.*

313. For further commentary on the distinction between adjudicative and legislative facts, see, for example, *Bulova Watch Co. v. Hattori & Co.*, 508 F. Supp. 1322, 1328 (E.D.N.Y. 1981), which explains that the "court's power to resort to less well known and accepted sources of data to fill in the gaps of its knowledge for legislative and general evidential hypothesis purposes must be accepted because it is essential to the judicial process." *See also* ROBERT E. KEETON, JUDGING 38–39 (1990) (discussing the distinction between legislative and adjudicative facts).

314. State v. Rasabout, 2015 UT 72 ¶ 106, 356 P.3d 1258, 1285 (Lee, Associate C.J., concurring in part and concurring in the judgment).

in resolving a difficult question under the common law. Linguistic analysis is no different; to the extent we charge our judges with resolving ambiguities in language, we cannot (*and do not*) reasonably restrict their ability to do so on a well-informed basis — even on grounds not presented by the parties, and not within the domain of judges' professional training.[315]

For better or worse, judges do that all the time. State court judges decide questions of common law that require us to consider and weigh questions implicating literature in fields of social science about which we are hardly experts.[316] No one bats an eye when judges do their own research and thinking on a broad range of "legislative facts." The matter should be no different for linguistic analysis of ordinary meaning.

C. *Practicality: Corpus Linguistics Will Impose an Unbearable Burden*

The widespread use of corpus linguistics could put a strain on parties and the courts. This is another criticism that has appeared in majority opinions in the Utah Supreme Court. The argument is that turning the analysis of ordinary meaning into an empirical, data-driven enterprise will introduce the "dueling expert" problem and make statutory cases more costly and time-consuming.[317] The effects of the proliferation of expert testimony are a matter meriting careful consideration. For every question on which we require expert analysis, we compound the expense and time it takes for a case to be resolved. We should not do that without a good reason.

Yet we find this objection to corpus analysis unpersuasive for several reasons. First is the fact that not all problems of statutory interpretation lend themselves to corpus linguistic analysis. The utility of this tool, as currently conceived, is limited to problems of *lexical ambiguity* — of a contest between two meanings of the terms of the statutory text. That excludes a category of problems of *semantic* (or in other words *structural ambiguity*) — a problem, for example, as to whether a modifier is understood to apply to all items in a statutory list or only the "last

---

315. *Id.*

316. Recent examples in the Utah Supreme Court include the question whether a medical practitioner owes a duty to third parties who are foreseeably injured by the negligent prescription of pharmaceuticals, B.R. & C.R. *ex rel.* Jeffs v. West, 2010 UT 11, 275 P.3d 228; and the question of the appropriate age at which a child may be required to defend against a claim for negligence, Nielsen *ex rel.* C.N. v. Bell *ex rel.* B.B., 2016 UT 14, 370 P.3d 925.

317. *Rasabout*, 2015 UT ¶ 19, 356 P.3d at 1265 (majority opinion).

antecedent."[318] Even as to the class of cases of lexical ambiguity, moreover, not all cases will call for corpus analysis. In our view "[c]orpus analysis is something of a last resort."[319] "It comes into play only if we find that the legislature is not using words in some specialized sense, and only if we cannot reject one of the parties' definitions based on the structure or context of the statute."[320] This yields a limited but important domain for corpus linguistics. Judges should turn to an empirical analysis of frequency only in cases in which they have "no better way" of resolving a contest between probabilities of meaning.[321] That is a relatively rare case.[322]

Second, corpus-based analysis will not always require an expert. This "isn't rocket science."[323] Lawyers are crafty, ingenious creatures with the capacity to learn and even master new tools, technologies, and methodologies. Witness the way attorneys have learned to parse historical materials and present them when litigating the original meaning of the Constitution. In a way, lawyers have been doing corpus analysis for a long time; they scour Westlaw or Lexis to determine how courts have interpreted a phrase or concept. So it is undoubtedly true that lawyers will have to bone up on some basic linguistic methodology. But continuing education is an ongoing element of the legal profession. And a familiarity with and capacity for corpus analysis can take root just like Westlaw and Lexis searches did.[324] The rising generation of millennials is particularly suited to the

---

318. *See* Lockhart v. United States, 136 S. Ct. 958, 962 (2016) (describing the "rule of the last antecedent" and applying it to interpret a statute imposing a mandatory minimum sentence on defendants who violate the federal child pornography statute and have previously been convicted of certain crimes); CRUSE, *supra* note 125, at 107-08 ("Ambiguity has been presented here as a lexical phenomenon; it is important to emphasize, however, that there are other sources of ambiguity. One of these, of course, is syntax, as in *Mary saw the man with the telescope*. Many syntactic ambiguities arise from the possibility of alternative constituent structures, as here: *with the telescope* is either a manner adverbial modifying *saw*, or a prepositional phrase modifying *the man*.").

319. *Rasabout*, 2015 UT ¶ 118, 356 P.3d at 1286 (Lee, Associate C.J., concurring in part and concurring in the judgment).

320. *Id.*

321. *Id.*

322. *Id.* ¶ 118, 356 P.3d at 1287 (asserting that in "five years" on the Utah Supreme Court, Associate Chief Justice Lee had "employed such analysis only a very few times," and that "[i]n the many other statutory cases" that have arisen, he "disposed of the matter using more traditional tools of interpretation").

323. *Id.* ¶ 114, 356 P.3d at 1286.

324. The advent of computer-aided legal research is now an accepted staple. But it wasn't always thought to be so. Early reactions paralleled some of the responses to corpus linguistics. *See, e.g.*, Robert C. Berring, *Legal Research and the World of Thinkable Thoughts*, 2 J. APP. PRAC. & PROCESS 305, 316 (2000) (declaring that it "scares" the author "[i]f search engines like *Google*

task. They have never known life without a computer, and are constantly embracing new applications and tools for computer analysis. In time we may see competing corpus presentations as a matter of course in adversary briefing.[325] There will often be no need for dueling experts, just as there is often no need for dueling historical experts in constitutional litigation, or dueling dictionary experts on a statutory question.

Finally, if in the rare case there is a need for the parties to retain corpus linguistic experts, that is hardly cause for alarm. Where the issue is complex enough and the stakes are high enough, expert analysis could be helpful — and certainly preferable to deciding a matter as significant as, say, the applicability of a federal sentencing enhancement on the basis of an unreliable source like a dictionary or an opaque one like a judge's intuition. Some problems are important enough to merit expert analysis. We should leave that matter to the marketplace — to the clients and lawyers who decide how best to formulate and present a legal position.

D. Corpus Data Represents Only "More Factually Common Iterations"

Solan and Gales have observed that corpus data may reflect only the fact that a given sense of a certain term is a more factually common iteration of that term

---

move into legal information"); Barbara Bintliff, *From Creativity to Computerese: Thinking Like a Lawyer in the Computer Age*, 88 LAW LIBR. J. 338, 339 (1996) (warning that computer-aided legal research will undermine the ability to think like a lawyer); Molly Warner Lien, *Technocentrism and the Soul of the Common Law Lawyer*, 48 AM. U. L. REV. 85, 85-86 (1998) (arguing that computer-aided legal research "may be harmful to the process of legal reasoning" and that lawyers should be aware of the "negative impacts" of using technology in this way); Scott P. Stolley, *Shortcomings of Technology: The Corruption of Legal Research*, FOR THE DEFENSE 39 (Apr. 2004) (viewing the likes of LexisNexis and Westlaw as leading to a generation of lawyers who can't find cases on point). Most of us view this criticism as downright silly today. We realize that computer research tools can be misused, and may be improved if supplemented by more traditional methods. But they cannot properly be rejected on the basis of their unfamiliarity.

325. This will hold if, but only if, our courts continue to embrace this methodology — as has happened recently in Michigan. If we (judges) build it, they (attorneys) will surely come. *See* Appellants' (Third) Supplemental Authority, *In re* Estate of Cliffman, 892 N.W.2d 380 (Mich. 2017) (No. 67-151998), 2016 WL 4480882 (submitting supplemental authority with exhibits "show[ing] the relative frequency with which the words or word combinations appear in the COCA database," and noting that "[i]n *Harris*, this Court approved the use of the *corpus linguistics* in determining the common usage and meaning of statutorily undefined words"). But lawyers need not wait on the courts to begin incorporating corpus analysis in their briefing. Courts, including the United States Supreme Court, have already proven amenable to well-executed corpus-based analysis. *See* Zimmer, *supra* note 281 (discussing Neal Goldfarb's influential, corpus-based amicus brief in FCC v. AT&T, Inc., 562 U.S. 397 (2011)).

in the real world.[326] If that is true, there may be reason to doubt the probity of the data in establishing the semantic meaning[327] perceived by lawmakers or the public.

These are important concerns. And anyone turning to corpus analysis would do well to consider these limitations before jumping too quickly to an inference about ordinary meaning. But we do not view the sense-division problems noted here to be fatal to the probity of corpus linguistic analysis (even for related senses of a statutory term). We propose a range of responses to this concern.

First, the Solan-Gales point seems overstated. Let's apply it to the *carry* data. It may be likely, as Solan and Gales might suggest, that the corpus data we found is indicative of the fact that most iterations of carrying a firearm in the real world involve personally *bearing* it. Yet we do not see that as depriving the data of probative value. If most iterations of firearm carrying involve personally *bearing*, then that sense of carrying seems likely to be the one that first comes to mind when we think of this term. That top-of-mind sense, as noted, may not exhaust the breadth of human perception of this term. If pressed, some people might concede that the term encompasses the *transport* sense too. As discussed below, there may be a way to measure such perceptions of meaning.

This raises the question of whether to credit only the top-of-mind sense or a possibly broader, "reflective" sense as ordinary. But this is not a deficiency in corpus data — or even in linguistic theory. It is a question for law — "we have to decide *which* meaning, produced by which *theory* of meaning, we ought to pick."[328] We think the answers to these questions are dictated in part by the rationales that drive us to consider ordinary meaning. A concern for fair notice and protection of reliance interests may well direct us to stop at the top-of-mind sense of a statutory term. If the personally *bear* sense of *carry* is the first one that comes to mind, then that may be the sense that the public will have in mind upon reading the terms of a statute, and if we are interested in protecting reliance interests and avoiding unfair surprise, we may want to stop short of including the broader *transport* sense that the public might concede to be covered upon reflection.

We emphasize that corpus analysis does not take place in an acontextual vacuum. A corpus-based approach to ordinary meaning, as noted, does *not* simply evaluate which of two competing uses is the most common. Instead, the corpus

---

326. Lawrence M. Solan & Tammy Gales, *Corpus Linguistics as a Tool in Legal Interpretation*, 2017 BYU L. Rev. (forthcoming 2018).

327. The point is not to suggest that mere semantic meaning is the right framing. Above we conceded that the pragmatic context of relevance to so-called "intended" or "public" meaning is the correct focus. But for now we are speaking only of semantic meaning. We add the wrinkle of pragmatic context below.

328. Baude & Sachs, *supra* note 3, at 1089–90.

allows us to examine the use of a word or phrase in a particular syntactic, semantic, or pragmatic context, in the speech or writing of a particular speech community or register, and at a particular point in time. Our analysis of *carry*, for example, does not simply examine the use of *carry* at large. We look for sentences in which the verb *carry* has a human agent performing the carrying and a weapon object (*firearm* or one of its synonyms) being carried. We look for such instances in what we have argued is the relevant speech community and in texts dating from the era in which the relevant statute was enacted. With this level of granularity, we are often able to find not only common ways to describe common real-world occurrences, but also the most common ways in which highly particularized and highly contextualized occurrences are described in a given speech community at a given point in time. If there are cases where "it is natural to use a particular expression, but the circumstances do not arise often,"[329] as Solan and Gales suggest, an appropriately designed corpus search (performed in a sufficiently robust corpus) will help us identify these instances and make informed, evidence-based judgments about them.

Second, above we were considering data at the right end of the frequency continuum — an indication that one of two senses is clearly the most frequent, or even almost exclusive. But what if the data is less clear? What if the data suggests that each of two senses is about equally possible? Or that one is a bit more frequent but not clearly so?

Sometimes an indication that *both* senses of a term are relatively frequent will be telling. If two senses are closely related and both appear relatively equally in the data, that may tell us that both are about equally likely to be called to mind. In that event it may be difficult to exclude either as *extra*ordinary.

The salience of inconclusive data may also depend on the nature of the question presented. We have been speaking here of isolated questions of ambiguity and ordinary meaning. But sometimes the question of whether the language of a statute is plain or ordinary is bound up with questions of whether or not the court will consider extrinsic evidence of meaning like legislative history, or apply a substantive canon, or defer to an agency interpretation. In such cases, inconclusive data about which meaning is ordinary may be quite conclusive — it may tell us that there is ambiguity sufficient to proceed past the threshold "standard picture." Corpus data can bring rigor to this range of questions too; instead of guessing about plainness we can summon data.

In some cases, that data may be too mixed to yield any helpful answers. Even then that does not require us to abandon the standard picture. We could, for example, look to other empirical methods for measuring perceptions of meaning. Barring that kind of help, we can fall back on a principle of interpretation

---

329. Solan & Gales, *supra* note 326 (manuscript at 3).

framed by something other than a view of the standard picture — as in a rule of interpretation that has to do with "legal content" of the law, like the rule of lenity. But we see no reason to fall back too quickly. The law commits to the standard picture for good reason. We think the courts should try their best to find real answers to linguistic questions before falling back on fake ones.

E. Political Neutrality

A final potential concern goes to the utility of the methodology of corpus linguistics across a range of theories of interpretation. The utility of this methodology may be most apparent to the textualist or the originalist. But we see much broader applications for corpus linguistics. We also see reason for those who are skeptical of textualism and originalism to resort to this new tool with equal alacrity.

The textualist finds statutory meaning in the words of a legal text. For that reason, the textualist would have a natural affinity for a tool that promises to help uncover the meaning of the text. That affinity would be sharpened to the extent the tool can help deliver on the promise of determinacy — a promise at the heart of this theory.

The originalist's expected attraction to corpus analysis rests on similar grounds. An originalist who seeks the original public meaning of the words of the Constitution,[330] for example, would quickly see the value of data-based answers to questions previously left to more impressionistic analysis. The data would be viewed as delivering on a core promise of originalism — "fixation," which is the idea that written constitutions are viewed as carrying fixed content as of the time of their adoption.[331]

Yet it would be a mistake to dismiss corpus linguistic analysis as a methodology of likely interest only to the textualist and the originalist. Even the most jaded skeptics of these two approaches should find this methodology significant. A threshold reason was discussed above: even those who search for legislative intent or purpose view the text as the "best evidence" of intent or purpose.[332]

---

330. *See* John O. McGinnis & Michael B. Rappaport, *Original Methods Originalism: A New Theory of Interpretation and the Case Against Construction*, 103 Nw. U. L. Rev. 751, 761 (2009) ("[O]riginal public meaning, in contrast to original intent, interpret[s] the Constitution according to how the words of the document would have been understood by a competent and reasonable speaker of the language at the time of the document's enactment . . . . [and] is now the predominant originalist theory.").

331. *See* Lawrence B. Solum, *The Fixation Thesis: The Role of Historical Fact in Original Meaning*, 91 Notre Dame L. Rev. 1, 1 (2015).

332. *See id.* at 52 n.130, 65 n.154 and accompanying text.

"We're all textualists now" in that most all of us at least start with the text.[333] If we're going to start with the text, we should seize the best tools for discerning its meaning.

Yet even an avowed "anti-textualist" should be attracted to corpus linguistics, here for an instrumental or strategic reason. A key move for the anti-textualist is to challenge the purported determinacy of statutory text (or fixation of constitutional language). Corpus analysis can often help in that endeavor.[334] Where the data show that there is no ordinary meaning, or that there is a wide range of ordinary meanings, the interpreter will be free to dismiss the notion of determinacy (or fixation) and turn to other theories or tools of interpretation.

For these reasons, we see corpus linguistic analysis as a tool without any necessary connection to a theory of interpretation or an often-corresponding political ideology. It is a neutral tool with broad utility for anyone interested in data of relevance to the analysis of ordinary meaning.

F. Potential: The Role for Corpus Linguistic Analysis in Addressing Problems of Ordinary Meaning

Another critique is one not yet made by critics but implicitly acknowledged in our analysis throughout this Article. For all our bemoaning of the deficiencies in the law's construct of ordinary meaning, and touting of the insights provided by corpus linguistic analysis, we have not really offered our own grand theory of corpus-based ordinary meaning. Because we acknowledge that corpus data may at least sometimes be indeterminate, we cannot claim that corpus linguistic analysis will definitively resolve ongoing debates about the ordinary meaning of the language of the law.

In that sense, we are left to concede that the methodology we propose is not an answer to the many facets of the ordinary meaning problem highlighted herein. That said, we are not shy in asserting that corpus linguistic analysis is an essential step in improving the quality of the ordinary meaning inquiry going forward. At a minimum, the data that can be compiled through corpus linguistic analysis will allow lawyers and judges to have a transparent debate informed by real data instead of inferences from sources (like dictionaries or etymology or intuition) that are both opaque and ill-suited to the task to which they are applied. The corpus methodology that we have introduced promises three contri-

---

333. Kagan, *supra* note 10.

334. Mouritsen, *supra* note 271, at 161 (arguing that textualist analysis may be "vulnerable" to attack from corpus data because textualist claims about meaning "can be proven true or false using empirical linguistic methods").

butions to ordinary meaning analysis going forward. First is a diagnostic contribution: the methodology of corpus linguistics helps to identify shortcomings in the law's current approach to identifying and assessing ordinary meaning. Second, corpus linguistic analysis can help advance the theory of interpretation. The tools and methodologies presented herein will aid in the development of a more sophisticated legal conception of ordinary meaning. Third, having identified the problem and laid out the requirements for a proposed solution, we advance the methodology of corpus linguistics as the best mechanism for yielding a satisfactory solution.

We grant that some problems of ordinary meaning will require resort to other tools or principles of interpretation for their resolution. Where the corpus data are inconclusive, or the distinction between two proposed definitions seems so thin that we doubt that it represents any real difference in human perception, we may need to look elsewhere to resolve the interpretive question presented. But that does not mean that the corpus data were unhelpful. It means that we looked at data—at comparatively empirical, falsifiable grounds for assessing ordinary communicative content—as a threshold step and decided we needed to go further to find a satisfactory answer.

CONCLUSION

Some points of analysis outlined here are necessarily tentative. That seems inevitable in the course of breaking new ground. We trust that some of the value in our contribution will be to spark further analysis and scholarship on the questions we have raised.

Moving forward, judges, lawyers, and linguists will need to collaborate to settle on some best practices in this emerging field. Some important questions to answer include methods for selecting the best corpus for a given type of ambiguity, standards for the appropriate sample size for a given search, standards for determining appropriate search terms and search methods for various types of inquiries, and the identification of suitable coding methods. Scholars have begun to explore these and other related questions.[335] Further work is in order. But we are confident that lawyers and linguists can work together to develop an orthodox set of methods that will refine an approach that is now in its infancy.

---

335. *See* James C. Phillips & Jesse A. Egbert, *A Concise How-To Guide for Law and Corpus Linguistics: Importing Principles and Practices from Survey and Content-Analysis Methodologies To Improve Corpus Design and Analysis*, 2017 BYU L. Rev. (forthcoming 2018) (arguing that principles and methodologies from survey and content-analysis methodologies need to be applied to corpus design, selection, and coding).

Linguists have observed that corpus linguistics generally "has not yet reached the stage where it can present a stable set of methodological procedures coupled to specific descriptive questions."[336] That is undoubtedly all the more true for the application of this tool to a brand new field. The law, after all, asks questions that linguists historically have not deemed important—concerning the average or "ordinary" understanding of a given term in a given linguistic setting. The methodology of corpus linguistics will undoubtedly experience growing pains as it is employed for new purposes. Yet linguists have noted elsewhere (more generally) that "[t]he observation that distributional corpus analysis has not reached" the stage at which we have embraced a set of widely accepted norms "is certainly not a reason to abandon the approach; rather, it defines a promising and exciting research [program]."[337] That is certainly true as to the application of corpus linguistics to the enterprise of judging ordinary meaning. Whatever its current limitations, "semantic analysis can, and indeed, should, turn to corpus methods."[338]

The need is acute when the interpretive task involves questions of law. Too much rides on the resolution of legal ambiguity to resolve the matter by means "fraught with the potential for bias and error."[339] If and when the law turns on an assessment of ordinary communicative content we must at least try to define and operationalize the inquiry with greater care. We see the approach outlined here as a step in that direction.

---

336. GEERAERTS, *supra* note 247, at 178.

337. *Id.*

338. Glynn, *supra* note 136, at 7.

339. State v. Rasabout, 2015 UT 72 ¶ 134, 356 P.3d 1258, 1290 (Lee, Associate C.J., concurring in part and concurring in the judgment).