JUDGING ORDINARY MEANING

Thomas R. Lee & Stephen C. Mouritsen[*]

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 Steve 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*) and a Seventh Circuit opinion of 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 of 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 of interpretation. And we consider and respond to criticisms of their use by lawyers and judges.

---

[*] Thomas R. Lee is Associate Chief Justice of the Utah Supreme Court and a Distinguished Lecturer of Law at Brigham Young University. Stephen C. Mouritsen is an Associate at the University of Chicago Law School and Adjunct Professor of Law and Corpus Linguistics at Brigham Young University.

The authors express thanks to those who commented on earlier drafts or offered insights in response to presentations in various conferences, symposia, and talks, including Randy Barnett, Will Baude, Ryan Doerfler, Ed Finegan, Jonah Gelbach, Dan Klerman, Kurt Lash, John Manning, Noah Messing, Gordon Smith, Larry Solan, Larry Solum, Eugene Volokh, and John Yoo. Thanks also to Brigham Young University and to the Georgetown University Law Center for co-sponsoring a conference on law and corpus linguistics, at which the ideas in this piece were initially vetted, and to the Olin-Searle-Smith Fellows in Law program for making possible Mr. Mouritsen's association with the University of Chicago Law School.

INTRODUCTION

A key component of the meaning we ascribe to law concerns its "communicative content"—the "linguistic meaning" of the words of a statute or regulation,[1] encompassing the "intended" meaning of the lawmaker or the "contextual meaning" understood by the public.[2] This is the threshold question for many problems of legal interpretation. It is the "standard picture" painted by most theories of interpretation, which starts with a search for the "ordinary communicative content" of the words of the law.[3]

"There are excellent reasons for the primacy of the ordinary meaning rule."[4] Most of them stem from the purported determinacy of the ordinary meaning inquiry. We speak of a search for meaning "not in the subjective, multiple mind of Congress but in the understanding of the objectively reasonable person."[5] And we generally conclude that the search for such meaning "matches up well with our understanding of what the *rule of law* entails:"[6] It assures notice to the public, protects reliance interests, assures consistency of application, and respects the will of the legislative body.[7] 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."[8]

---

[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, __- __ (2015) (speaking of these and other conceptions of the communicative or "conversational" content of the words of the law).

[3] *See* William Baude & Stephen E. Sachs, *The Law of Interpretation*, 130 HARV. L. REV. 1079, 1086 (2017) (speaking of the "standard picture," or "view that we can explain our legal norms by pointing to the ordinary communicative content of our legal texts," or in other words "an instrument's meaning as a matter of language"); *see also id.* at 1082 n.2 (borrowing the "standard picture" terminology from Mark Greenberg, *The Standard Picture and Its Discontents*, in 1 OXFORD STUDIES IN PHILOSOPHY OF LAW 39, 48 (Leslie Green & Brian Leiter eds., 2011)).

[4] WILLIAM N. ESKRIDGE, JR., INTERPRETING LAW: A PRIMER ON HOW TO READ STATUTES AND THE CONSTITUTION 39-40 (2016).

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

[6] ESKRIDGE, *supra* note 4, at 35.

[7] *See id*. at ("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.").

[8] *Id*.

This premise has taken hold in our courts. "[W]e're all textualists now."[9] That holds true at least in the sense that most all 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."[10]

Yet the academy has been less sure of the premises of this trend. Scholars generally endorse the value of determinacy but roundly doubt the judge's ability to find it in mere "communicative content" or "ordinary meaning" of statutory text.[11] There are two dimensions to this 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.[12]

We share these concerns but offer a different solution. In Part I of this article we show that the law has done a poor job conceptualizing the notion of ordinary meaning. And we demonstrate that "[u]ncertainty and division" in assessing such meaning "seem inevitable" under the methods resorted to by judges in this field.[13] But we do not see these problems as an invitation to *abandon* the search for the ordinary communicative

---

[9] Elena Kagan, *The Scalia Lecture: A Dialogue with Justice Kagan on the Reading of Statutes* at 8:28 (Nov. 17, 2015), http://today.law.harvard.edu/in-scali-lecture-kagan-discusses-statutory-interpretation [http://perma.cc/3BCF-FEFR].

[10] *See* JOHN F. MANNING & MATTHEW C. STEPHENSON, LEGISLATION AND REGULATION 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 such courts are engaged in an "effort[] to increase predictability in statutory interpretation," that they give primacy to text and decline to look to external sources of meaning if they find the text "plain," and asserting that "these state efforts . . . respond directly to the leading academic proposals advanced to make federal statutory interpretation more determinate").

[11] *See* Fallon, *supra* note __, at 1255–63 (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 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).

[12] *See* Fallon, *supra* note ___, 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 1269 (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).

[13] *Id*. at 1269.

content of the law in favor of "case-by-case" "interpretive eclecticism"—of choosing 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."[14] 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 worse."[15] This kind of "interpretation" overrides rather than protecting 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.

Granted, "we can't treat the meaning of [the law's] language as the only source of its legal effect."[16] Our law of interpretation may have good reasons to depart from the "standard picture"—to substitute "fake" answers to linguistic questions for real ones.[17] It is doubtless true, moreover, that some of our rules of interpretation dictate a "process" that "often looks nothing like a search for linguistic meaning."[18] But that is no reason to abandon the enterprise entirely. The better response—or at least the first step in the response—is to examine the "standard picture" more carefully. To do so we should seek a more careful conception of ordinary meaning and see if we can find better ways to measure it.

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 non-linguistic answer that vindicates policies more important than the ones advanced by the "standard picture." But it skips the key threshold question to assume that the "law of interpretation" *just is,*[19] or that its "fake" (non-linguistic) premises are sufficient to override the search for ordinary meaning.[20]The search for ordinary meaning is hard. But 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.

---

[14] *Id*. at 1305.

[15] Sunstein, *supra* note __, at 193–94.

[16] Baude & Sachs, *supra* note __, at 1088.

[17] *Id*. at 1096.

[18] *Id*. at 1088.

[19] *Cf.* Sunstein, *supra* note __, at 193 (asserting that there is nothing about interpretation that "just is").

[20] In other words maybe the "standard picture" doesn't "claim[] to be a picture of American law." Baude & Sachs, *supra* note __, at 1089. Maybe American law claims to be a reflection of the standard picture—of ordinary meaning. We should entertain that possibility. Or at least we should ask whether there are good reasons for flipping the picture.

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 of 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 law of interpretation. And we identify problems in the methods the law has been using to address these issues.

We begin by noting the circumstances in which the "standard picture" controls under the law of interpretation, highlighting three exemplary cases in which the ordinary communicative content of the words of a statute seem 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 our attempts to operationalize (or measure) it. After outlining these two sets of problems we introduce some theories and empirical methods utilized by linguists that may help us do a better job of delivering on the promise of an objective inquiry into ordinary meaning.[22] We then apply these tools—imported from a field called corpus linguistics—to our three exemplary cases. And we close by responding to actual and anticipated criticisms of our approach and by highlighting unresolved issues that must be addressed going forward.

---

[21] Judge Richard Posner framed the ordinary meaning question in this (empirical) way in his opinion in United States v. Costello, 666 F.3d. 1040, 1042 (7th Cir. 2012). There he proposed to answer this question using the results of a Google search. We think Posner's instincts were right but his methods fell a bit short, as explained below. See *infra* __.

[22] 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.*, State v. Rasabout, 2015 UT 72, ¶¶ __ - __; 356 P.3d. 1258 (Lee, J., concurring in the judgment and 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, ¶ __, n. __, 308 P.3d. 517 (Lee, J., for the majority; 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 limitation under Utah law); Baby E.Z. v. T.I.Z., 2011 UT 38, ¶ __ - __, 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)); People v. Harris, 885 N.W.2d 832, 838–39 n.29 (Mich. 2016) (opinion of the court per Zahra, J.) (citing Utah Supreme Court opinions in support of the methodology of corpus linguistics; relying on corpus linguistic data in support of the court's interpretation of the term "information" in Michigan statute forbidding use of "information" provided by law enforcement officer if compelled under threat of employment sanction); *id.* at 850 n.14 (Markman, J., dissenting) (also citing Utah Supreme Court opinions and also relying on corpus linguistic data, but drawing a different inference from the data).

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.[23]

Judges generally are even more sanguine about the matter. The caselaw 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.[24] Where such meaning is viewed as "plain," moreover, the courts consistently declare the interpretive enterprise to be at an end.[25] 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 extra-textual sources of meaning or intent.[26]

A variation on the theme applies in the realm of substantive canons of construction or principles of agency deference. The rule of lenity, for example, says that genuine ambiguities in criminal laws are resolved in favor of the defendant[27]; the converse is the notion that "the rule of lenity

---

[23] *See* Fallon, *supra* note 2, at __; and Baude & Sachs, *supra* note __.

[24] *See, e.g.*, Clark v. Rameker, 134 S. Ct. 2242, 2246 (2014) (". . . we 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., 132 S. Ct. 1702, 1706 (2012) ("Because the TVPA does not define the term "individual," we look first to the word's ordinary meaning."); Mac's Shell Serv., Inc. v. Shell Oil Products Co. LLC, 559 U.S. 175, 175–76 (2010) ("Because [the Act] does not further define ["termination" and "cancel"], they are given their ordinary meanings . . .").

[25] *See* William Baude & Ryan Doerfler, *The (Not So) Plain Meaning Rule*, 84 U. CHI. L. REV (forthcoming 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" "[i]f it's not plain").

[26] *See, e.g.*, Gluck, *supra* note 12, 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 analysis"); KENT GREENAWALT, STATUTORY INTERPRETATION: 20 QUESTIONS 35 (1999) ("No one seriously doubts that interpretation of statutes turns largely on textual meaning").

[27] How much ambiguity, of course, is a difficult question. *See* Abramski v. United States, 134 S. Ct. 2259, 2272 (2014) (asserting that the rule applies only if the "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"); *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); ANTONIN SCALIA & BRYAN GARNER, READING LAW: THE INTERPRETATION OF LEGAL TEXTS 299 (2012) (decrying the "multiplicity of expressed standards" for invoking the rule of lenity,

has no application when the statute is clear."[28] And *Chevron* deference is similar. The courts defer to agencies only where the terms of the statute are ambiguous.[29] So both of these principles likewise implicate the question of the degree of plainness of statutory language.

This is not an exhaustive list of the courts' consideration of the ordinary communicative content of statutory language. But it is enough to highlight the point that the law at least sometimes credits that content in the enterprise of legal interpretation.

That said, 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."[30] The same goes even for "plain" meaning. "Courts and commentators 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 only very limited information about the context of the utterance."[31] "Other times 'plain meaning' is used to denote *obvious* meaning—i.e., the meaning that is clear."[32] (This is the sense at work in the "plain meaning rule."[33])

Second is a problem of operationalization or measurement. The concern here is that even if we could settle on a sense of ordinary or plain meaning we are unsure how to assess it. "Uncertainty and division seem inevitable."[34] 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.[35] And the problem is underscored by the tools (mis)used by

---

"leav[ing] open the crucial question . . . of how much ambiguousness constitutes an ambiguity").

[28] *See* SCALIA & GARNER, *supra* note __, at 301.

[29] *See* Michigan v. E.P.A., 135 S. Ct. 2699, 2707 (2015).

[30] WILLIAM N. ESKRIDGE, JR., PHILIP P. FRICKEY & ELIZABETH GARRETT, 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"; 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 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").

[31] Baude & Doerfler, *supra* note __, at 5.

[32] *Id.*

[33] *See id.*

[34] Fallon, *supra* note __, at 1268.

[35] *Id.*

judges to try to answer this empirical question (resort to a dictionary or a word's etymology, for example, as explored further below).

The theoretical and measurement problems with 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 these:

· Is a person guilty of *carrying a firearm* (under a federal sentencing enhancement provision) in connection with a drug crime if he merely 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*, 524 U.S. 125 (1998). 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. And 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 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 them in litigation? This question arose in *Taniguchi v. Kan Pacific Saipan, Ltd.*, __ U.S. __, 132 S. Ct. 1997 (2012). Again 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 both 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. And they also disagreed about what the search for *ordinary meaning* ultimately entailed, with the majority insisting that only the more common sense of the term was covered, and the dissent asserting that a permissible sense should also count.

· Is a woman who allows her "illegal alien" boyfriend 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*, 666 F.3d. 1040. 1042 (7th Cir. 2012). *Costello* is like *Muscarello* and *Taniguchi*: it involves 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? Judge Richard Posner authored the opinion for the court. And he recognized the deficiencies of standard methods—principally, dictionaries—in answering that question. So he proceeded to a search

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

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 spectrum:

POSSIBLE ⟹ COMMON ⟹ MOST FREQUENT ⟹ EXCLUSIVE

At the left end of the spectrum 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; so we might note that a given sense of a term is not common or usual 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 one more point to the continuum. When courts speak of plain meaning (as a construct distinct from ordinary meaning) they generally mean to "denote *obvious* meaning" or "meaning that is clear."[36] 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.[37] One attested sense of *vehicle* is the notion of a "carrier," "agent of transmission."[38] And 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 pet dog or cat, which could be referred to as a *vehicle* (as in a pet 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

[36] Baude & Doerfler, *supra* note __, at 5.

[37] H.L.A. Hart, *Positivism and the Separation of Law and Morals*, 71 HARV. L. REV. 593, 606–15 (1958).

[38] WEBSTER'S THIRD NEW INTERNATIONAL DICTIONARY 2538 (1961).

plain or clear, or in other words 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."[39] This sense of *vehicle* could easily be viewed as a common sense of vehicle—certainly more common than the infection *carrier* sense noted above. But even this sense 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,[40] we might understand the term to encompass only *motor vehicles*, and thus not to cover the bicycle. (Here we are making 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.[41])

The four points on the frequency continuum do not completely capture the range of senses of ordinary meaning embraced by our courts. Sometimes our judges seem to have reference to a fifth notion of ordinary—a notion of linguistic *prototype*.[42] 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 of prototype (the clearest cases, best examples) of the category."[43] Thus, prototype analysis has shown that people consider *chair* to be a more prototypical example of *furniture* than *stool*,[44]

---

[39] *Id.*

[40] This notion of "ordinariness" is attested in the Oxford English Dictionary. *See* 10 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 A. Gales, *Finding Ordinary Meaning in Law: The Judge, the Dictionary or the Corpus?* 1 INT'L J. OF L. 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.").

[41] *See infra* Part III.C.

[42] *See* McBoyle v. United States, 238 U.S. 25, 27 (1931) (per Holmes, J.) (determining whether an "airplane" was a "vehicle" for the purposes 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.

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

[44] *Id.* at 229.

automobile to be a more prototypical *vehicle* than *yacht*,[45] and *robin* to be a more prototypical bird than *ostrich*.[46]

And that is another way to conceive of the notion of ordinary meaning in the law.[47] 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.[48] 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.[49]

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

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."[50] Frank Muscarello was convicted on drug charges in a case in which he was shown to have transported a gun to the drug deal in the locked glove compartment of his car. The question presented was whether that counted as carrying. Both the majority (Breyer) and dissenting (Ginsburg) opinions agreed that the proper interpretation of "carries a firearm" came

[45] *Id.* at 230.

[46] *Id.* at 232.

[47] *See* Lawrence M. Solan, *Why Laws Work Pretty Well, but Not Great: Words and Rules in Legal Interpretation*, 26 L. & SOC. INQUIRY 243, 258 (2001) (citing Smith v. United States, 508 U.S. 223 (1993) ("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."); *see also* ESKRIDGE ET AL., *supra* note 27, at 850 (discussing prototypical meaning in the context of statutory interpretation); Lawrence Solan, Terri Rosenblatt & Daniel Osherson, *False Consensus Bias in Contract Interpretation*, 108 COLUM. L. REV. 1268, 1276–80 (2008); Lawrence M. Solan, *The New Textualists' New Text*, 38 LOY. L.A. L. REV. 2027, 2042–46 (2005); Note, *The Supreme Court 1997 Term, Leading Cases*, 112 HARV. L. REV. 355, 361–62 (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.").

[48] Lawrence M. Solan, *Law, Language, and Lenity*, 40 WM. & MARY L. REV. 57, 67–68 (1998).

[49] 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.

[50] 18 U.S.C. § 924(c)(1).

down to the "ordinary English meaning" of that "phrase."[51] Yet neither opinion settled on a single sense of "ordinary." Both opinions slide back and forth along the continuum, without acknowledging that they are doing so.

At one point Breyer seems to employ a merely "common" sense of ordinary; he does so in asserting 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 a New York Times database reflect that sense,[52] and in concluding that "the word 'carry' in its ordinary sense *includes* carrying in a car."[53] Yet Elsewhere Breyer seems to speak of the car-carrying sense as most frequent. He does so in asserting (a) that the "ordinary English" sense of *carry* is to *transport* it in a vehicle; (b) the *bear* personally sense is "special"; and (c) "we believe Congress intended to use the word in its primary sense and *not* in this latter, special way."[54]

Justice Ginsgurg's dissent is also inconsistent. In asserting 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.[55] That is the language of possibility or commonality.[56] Elsewhere, however, Justice Ginsburg seems to speak in terms of personally *bearing* as the most frequent sense of the term—in noting, in response to Justice Breyer's statistics, "what meaning showed up some two-thirds of the time."[57]

2. *Taniguchi v. Kan Pacific Saipan*

Taniguchi was a personal injury case.[58] 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.[59] The defense "paid to have various documents translated from Japanese to English."[60] 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

[51] Muscarello v. United States, 524 U.S. 125 at 127–28 (1998).

[52] *Id*. at 129.

[53] *Id*. at 131 (emphasis added).

[54] Muscarello, 524 U.S. at 128 (emphasis added).

[55] *Id.* at 148 (Ginsburg, J., dissenting) (emphasis added).

[56] *See 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).

[57] *Id.* at 143.

[58] Taniguchi v. Kan Pac. Saipan, 132 S.Ct. 1997 at 2000 (2012).

[59] *Id.*

[60] *Id.*

operative language was from a statute allowing the prevailing party in federal litigation to recover certain costs, including those incurred by an "interpreter."[61]

And here the case seemed to turn on the operative notion of ordinary meaning. Justice Alito wrote for the majority and concluded that the ordinary sense of *interpreter* is *oral translator*. Alito finds *written translator* possible but "hardly a common or ordinary meaning."[62] He says that "an interpreter is normally understood as one who translates orally from one language to another."[63] And he concludes that the *written translator* sense is not "truly common," but is rather "obsolete," citing dictionaries to support that conclusion.[64]

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."[65] This is a core basis of the view of the *Taniguchi* dissenters. They do not expressly disagree with Alito's assertion that the *oral translator* notion is most common; they are simply saying that either of two common senses of a term should count as ordinary.[66]

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."[67] Her alleged crime was essentially in "having permitted" her "illegal alien" "boyfriend to live with her."[68] And 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, for the majority, warns of the perils of overreliance on the dictionary to resolve questions of ordinary meaning. (More on that below.) And he directs the ordinary meaning analysis to an empirical inquiry—which he proposes to resolve by means of a Google search.

---

[61] *Id.* at 2003.

[62] *Id.*

[63] *Id.* at 2004.

[64] *Id.*

[65] *Taniguchi*, 132 at 2007 (Ginsburg, J., dissenting).

[66] *See id.* at 2008 (asserting that the *written translator* sense is an "acceptable usage" even if it is "not 'the most common usage'").

[67] 8 U.S.C. § 1324(a)(1)(B)(ii).

[68] United States v. Costello, 666 F.3d 1040, 1043 (2012).

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." And because 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."[69] This is a "most frequent" sense of ordinariness—and a blatantly empirical sense 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.'"[70] In support of this point Manion cites definitions from dictionaries in print 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."[71]

* * *

Our judges purport to be speaking of a consistent, common sense of ordinary meaning. But we switch back and forth between different senses of ordinary meaning, usually without acknowledging the inconsistency. Sometimes (as in *Muscarello*) our 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—with the majority selecting one point on the frequency continuum that leads to one result and the dissent opting for a different point that yields the opposite outcome. This is problematic—not just for statutory interpretation, but for the rule of law.[72]

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

---

[69] *Id.*

[70] *Id.* at 1052 (Manion, J., dissenting) (emphasis added).

[71] *Id.* (citing WEBSTER'S NEW INTERNATIONAL DICTIONARY OF THE ENGLISH LANGUAGE 981 (1917) (with a definition of *harbor* of "[t]o afford lodging to; to entertain as a guest; to shelter; to receive; to give refuge to")); WEBSTER'S NEW COLLEGIATE DICTIONARY 376 (John P. Bethel et al. eds., 1953) (including a definition of *harbor* as "to entertain as a guest; to shelter; to give a refuge to").

[72] *See* William Baude & Stephen E. Sachs, *The Law of Interpretation*, 130 HARV. L. REV. 1079, 1089 (2017) ("[W]e have to decide which meaning, produced by which theory of meaning, we ought to pick.").

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 doesn't count.

These conclusions seem uncontroversial. But the judge who makes them is making an empirical assessment. Gut-level empirics probably won't bother us if they go only to a holding that a pet or a bicycle is not a *vehicle* prohibited in the park. But what about a motorized scooter or a golf cart? Are they covered? 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. Two primary grounds are dictionaries and 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."[73] 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 of course troubling; a judge who cherry-picks a preferred dictionary definition while ignoring an alternative is clearly 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 both the automobile and conveyance notions but find a basis for embracing one as ordinary. We see two principal means of doing so in the cases: (a) preferring the definition that appears first in a dictionary's list of senses, and (b) citing

[73] State v. Rasabout, 2015 UT 72, ¶ 53, 356 P.3d 1258 (Lee, 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 "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, 899 F.2d 15, 15 (6th Cir. 1990) (citing only one definition of "operator" in determining the ordinary meaning even though opposing definitions existed).

the etymology of the statutory term. Neither of these approaches is defensible, however, for reasons explained 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 here and turns to Google, albeit in a manner raising 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.[74] But he embraces the former sense as the "primary" one and dismissed the latter as "special."[75] And his first argument in support of that conclusion is that "[t]he Oxford English Dictionary gives as its *first* definition 'convey, originally by car or wagon, hence in any vehicle, by ship, on horseback, etc.'"[76] The italicized emphasis on "first" is Breyer's. Breyer's 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)"[77]) and the "*first* definition" in the Random House Dictionary ("'to take or support from one place to another; convey; transport.'"[78]

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 "*twenty-sixth* definition" in the OED.[79] This seems clearly to be the threshold basis of Breyer's conclusion that "[t]he relevant linguistic fact[]" is that "the word 'carry' in its ordinary sense includes carrying in a car."[80]

Breyer also turns to etymology, in asserting that "[t]he origin of the word 'carries' explained why the first, or basic, meaning of the word 'carry' includes conveyance in a vehicle."[81] 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."[82] The precise premises of Breyer's analysis are left implicit. But the implicit point seems clear—the etymology of the verb *carry* confirms that

[74] 524 U.S. at 128.

[75] *Id.*

[76] *Id.* (citing 2 OXFORD ENGLISH DICTIONARY 919 (2d ed. 1989)).

[77] *Id.* (citing WEBSTER'S THIRD NEW INTERNATIONAL DICTIONARY 343 (1986)).

[78] *Id.* (citing RANDOM HOUSE DICTIONARY OF THE ENGLISH LANGUAGE UNABRIDGED 319 (2d ed. 1987)).

[79] 524 U.S. at 130 (quoting 2 OXFORD ENGLISH DICTIONARY at 921).

[80] *Id.*

[81] *Id.* at 128 (citing BARNHART DICTIONARY OF ETYMOLOGY 146 (1988)).

[82] *Id.*

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 going to 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 relative frequency of listed senses of a given word.[83] Many "dictionaries simply rank their definitions according to evidence of *historical* usage."[84] Others openly "disavow[] any attempt to establish a hierarchy of ordinariness in the ranking of . . . senses, admitting that sometimes an 'arbitrary' listing of senses is used."[85] Thus, historical ordering is unhelpful in the assessment of ordinariness. If anything, an older sense may be less likely to be the more common one today.

That is also the problem with etymology. If our usage and understanding of a word have evolved over time, as they often will, the historical pedigree of a word may direct us to an outmoded or even obsolete definition—to the notion that *December* is the tenth month, or that an *anthology* is a bouquet of flowers.[86] That is entirely possible as to

---

[83] *See* Stephen C. Mouritsen, *The Dictionary Is Not a Fortress: Definitional Fallacies and the Corpus-Based Approach to Plain Meaning*, 2010 BYU L. REV. 1915, 1924-25 (discussing problems with dictionary usage by courts; identifying the "sense-ranking fallacy"). The *Random House Dictionary of the English Language, Unabridged* appears to be an exception. Its front matter states that "the most frequently encountered meaning generally comes before less common ones. Specialized senses follow those in the common vocabulary, and rare, archaic, and obsolete senses are listed last." RANDOM HOUSE DICTIONARY OF THE ENGLISH LANGUAGE UNABRIDGED xxii (2d ed. 1987). But we see grounds for skepticism of these sorts of claims. *See infra* ___. And 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.

[84] *Id.*; *see also* 1 OXFORD ENGLISH DICTIONARY xxix (2d ed. 1989) ("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 17a (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").

[85] *Id.*

[86] *December*, THE BARNHART CONCISE DICTIONARY OF ETYMOLOGY 188 (Robert K. Barnhart ed,, 1995) ("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)."); *Anthology*, THE BARNHART CONCISE DICTIONARY OF ETYMOLOGY 29 (Robert K. Barnhart ed., 1995) ("1640, collection of the 'flowers' of verse (i.e. small, choice poems) by various authors; borrowed, perhaps by influence of French *anthologie*, from Greek *anthologí*ā flower-gathering (*ánthos* flower + *légein* gather).").

Breyer's analysis of *carry*; the etymology of the word tells us nothing about its ordinary meaning today.

2. *Taniguchi v. Kan Pacific Saipan*

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 *oral translator*.[87] And he says that the "sense dividers" in the cited dictionaries confirm the court's holding—in designating the *oral translator* notion one that is "especially" indicated and flagging the *written translator* sense as "obsolete."[88]

Yet Alito's approach is still problematic. The "survey" of dictionaries is far from systematic. 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 presented for our review.[89]

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.[90]

Second, sense dividers are not reliable measures. Dictionaries tell us very little about the basis for the "obsolete" sense designation. And 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 be a matter no

---

[87] 132 S.Ct. at 2003.

[88] *Id.* at 2002–03; *id.* at 2003 (noting that the *Oxford English Dictionary* "designated [the written translator] meaning as obsolete"); *id.* at 2003 & 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).

[89] Even a documented survey of every known dictionary might not be sufficient, moreover, for reasons explained below. *See infra* ___.

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

longer in use; but without more such designation gives us only the opinion of those lexicographers, and not a hard basis for an empirical conclusion.[91]

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.[92] It suggests the notion that the two senses are not highly distinct from each other, but instead are more 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.'"[93] 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.

So for these reasons the *Taniguchi* opinion also employs inadequate tools of measurement. Justice Alito's "survey" and sense designations seem more sophisticated but they are not ultimately good tools of assessing empirical questions of ordinary meaning.

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."[94] And for that reason Posner turns to Google to get a "rough index of the frequency of [*harbor*'s] use."[95] 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 web is enormous, free, immediately available, and largely linguistic.[96] And "it is appealing to use the web as a data source"

---

[91] *See also* DOUGLAS BIBER, CORPUS LINGUISTICS: INVESTIGATING LANGUAGE STRUCTURE AND USE 39 (1998) (observing that "citation slips" used by lexicographers "represent only those contexts that a human reader happens to notice").

[92] *See infra* Part III.

[93] 132 S.Ct. at 2003 (citing WEBSTER'S THIRD NEW INTERNATIONAL DICTIONARY 1182 (1976)).

[94] *Id.*

[95] *Costello*, 666 F. 3d at 1042 (2012).

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

because "language analysis and generation benefit from big data.[97] 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 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. If we wish to investigate the biases, the area we become expert in is googleology not linguistics."[98] Google returns can vary by geography, by time of day, and from day to day.[99] So Google search results are rather unscientific.

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 (but only contemporary web pages). 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 Posner formulated his own set of search terms—comparing hit counts for phrases like "harboring fugitives" and "harboring guests," and comparing relative hit counts for his two sets of search terms. But that introduces another set of problems. Posner gives no basis for his chosen set of search terms, and the terms he chose seem likely to affect the outcome.

Finally, even assuming away the above problems, the hit counts that Posner relies on may not be indicative of ordinariness in the sense of frequency of usage. 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.[100] A single web page may have 10s, 100s or 1,000s 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.

For all these reasons we think Judge Posner was onto something in seeking an empirical method of measurement. But we think his Google search was inadequate.

II. THEORIZING ORDINARY MEANING

---

[97] *Id.*
[98] *Id.*
[99] *Id.*
[100] *Id.*

Before we turn to the measurement problem we must start by refining our theory of ordinary meaning. 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"[101] and the Baude-Sachs idea of the "reader's understanding"[102]); 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.[103]

Not all of these conceptions fit within the ordinary meaning construct that is our subject 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, the law of interpretation—informed by principles of stare decisis—can (and should) yield due deference to the "interpreted meaning" established by precedent. But our reasons for respecting such meaning have nothing to do with the rule of law premises behind the law's search for the ordinary communicative content of the words of the law. They stem from principles of stare decisis.[104]

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 Scalia and Garner in *Reading Law*. And the inquiry is framed in objective-sounding terms—in a search for "objectified intent." But on closer review this notion of meaning has nothing to do with actual communicative content of the words of the law (or, accordingly, of intentions to attribute to lawmakers or reliance interests developed by the public). It is a highly constructive inquiry aimed at an impossibly well-informed ideal 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."[105]

---

[101] Sunstein, *supra* note __, at 198.

[102] Baude & Sachs, *supra* note __, at 1090 (distinguishing the "author's intent" and the "reader's understanding").

[103] *See* Fallon, *supra* note __, at 1255–63.

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

[105] *Id.*

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.[106]

That leaves (a) semantic meaning, (b) contextual meaning (public meaning or the reader's understanding); and (c) intended meaning. Semantic meaning, in Fallon's taxonomy, 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."[107] And 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."[108]

The point here is the notion that "literalist" textualism excludes non-semantic 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."[109] Here he cites an example from *Reading Law*: "'*Nail* in a regulation governing a beauty salon has a different meaning from *nail* in a municipal building code.'"[110]

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 non-semantic context—pragmatics—into account in deriving meaning from language.[111] And for that reason we see no basis to credit semantic

---

[106] *See* Gary Lawson & Guy Siedman, *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").

[107] Fallon, *supra* note __, at 1245.

[108] *Id.*

[109] *Id.* at 1246.

[110] *Id.* (quoting SCALIA & GARNER, *supra* note __, at 20).

[111] In the law, pragmatism can refer to "basing judgments (legal or otherwise) on consequences, rather than on dedication from premises in the manner of syllogism." Richard A. Posner, How Judges Think 40 (2010). In linguistics, pragmatics can refer to the study of specific linguistic phenomena like conversational implicature or deixis.

meaning without consideration of pragmatic context. 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 would go for the public governed by the law. If no one reads laws literally by pure semantics, we should have no reason to protect reliance interests or notice concerns rooted in that kind of understanding.

That does not make the elements of semantic meaning irrelevant. We humans do take verbal, semantic context into account in interpreting language. It's just that we also take nonverbal, pragmatic context into account. And we think it important to highlight each of these elements of context that might affect our understanding—further 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. 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.[112] 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.[113]

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. And a decision to credit such meaning is appropriate to the extent we find the policy of protecting the intentions of lawmakers to predominate. 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."[114] And 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 for this meaning.[115]

Alan Cruse, Meaning In Language: An Introduction to Semantics and Pragmatics 317-97 (2004). Here we use the term generally with reference to non-verbal context that may affect meaning.

[112] Baude & Sachs, *supra* note __, at 1091.

[113] *Id.* at 1090.

[114] Larry Alexander, *Telepathic Law*, 27 CONST. COMMENT. 139, 140 (2010).

[115] *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 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

But 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.[116] Enforcing "hard-to-find intentions" of lawmakers "would make the law unpredictable and arbitrary."[117] 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 (again, as properly informed by relevant elements of semantic and pragmatic context—as described below).

So again this is a viable construct for our inquiry into ordinary meaning. Before framing the theory of meaning in 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 also tell us that we must 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 a range of components: lexicography, syntax, and semantics.

1. Lexicography

---

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 supra* __.

[116] *See* Antonin Scalia, Common-Law Courts in a Civil-Law System: The Role of United States Federal Courts in Interpreting the Constitution and Laws, Address at the Tanner Lectures on Human Values at Princeton University (March 8–9, 1995) at 92, *available at* http://tannerlectures.utah.edu/_documents/a-to-z/s/scalia97.pdf. (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"); 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."); Michael Herz, *Purposivism and Institutional Competence in Statutory Interpretation,* 2009 MICH. ST. L. REV. 89, 102 (2009) (arguing that "the case for textualism" is in part "[t]he claim . . . that if legal rules are embedded in publically available texts, affected person 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").

[117] Baude & Sachs, *supra* note __, at 1091.

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 macrostructure of 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 put it, "that someone sat there and wrote the dictionary, and we speak as though there were only one dictionary, whose lexicographer got all the definitions 'right' in some sense that defies analysis."[118] But that is not linguistic reality. Dictionaries may differ sharply in the number of senses they assign to a given term or in the divider they use to distinguish senses. Dictionaries do not "emerge from some lexicographical Sinai; they are the products of human beings. And human beings, try as they may, bring their prejudices and biases into the dictionaries they make."[119]

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."[120] This "has always been a burning issue of lexical semantics."[121] And linguists also acknowledge that the sense distinctions reflected in dictionaries are "more of a descriptive device rather than a claim about psycholinguistic reality."[122]

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*).[123] Words are then divided into senses based on a variety of factors, including their part of speech, pronunciation, inflection, etymology, and shades of meaning.[124] This approach to defining words and dividing them

---

[118] Lawrence Solan, *When Judges Use Dictionaries*, 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.").

[119] JONATHAN GREEN, CHASING THE SUN: DICTIONARY MAKERS AND THE DICTIONARIES THEY MADE xiv (1997).

[120] Nikola Dobric, *The Predictive Power of the (Micro)Context Revisited—Behavioral Profiling and Word Sense Disambiguation*, UDC 77 (2014), https://www.researchgate.net/publication/270285812_The_Predictive_Power_of_the_MicroContext_Revisited_-_Behavioral_Profiling_and_Word_Sense_Disambiguation.

[121] *Id.*

[122] Dylan Glynn, *Polysemy and Synonymy*, *in* CORPUS METHODS FOR SEMANTICS: QUANTITATIVE STUDIES IN POLYSEMY AND SYNONYMY 10 (Dylan Glynn & Justyna A. Robinson eds., 2014).

[123] SIDNEY I. LANDAU, DICTIONARIES: THE ART & CRAFT OF LEXICOGRAPHY 153 (2d ed. 2004).

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

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*.)[125] Moreover, the citation or quotation files from which many contemporary dictionaries are derived were collected without the benefit of modern sampling methods. Accordingly, as noted in Part I above, they cannot not be relied upon for information about the frequency of a given word or word sense.[126]

Contemporary lexicographers have moved past reliance on citation files and have begun to rely on electronic "corpora"—large bodies or databases of naturally occurring language—to gather linguistic data. 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.[127]

We should be sensitive to the challenges presented by sense division in our assessment of ordinary meaning. We should not jump immediately to the conclusion that two senses listed separately in a dictionary reflect "psycholinguistic reality"—that a distinct *listing* in a dictionary tells us that human beings will understand distinctly listed senses as mutually exclusive constructs. Nor should we assume the converse—that that the lack of distinct listing tells us that human beings will not perceive a distinction.

Instead we should look to what is less arbitrary and more readily measurable—to patterns that emerge from corpus linguistic data, or to what we can glean about human perception from psycholinguistic surveys and experiments. The point can be made 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 distinctly separate from each other.[128] But if the question of the dividing line is in some sense arbitrary and not reflective of

---

[125] ANN O'KEEFFE & MICHAEL MCCARTHY, THE ROUTLEDGE HANDBOOK OF CORPUS LINGUISTICS 433 (2010).

[126] LANDAU, *supra* note __, at 153.

[127] 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 [a word] and other expressions of the same language.").

[128] *See* DOUGLAS BIBER, SUSAN CONRAD, RANDI REPPEN, CORPUS LINGUISTICS: INVESTIGATING LANGUAGE STRUCTURE AND USE 40 (1998) (documenting the differences in the definition and sense distribution of the noun "deal" as recorded in five general-use dictionaries).

psycholinguistic reality, then we ought to seek to measure psycholinguistic reality 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 find distinct definitions encompassing *oral translator* and *written translator*. But that may not tell us that these two senses represent psycholinguistic reality—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." And the *written translator* idea could certainly be viewed as an *atypical* example. That may be all that dictionaries are telling us by the indication 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*. (More on solving this puzzle in Part III.C.)

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.[129] Yet they rarely say what they mean by context. Linguistic theory can help identify which elements of context may matter.

Context can be viewed as encompassing both verbal and non-verbal components.[130] The verbal context of a word or phrase in a statute

[129] *See* Johnson v. United States, 559 U.S. 133, 139 (2010) (observing that in the ordinary meaning inquiry, "[u]ltimately, context determines meaning"); Braunstein v. C.I.R., 374 U.S. 65, 70 (1963) (defining the ordinary meaning of "gain" in a particular context); 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"). 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.").

[130] Charles Goodwin & Alessandro Duranti, *Rethinking Context: An Introduction*, *in* RETHINKING CONTEXT: LANGUAGE AS AN INTERACTIVE PHENOMENON 6–9 (1991).

includes its *syntactic* and *semantic* environments. Syntax is a set of rules and principles that governs sentence formation and determinates which sentences will convey meaning to members of the same speech community.[131] One aspect of syntax is *argument structure*, a linguistic term of art that has reference to the participants in the action of a verb.[132] 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.[133] 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.[134] These features include concepts like number, animacy, gender, humanness, and concreteness (*i.e.*, tangibleness).[135] In semantic theory, words can also be understood with reference to their functional role in an utterance. A word has an *agentive* function if it is 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. [136]

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—

---

[131] 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.").

[132] ALAN CRUSE, MEANING IN LANGUAGE: AN INTRODUCTION TO SEMANTICS AND PRAGMATICS §§ 14.1-5 (2011). An intransitive verb (*fall, die, yawn*) has one argument—the subject. A di-transitive verb (*throw, send*) has three arguments—the subject, the direct object, and the indirect object.

[133] MICHAEL MORRIS, AN INTRODUCTION TO THE PHILOSOPHY OF LANGUAGE 152 (2007) ("Semantics is the attempt to give a systematic explanation of how the meaning of sentences depends upon the meaning of their parts.").

[134] JOHN I. SAEED, SEMANTICS 260–62 (20XX).

[135] *Id.*

[136] 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 involve 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 ___, at § 14.5.

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 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 doesn't instigate it and doesn't 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.[137]

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.[138] Such considerations may encompass the physical or social setting of an utterance, and even an inference about the intent of the speaker.[139]

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

---

[137] *See* Muscarello, 524 U.S. at 143 (Ginsburg, J., dissenting) (noting that the issue presented "is not 'carries' at large but 'carries a firearm'").

[138] *See* Goodwin & Duranti, *supra* note __, at 6-9; CRUSE, *supra* note __. 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).

[139] 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* Larry Alexander & Saikrishna Prakash, *"Is That English You're Speaking?": Why Intention Free Interpretation is an Impossibility*, 41 SAN DIEGO L. 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."); Scalia, *Response*, *in* A MATTER OF INTERPRETATION: FEDERAL COURTS AND THE LAW 144 (1998) (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").

grass."[140] 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's possible. And we may need to take such context into account in assessing ordinary meaning.

C. Meaning as of When?

Human language is in a constant state of change.[141] But it does not change at a predictable rate.[142] Nor do different linguistic features change at the same time.[143] Our theory of ordinary meaning must take account of this variation and allow us to examine the linguistic norms prevailing at different historical periods.

We 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*.[144] 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 was reflective of 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

---

[140] *See* RICHARD A. POSNER, LAW, PRAGMATISM, AND DEMOCRACY (2005) (making the case for a pragmatic theory of government); RICHARD A. POSNER, HOW JUDGES THINK 40 (2010) ("[Pragmatism] refers to basing judgments (legal or otherwise) on consequences, rather than on dedication from premises in the manner of syllogism.").

[141] JOHN LYONS, INTRODUCTION TO THEORETICAL LINGUISTICS 43 (1968) ("All language 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.").

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

[143] *Id.*

[144] *Costello*, 666 F.3d at 1043–44.

modern times—so Judge Posner appears to have concluded that that gave him license to do the same thing.

That phenomenon is at least 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*.[145] That 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 we are bound by their views.[146] 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."[147] And 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 at least arguably leads in a different direction. If we are seeking to protect reliance interests and fair notice we are at least arguably opting for contemporary (not historical) evidence of ordinary meaning. "Normal" English speakers are guided by their contemporary understanding. And they lack a sophisticated understanding of historical usage. So if we are trying to assess public meaning we may be assuming away the temporal premises of originalism, and searching only for contemporary meaning.

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, no more than contemporary ones, cannot yield reliable information about which of various senses is more ordinary. And the problems are compounded as to historical dictionaries.[148]

---

[145] *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 v. Ziska, 334 P.3d 964, 967 (Or. 2014); State Bd. of Nursing v. Ruebke, 913 P.2d 142, 157 (Kan. 1996).

[146] Alexander, *supra* note __, at 141.

[147] *Id.*

[148] *See* Rickie Sonpal, *Old Dictionaries and New Textualists*, 71 FORDHAM L. REV. 2177, 2209 (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.").

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. But we ought to regard them with the skepticism when they are offered as evidence of "ordinary" or "original" meaning.

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.[149] 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 must be able to address the differences in various linguistic registers.

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 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. And we may also wish to take into account the pragmatic consideration that the more formal nature of legal language can affect human understanding of meaning.[150]

---

[149] 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."); SANDRA LEE MCKAY, SOCIOLINGUISTICS AND LANGUAGE TEACHING 49 (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 of the conduct and interpretation of speech."); John Sinclair, *Meaning in the Framework of Corpus Linguistics*, 20 LEXICOGRAPHICA 1, 22 (2004) ("The differences in interpretation between members of a speech community are small and they do not interfere much with normal communication."); Reed Dickerson, *Statutory Interpretation: Dipping Into Legislative History*, 11 HOFSTRA L. REV. 1125, 1154 (1982) (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").

[150] *But see* Doerfler, *supra* note __, at 983–84 (articulating a "conversation" model of "fictionalized" legislative intent, in which "[a]n interpreter occupies the position of

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. 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 semantic, pragmatic, temporal, and speech community considerations that may be relevant.

That leaves the question of measurement or operationalization. Again we can find help in linguistics. We propose the use of tools employed in corpus linguistics. Corpus linguistics is an empirical approach[151] to the study of language that involves large, electronic databases of text known as corpora (the plural of *corpus*).[152] A corpus is a body or database of naturally occurring language.[153] 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.[154] The defining characteristic of corpus linguistics is "the claim that it is possible to actually 'represent' a domain of language

---

conversational participant, hearing statements directed at her and other participants" and credits "information salient to both members of Congress and to citizens").

[151] TONY MCENERY & ANDREW HARDIE, CORPUS LINGUISTICS: METHOD, THEORY AND PRACTICE 49 (2011) ("Empiricism lies at the core of corpus linguistics"); Paul Baker, 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 analyzing real examples of language as it is actually used. Corpus linguistics is therefore a strongly empirical methodology.").

[152] MCENERY & HARDIE, *supra* note ___, at i.

[153] Douglas Biber, *Corpus-Based and Corpus-driven Analyses of Language Variation and Use*, *in* THE OXFORD HANDBOOK OF LINGUISTIC ANALYSIS 159 (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 . . . . It utilizes a large and principled collection of natural texts, known as a "corpus", as the basis for analysis.").

[154] *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 include casual face-to-face conversation, service encounters, lectures, sermons, political debates, etc.; and, in writing, these include email messages, text-messaging, newspaper editorials, academic research articles, etc.").

use with a corpus of texts, and possible to empirically describe linguistic patterns of use through analysis of that corpus."[155]

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.[156]

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.[157] And if the search for ordinary meaning entails an analysis of the relative frequency of competing senses of a given term, then corpus linguistics seems the most promising tool.

Such data can also inform our assessment of linguistic prototype.[158] 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 measurement of the components of ordinary meaning identified above. Then we illustrate the utility of those tools by applying them to the cases and examples discussed throughout the article. And we conclude this part of the article with some observations about inferences that can be drawn from the data about the ordinary meaning of *vehicle*, *carry a firearm*, *interpreter*, and *harbor*.

---

[155] BIBER & REPPEN, *supra* note ___, at [***].

[156] McEnery & Hardie, *supra* note __, 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 the data they produce is made available to analysts in the future.").

[157] Assuming, of course, the corpora used are properly constructed such that they enable us to make generalizations about a larger population. *See* Douglas Biber, *Representativeness in Corpus Design*, 8 LITERARY & LINGUISTIC COMPUTING 243 (1993); JESSE EGBERT, B. GRAY, & DOUGLAS BIBER, DESIGNING AND EVALUATING LANGUAGE CORPORA (forthcoming).

[158] 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 649–50 (Claudia Maienborn, Klaus von Heusinger & Paul Portner eds., 2011) ("In response to the question 'where does prototypicality come from?' (Geeraerts 1988), 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 the 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.").

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. S*pecial* corpora are limited to a particular genre, register, or dialect.[159] 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 that 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 *carries*, *carrying*, and *carried*) and to search for results related 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 that searches for the base form of a word and 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. [160]

---

[159] For example, the Linguistic Data Consortium at the University of Pennsylvania produced a corpus of recorded Egyptian Arabic telephone calls. *See* Alexandra Canavan, George Zipperlen & David Graff, *CALLHOME Egyptian Arabic Speech*, LINGUISTIC DATA CONSORTIUM (1997),https://catalog.ldc.upenn.edu/LDC97S45.

[160] 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. OF CORPUS LINGUISTICS 159, 164 (2009).

2. Corpus Tools—Frequency, Collocation, and KWIC

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.[161] Whether we regard the ordinary meaning of a given word to be the *possible*, *common*, or the *most common* sense of that word in a given context, linguistic corpora can allow 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."[162] As we have seen, words are often interpreted according to the semantic environment in which they are found. And a collocation program 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.[163]

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. This sort of systematic observation of a word in context can allow us to gain meaningful and quantifiable insight about the range of possible uses of a word and the frequency of its different senses.

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. This is a defining characteristic of corpus linguistics—"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."[164] As professor Larry Solan has noted, "[w]hen 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."[165] In this article

---

[161] TONY MCENERY & ANDREW WILSON, CORPUS LINGUISTICS: AN INTRODUCTION 82 (2d ed. 2001).

[162] *See* SUSAN HUNSTON, CORPORA IN APPLIED LINGUISTICS 78 (2002).

[163] *Id.* at 69.

[164] BIBER & REPPEN, *supra* note __..

[165] *Solan, The New Textualists' New Text*, *supra* note ___, at 2059.

we rely on a pair of broad-based corpora of standard written American English (one contemporary and one historical). But a corpus can be constructed to represent the language use of any speech community or register.[166]

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 (the "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 "3.7 billion words of data from web-based newspapers and magazines from 2010 to the present time."[167] It is a monitor corpus that "grows by about 4-5 million words of data each day (from about 10,000 new articles), or about 130 million words each month."[168] The NOW Corpus downloads content every night from dozens of websites listed on Google News, using an automated software program.[169] These texts are then automatically tagged and lemmatized (adding part-of-speech metadata to each word) and integrated into the existing corpus.[170] Because of this extraordinary rate of growth, the NOW Corpus is currently the largest tagged corpus of English in the world and seems likely to keep that title for the foreseeable future.

---

[166] While corpora vary in size and sophistication, anyone can build a corpus using freely available software like AntConc. *See AntConc Homepage*, LAURENCE ANTHONY, http://www.laurenceanthony.net/software/antconc/ (last visited Mar. 8, 2017).

[167] *See NOW Corpus*, BYU, http://corpus.byu.edu/now/ (last visited Mar. 8, 2017).

[168] *Id.*

[169] *Id.*

[170] *Id.*

"There is no data like more data,"[171] 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, and 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,[172] 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 a corpus 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 of now (and over the past decade). So if we want to evaluate interpretive problems against the backdrop of linguistic norms prevailing at the time of enactment, we will need to turn elsewhere.

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

The COHA is "the largest structured corpus of historical English."[173] 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."[174] 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, non-

[171] Kenneth Church, *Has Computational Linguistics Become More Applied?* in Computational Linguistics and Intelligent Text Processing: 10th INT'L CONF., CICLING 1, 3 (2009) (attributing the statement to Robert Mercer).

[172] WILLIAM LABOV, DIALECT DIVERSITY IN AMERICA: THE POLITICS OF LANGUAGE CHANGE 1–2 (2012).

[173] *Id.*

[174] *Id.*

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.

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. And 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.[175]

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 HLA Hart[176] in his famous debate with Professor Lon Fuller.[177] But seemingly everyone has treated the problem since then.[178] And there is no shortage of extensions of the hypothetical. Hart says that "[p]lainly" the rule "forbids an automobile," but asks "about bicycles, roller skates, [and] toy automobiles."[179] (The airplane example invokes an actual

[175] 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 corpus 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"). The BYU Law School, moreover, is seeking to fill this gap. It is currently working on a Corpus of the Founding Era American English ("COFEA").

[176] Hart, *supra* note ___, at 608.

[177] Lon L. Fuller, *Positivism and Fidelity to Law—A Reply to Professor Hart*, 71 HARV. L. REV. 630, 663 (1958).

[178] *See, e.g.*, SCALIA & GARNER, *supra* note __, at 37–38; Fallon, *supra* note __, at 1260–61; ESKRIDGE, *supra* note __, at 45–46.

[179] Hart, *supra* note __, at 607.

case—*McBoyle v. United States*,[180] in which the United States 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.[181])

The scholars cited throughout this article have offered their own views on the scope of "vehicle." 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 cars, baby carriages, tricycles, or perhaps even bicycles."[182] Professor Richard 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.[183] And Professor William Eskridge disagrees with Scalia's assertion as to bicycles; he says that "bicycles are commonly considered vehicles," a conclusion he claims to confirm using corpus data (more on this below).[184]

Yet no one has attempted to measure our assessment of *vehicle* with any data. We present some relevant data below—data of relevance to 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 *vehicles* appears and the kinds of vehicles that tend to appear in that environment.

We can view the most common contemporary collocates[185] of *vehicle* in the NOW Corpus.[186] In NOW the fifty most common collocates of *vehicle* are as follows:

---

[180] 283 U.S. 25, 27 (1931).

[181] 18 U.S.C. § 408 (1919).

[182] SCALIA & GARNER, *supra* note __, at 37–38.

[183] *See* Fallon, *supra* note __, at 1260 – 61.

[184] ESKRIDGE, *supra* note 4, at 45–46.

[185] The NOW corpus and other BYU corpora are available without a subscription. To access NOW you go to http://corpus.byu.edu/now/. To generate a list of collocates in NOW you take the following steps: (1) Select "Collocates" on the NOW Corpus homepage; (2) Enter "VEHICLE_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 noun forms); (3) Enter an asterisk "*" (a wildcard) in the "Collocates" field; (4) Select "Sections" and select "United States"; (5) Select "Sort/Limit" and set the "Minimum" to "Mutual Info"; and (6) Click "Find collocates."

[186] The following link will reproduce the search above, except that the user would need to repeated step four (4), "Select 'Sections' and select 'United States,'" which

> *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*[187]

Many of the collocates of *vehicle* in the NOW Corpus strongly indicate *automobile* as a likely candidate for *vehicle*'s most common use of vehicle. The NOW Corpus lists a number of automotive collocates like *motor, car, traffic, fuel, driving, gas, battery, batteries, emission, 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 the context almost always indicate an automotive meaning.[188] *Airplane* doesn't appear, though two particular types of aircraft are attested in the collocates—unmanned aerial vehicles (drones) and spacecraft.[189] Similarly, *bicycle* doesn't 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 and can be viewed at the link below.[190]

> *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,*

---

doesn't repopulate automatically. *See NOW Corpus*, BYU, http://corpus.byu.edu/now/?c=now&q=54596680 (last visited Mar. 8, 2017).

[187] The search results are saved at the following link, *see NOW CORPUS*, BYU, http://corpus.byu.edu/now/?c=now&q=52902048 (last visited Dec. 21, 2016).

[188] 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.

[189] 2 OXFORD ENGLISH DICTIONARY at 920.

[190] *See COHA*, BYU, http://corpus.byu.edu/coha/?c=coha&q=52600298 (last visited Dec. 21, 2016).

> *horse-drawn, high-powered, amphibious, administrators, tactical, registration, delivery*

We can see from this data that the meaning of *vehicle* evolved significantly from this period, though the automotive use of *vehicle* still predominated. The decade of the 1950s is remarkable as the first decade in which the spacecraft sense of *vehicle* appears, but also the last decade in which the *horse-drawn* collocate of *vehicles* appears.[191] *Unmanned* no longer appears, but *manned* vehicles appear (*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. And, again, none of the top fifty collocates of *vehicle* include the notion of *airplane* and *bicycles*.

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,[192] and because the COHA only allows us to search in 10-year increments, it may make sense to include data from the decades of 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. 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, sqdriver, passing, moving, various, horse, automobiles, tax, heavy, drawn, carry, roadless, rickety, trucks, communication, approaching, traffic, electric, mental, physical, 3500000, astral, belonging, steam, transportation, commissioner, rear, total, carrying, propulsion, propelled, oncoming, carriages, registration, ego, conceivable, tires, drivers, vehicle, carriers, 45, loaded, halted, manufacturers*

The collocates from this period add a few interesting *vehicles* to our growing list, including *astral vehicle* (a reference to the theosophical notion of a "supersensible substance" that "accompanies the individual through life . . . and survives the individual after death").[193] It should also

---

[191] An additional *vehicle* is added to our collection with *amphibious vehicle*, and *tanks* makes an appearance again. And two collocates from (the abbreviation *b.g.* for background; and *robot*) both are the result of the inclusion of science fiction screenplays in the corpus. In both cases the vehicles in question are spacecraft.

[192] *See* McBoyle v. United States, 283 U.S. 25, 25–26 (1931) (quoting National Motor Vehicle Theft Act , 18 U.S.C. § 408 (1919)).

[193] This notion that also explains the presence of *ego* and *mental* in the collocates of *vehicle*.

be noted that only a few of the collocates in their 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.[194]

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 Key Word In Context 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* raises an important question for our frequency continuum. If we accept that the necessary and sufficient conditions of *vehicle* are "[a]ny means of carriage, conveyance, or transport"[195] or "a means of carrying or transporting something,"[196] 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 Key Word In Context

We can extract concordance data from the NOW Corpus.[197] A NOW search for concordance lines of *vehicle* will yield an output along these lines:

---

[194] *See COHA*, BYU, http://corpus.byu.edu/coha/?c=coha&q=53847214 (last visited Feb. 8, 2017).

[195] 19 OXFORD ENGLISH DICTIONARY 480 (2d ed. 1989).

[196] WEBSTER'S THIRD NEW INTERNATIONAL DICTIONARY (1961).

[197] 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*, 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. So 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 No. 1—KWIC of *Vehicle*

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 will allow us to select anywhere from 100 to 1,000 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 context in the article referenced in a given concordance line.

In order to examine the sense distribution of *vehicle*, we reviewed one hundred (100) randomized concordance lines of *vehicle* in the NOW Corpus. Of those, ninety one (91) 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 development, celebration, connectivity and more*), and there was a reference to the military's efforts to create a flying Humvee / helicopter hybrid.[198]

There were no *airplanes, bicycles, tricycles, skateboards, roller-skates, toy cars* or any of what Hart and others have characterized as penumbral, disputed cases in the NOW Corpus data. To the extent our notion of ordinary meaning has a frequency component, this data suggests that *automobile* is the overwhelmingly *most common* use of the word *vehicle* in the contemporary, 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 period of the 1950s showed a wider range of *vehicles*. Still approximately sixty-five percent

[198] We are not making this up. *See Aerial Reconfigurable Embedded System* WIKIPEDIA,https://en.wikipedia.org/wiki/Aerial_Reconfigurable_Embedded_System (last visited Mar. 5, 2017).

of the instances of vehicles during this timeframe had reference to *automobiles.* Another thirty percent had reference to the space program or missile defense, while the remaining five percent had reference to metaphorical uses of *vehicle* (*e.g.*, a *film as a starring vehicle* for an actor). For the period of 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 5th Avenue was crowded with vehicles, it is not clear from context whether automobiles or carriages or both were intended.) Finally, there were a number of references to theosophy and the notion of an astral vehicle. Still, in the COHA data, there were no references to *airplanes, bicycles, tricycles, skateboards, roller-skates, or toy cars* in the data 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.*[199]

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 (nouns, verbs), but not specific senses. That means that our search for *vehicles in the park* must begin by eliminating the approximately forty (40) percent of the concordance lines that have reference to *vehicles* that are *in park* as opposed to *in reverse, neutral,* or *drive.* There are approximately five percent of the concordance lines that refer to recreational vehicles in recreational vehicle parks. Of the remaining instances of vehicle, more than fifty (50) percent have specific reference 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

[199] (1) Select "Collocates" on the NOW Corpus homepage; (2) Enter "VEHICLE_n" in the "Word/phrase" field; (3) Enter an asterisk "PARK_n" in the "Collocates" field; (4) Select "Sections" and select "United States"; (5) Select "Sort/Limit" and set the "Minimum" to "Mutual Info" and "3"; (6) Click "Find collocates"; and (7) Click "PARK" or "PARKS."

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 near Sunset Park.*

(2) *two males in another vehicle near a park 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.*

Still, 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.*[200]

Professor William 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."[201] 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 and other vehicles.*

(2) *In New Jersey, bicycles are considered vehicles and must follow the same laws as motorists.*

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

[201] Eskridge, *supra* note 4, at 45–46.

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.)*

The conclusion that *bicycle* is included in the ordinary meaning of *vehicle* depends on our view of ordinary meaning. Based on the corpus data reviewed above, *bicycle* is certainly a *possible* sense of *vehicle,* but from the stand point of statistical frequency, it is not a *common* meaning and certainly not the *most common.*

With respect to *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*. *Airplane* is neither the *most common* or even a *common* meaning of vehicle, and based on its absence from any of our corpus data we might well ask if it 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 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.

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 and 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.[202] 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,*[203] *express, burden, permit, thoughtful, load, bags, plane, firearms, virus, tradition, flag, capable, torch, handgun,*

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

[203] Collocates are not lemmatized in the BYU corpora, so the singular and plural form of a given collocate care counted separately. In this case, that means that *firearm* and *firearms* both make the list.

cargo, openly, permits, duties, pipeline, mosquitoes, touchdowns, ships, executions, loads, trucks, felony, tasks, handguns, experiments, knife

These collocates suggests that a number of uses of *carry* don't 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*). And there are a number metaphorical uses of *carry* (*felonies carrying certain penalties, people carrying opinions*). There's *carrying out of attacks* and *executions*, and there is a sporting references to *carry* as well (*carrying so many yards for so many touchdowns).*[204]

Yet "[a]t issue here is not 'carries' at large, but 'carries a firearm.'"[205] 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 doesn't get us much closer to determining which of these senses of *carry* is the most frequent.[206] But as we will see, this search reveals common collocates of *carry* that have the 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 will give us a clearer picture of the use of *carry.*[207] The NOW corpus will give us a randomized sample of concordance lines featuring *carry*.[208] And we can review these concordance lines to determine both the range of possible meanings of *carry* and the

[204] We can see similar results in the COHA, using the same instructions in note __ above, 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 COHA*, BYU, http://corpus.byu.edu/coha/?c=coha&q=54015512 (last visited Feb. 15, 2017). As we have noted, 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.

[205] *Mucarello*, 524 U.S. at 143 (Ginsburg, J., dissenting).

[206] 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 which would serve similar functional roles in a sentence. This will help us locate relevant concordancing data, but does not answer the question of which sense of *carry* is the most common.

[207] One way to examine carry in context is simply to enter a search similar to that in note __ above, but substituting "CARRY_v" for "VEHICLE_n."

[208] As we have already seen, *carry* has a transitive argument structure and in the relevant context of Section 924(c)(1) *carry* has a human subject and a non-human, inanimate, weapon object. 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.

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 Section 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.*[209] 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. 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 Section 924(c)(i) or similar statutes. In order to ensure that we have sufficient data from outside of a legal context, we have also examined instances of *carry* in the context of a number of common synonyms of *firearm* that were listed among the most common collocates of *carry—gun(s), pistol(s), handgun(s),* and *rifles(s).*These synonyms share the same semantic features with *firearm*, but less commonly appear in statutory prohibitions against carrying a firearm.

Our search parameters will eliminate a number of irrelevant uses of *carry*.[210] All that is left is to review the concordance lines and determine in how many instances *carry a firearm* has reference to *carry on one's person* or *carrying in a car*. Here, the physical and spatial context can be helpful, as when the physical location 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, had reference to statutory provisions similar to Section 924(c)(1).

All told, we reviewed two hundred seventy-one (271) concordance lines from the NOW Corpus in which *carry* co-occurred with *firearm(s), gun(s), pistol(s), handgun(s),* and *rifles(s).* Of these instances of *carry* we found that one hundred and four (104) instances of that indicated a sense

[209] Such a search can be executed as follows: (1) Select "Collocates" on the NOW Corpus homepage; (2) Enter "CARRY_n" 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 "Mutual Info" and "3"; and (6) Click "Find collocates." *See NOW Corpus*, BYU, http://corpus.byu.edu/now/?c=now&q=54477447 (last visited Feb. 15, 2017).

[210] The metaphorical sense (*carry a tune*) or senses where the subject / agent is an inanimate objection (*the ship carries cargo*) or the sporting sense (*carried the football nine yards*).

of *carry a firearm on one's person*, while only five (5) instances suggested a *carry a firearm in a car* sense. The remaining senses were either unclear or senses of *carry* or unrelated to the question at hand. As would be expected, much less data was available for *carry* in the COHA. We found twenty-eight (28) concordance lines from the COHA, in which *carry* co-occurred with *firearm(s), gun(s), pistol(s), handgun(s),* and *rifles(s).* Of these instances of *carry* we found that eighteen (18) instances of *carry on one's person*, two (2) instances of *carry in a car*. The remaining instances were either unclear or reflected a different sense of *carry.*

To the extent that we view the question of ordinary meaning has involving statistical frequency, the analysis above tells us that *carry on one's person* is overwhelmingly the *most common* use while the *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*. And 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:[211]

> *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*

A number of the collocates tend to support the *Taniguichi* majority's position that *interpreter* most commonly has reference to an *interpreter of spoken language*. These include *speaking, spoke, listen*. A number of the collocates have reference to battlefield *interpreters (Afghan, Iraqi),* 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 a writing and

[211] *See NOW Corpus*, BYU, http://corpus.byu.edu/now/?c=now&q=54018483 (last visited Feb. 15, 2017).

publishing context, including *bible, writer, poet, editor, literary, publisher, journalist, Borders,* and even *Wycliffe.*[212] In addition, the collocates from the decade of the 1970s, when the Court Interpreters Act was passed, suggest a similar conclusion.[213]

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.[214] A number of instances, *interpreter* had reference to artist expression or the interpretation of works of art (*a noted interpreter of modern music*). Another common sense has reference to the interpretation of documents the 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. And there was one instance of an interpreter translating a foreign language document into spoken English, in (3) 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.*

---

[212] *See NOW Corpus*, BYU, http://corpus.byu.edu/now/?c=now&q=54609539 (last visited Mar. 9, 2017).

[213] *Interpreter* has very few frequent collocates from during the decade the 1970. 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 reveal their origin in the extremely common phrase *speaking through an interpreter*, or related phrases. *See COHA*, BYU, http://corpus.byu.edu/coha/?c=coha&q=54495283 (last visited Mar. 5, 2017).

[214] 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. 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 being whether the language at issue is spoken or written.

Absent from all of these concordance lines was a single instance of anyone referred 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 we would look to the data 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 of 1910 through 1919, the period during which the relevant statute was enacted.

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* has reference 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, persons, thoughts, fugitives, germs, mutation, tumors, aliens, moon, bias, genes, gene, hole, diversity, grudges, resentments*[215]

Such uses of *harbor* do not match the semantic features in the relevant statute. We are looking for objects of harbor that are human, animate,

---

[215] *See NOW Corpus*, BYU, http://corpus.byu.edu/now/?c=now&q=54496834 (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 through 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 (*alien, refugee*), no other relevant collocate appears more than once in the COHA. *See COHA*, BYU, http://corpus.byu.edu/coha/?c=coha&q=54496926 (last visited Mar. 5, 2017).

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, refugees.*

b. *Harbor* as a KWIC

In the NOW Corpus, we examined one hundred and forty concordance lines in which *harbor* occurred in the same environment as *fugitives, terrorists, criminals, aliens, refugees*. Of these, twenty three (23) instances of *harbor* had explicit reference to *concealment* while thirty two (32) had reference to *shelter*. In additional eighty three (83) instances, the distinction could not be determined by context. There were also three (3) instances of unrelated senses of *harbor.* In the COHA, there were only three (3) clear cut cases of the *shelter* sense. The remaining five (5) instances of *harbor* could not be determined by context.

This data raises more questions than 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. Conclusions

Above 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 to 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).

And 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) we begin with a more careful synthesis of the theory of ordinary meaning introduced above. And we then offer some conclusions about what the corpus data tell us about the ordinary meaning of *vehicle*, *carry a firearm*, *interpreter*, and *harbor*.

1. Corpus Data and Meaning

Corpus analysis may be addressed to the range of senses of ordinary meaning identified above—to semantic meaning, to intended meaning, and to public meaning. 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. And 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.

a. Semantic meaning

Frequency data from a corpus may give us useful evidence of the ordinariness or plainness of a given semantic sense of a particular term. In constructing a corpus search we can control for relevant components of semantic meaning—of elements of syntax like argument structure and components of semantics like number, animacy, humanness, and concreteness. We can do that, as above, by constructing a controlled search for uses of the verb *carry* (a) as a transitive verb, (b) by a human subject, (c) acting on an inanimate, concrete object, and (d) that is a weapon or form of firearm. And the results of our search will show us that *carry* in this context almost always (90% of the time in the above data set) has reference to the *bear on your person* sense of the verb.

That provides an empirical ground for the conclusion that the ordinary semantic meaning of *carry* in this setting aligns with the view espoused in Justice Ginsburg's dissent. If a database of naturally occurring language tells us that our references to carrying a firearm in this linguistic setting are most always references to *bearing* on your person and very rarely to *transporting* in a vehicle, then that is an indication that the ordinary semantic sense of *carry* in this setting is the former sense.

Certainly this is a better way to solve the *Muscarello* puzzle than by resort to sense ranking or etymology. And if what we are looking for is semantic meaning—as the law sometimes says we are—then this may be an acceptable answer to the question.

That said, we see some possible limitations on the strength of the inference to be drawn from this sort of data.

Sense division. One possible limitation stems from the vagaries of word sense division. Sense division is subjective.[216] Linguists, as noted above, have no agreed-upon formula for distinguishing senses of a word.[217] They concede that distinctions among senses may be "more of a

[216] Nikola Dobric, *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").

[217] 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

descriptive device rather than a claim about psycholinguistic reality."[218]

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 governing a beauty salon has a different meaning from *nail* in a municipal building code.'"[219] 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 by reference to a fingernail or toenail. And 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 related. Both get at the idea of *transport*; the difference concerns the mechanism—on one's person or in a vehicle. And it accordingly seems hard to know whether this difference reflects "psycholinguistic reality." The *bear* personally sense seems to be the notion of carry that we speak of most 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 is also encompassed within the notion of carrying a firearm. Psycholinguistic reality could be such that most ordinary people first think of the *bear* personally sense but on reflection agree that the *transport* sense is included.

A variation on this problem has been highlighted by Professors Larry Solan and Tammy Gales. They have observed that corpus data along these lines may reflect only the fact that a given sense of a certain term is *a more factually common iteration* of that term in the real world.[220] And

---

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; some linguists speak of senses and sub-senses, Glynn, *supra* note __, at 17, others 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 (2013), and others of hyponomy and hypernomy in polysemy, *see* Glynn, *supra* note __, at 10.

[218] Stefan Th. Gries, *Polysemy*, *in* HANDBOOK OF COGNITIVE LINGUISTICS 482–83 (Ewa Dabrowska & Dagmar S. Divjak eds., 2015). 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." Dobric, *supra* note __, at 77.

[219] SCALIA & GARNER, *supra* note __, at 20.

[220] Lawrence M. Solan & Tammy Gales, Corpus Linguistics as a Tool in Legal Interpretation, Paper presented at Brigham Young University Law School's Law and Corpus Linguistic Conference (Feb. 3, 2017).

if that is true, there may be reason to doubt the probity of the data in establishing the semantic meaning[221] 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 linguist analysis (even for related senses of a statutory term). Instead 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 don't 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. (We could conceivably test that through a psycholinguistic experiment; more on that later.)

That 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."[222] And we think the answers to these questions are dictated in part by the rationale that drives 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. Meaning arrived at upon reflection is meaning that could catch the public unawares. So there may be reason to stick to top-of-mind meaning and not encompass reflective meaning if we are concerned about fair notice.

The legal authority rationale could at least arguably lead us in a different direction. An ordinary legislator could expect to have his

---

[221] 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.

[222] Baude & Sachs, *supra* note__, at 1089–90.

legislative enactments sweep more broadly than the first sense to come to mind. (To answer that question we would need to proceed to a sense of intended meaning; more on that below.) And where that holds, it may make sense to treat the broader sense of *carry* that occurs to us on reflection to count as ordinary semantic meaning.

Inconclusive data. 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 are less clear? What if the data suggest 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. And 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 a raw question of ordinariness. But sometimes the question is one of plainness—of whether the semantic meaning of a word or phrase is sufficiently clear to close the door on legislative history (or on a substantive canon or agency deference).[223] And where that is the question inconclusive data may be quite conclusive—it may tell us that there is ambiguity sufficient to proceed past the threshold "standard picture." So corpus data can bring rigor to this range of questions too; instead of guessing about plainness we can summon some data.

In some cases the data may just be too mixed to yield any helpful answers. Even then that does not require us to abandon the standard picture. We could look to another empirical tool for assessing semantic meaning (see the discussion below on psycholinguistic experiments). Barring that kind of help, we can fall back on a principle of interpretation 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 (more on that below too). But we see no reason to fall back too quickly. Again, the law's commitment to the standard picture is for good reason. We think the courts should try their best to find real answers to linguistic questions before falling back on fake ones.

b. Intended Meaning

Semantic meaning is a good starting point for getting at intended meaning. But linguistic theory tells us that a lawmaker's intent can properly be inferred only by also considering pragmatic context. So if we are interested in vindicating not just word meaning but perceived

[223] *See supra* at __.

intended meaning of the lawmaker we must proceed beyond semantics. We will have to consider pragmatics.

Classic illustrations are 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."[224] And Fallon understandably asserts that the lawmaker adopting the "no vehicles in the park" rule would "reasonably" be understood to intend to allow the "gatekeeper" at the park to allow in an ambulance in the event of an emergency.[225] We can think of these conclusions as departures from the standard picture—as the vindication of consequentialist concerns overriding the communicative content of the law.[226]

But we think they can easily fit into the standard picture. The pure semantic meaning of "Keep off the grass" would apply to the groundskeeper, just as the "no vehicles" prohibition would encompass an ambulance (which on any semantic understanding would count as a vehicle). But the pragmatic context of these rules implies clear limitations on this semantic meaning that leaves room for the groundskeeper and the ambulance.

The author of the "Keep off the grass" rule in Posner's example is the property owner. And that owner undoubtedly also hired the groundskeeper. That pragmatic context gives us a clear indication of intended meaning: If the property owner hired the groundskeeper to cut the grass he obviously intended for the groundskeeper to cut the grass (and not keep off it).

The ambulance hypothetical is a little harder. But here again we don't think we need pure consequentialism to justify the exception to the semantic rule. Again the rulemaker is the property owner. And any property owner can be expected quite clearly to intend an exception for ambulances dealing with emergencies. Whether out of altruism or concern for premises liability, the park owner in Fallon's hypothetical can be presumed to intend to allow emergency vehicles in to rescue fallen park-goers.

Fallon says that the rulemaker likely was "thinking about automobiles driven for recreational purposes, not ambulances."[227] And he accordingly says that ambulances responding to emergencies are not a "specifically expected application."[228] Fair enough. But that doesn't mean that "moral

---

[224] RICHARD A. POSNER, REFLECTIONS ON JUDGING 180 (2013)..

[225] Fallon, *supra* note __, at 1260–61.

[226] *See* POSNER, HOW JUDGES THINK, *supra* note __, at 40 (describing "pragmatism" as a process of "basing judgments (legal or otherwise) on consequences, rather than on dedication from premises in the manner of syllogism").

[227] Fallon, *supra* note __, at 1261.

[228] *Id*. at 1262.

reasonableness" is the only way to make room for the ambulance exception, as Fallon suggests.[229] In ordinary communication we often make inferences about a speaker's intent. And where (as here) such intent is clear, we naturally credit it as a matter of normal conversational meaning.

So in our view this is still ordinary meaning. (It is just inferred intended meaning that takes pragmatic context into account.) We are still on the standard picture. The question for present purposes is whether we can measure this kind of meaning.

In our view the answer is that it may be difficult but not impossible. If we had an unlimited corpus with a sufficiently targeted register (genre), we could find ways to glean data on the sort of pragmatic, intended meaning described above. If we had a database of park owners talking about groundskeepers we would undoubtedly find examples of statements indicating their intent to allow them to tread on the grass. Presumably we could find something similar for ambulances. If park owners talk routinely about ambulances responding to emergencies in a manner indicating approval (and never disapproval), we could infer park owner intent to allow such access. Perhaps the relevant corpus would have to be even more targeted to be probative. To find an intended exception to the "Keep of the grass" or the "no vehicles" rule we would need to be looking at park owners who have these rules in place. But the point is that corpus analysis is at least theoretically capable of getting at pragmatic, intended meaning.

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 get at the pragmatic components of the intended meaning of a sentencing enhancement for carrying a firearm in connection with a drug crime. Theoretically, the right corpus would reflect dialogue among the 535 members of the Congress that voted on the sentencing enhancement in 18 U.S.C. § 924(c)(1). If we had such a corpus, and if it recorded extensive discussion among the members of Congress 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, the *might* qualifier is necessary. The limitation here is the same one presented above as to semantic meaning—as to whether a predominance 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

[229] *Id.*

sense that might occur to a lawmaker on reflection should count as ordinary.[230]

How might a judge answer this question? For some such questions the question may be framed within the standard picture. On questions of likely intended meaning of rules like the "Keep of 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.[231] And if so, it seems safe to conclude that the intended communicative content of these rules would sustain the noted 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. And 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 an available gun may always be used in a harmful way if it is available,[232] or (b) only if the gun is being carried on the defendant's person, since that kind of availability is even riskier.[233]

---

[230] *See supra* Part IV.B.1.a.

[231] *See* 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; noting that "[l]aw is a vector rather than an arrow," "[e]specially when you see the hand of interest groups"); MANNING & STEPHENSON, *supra* note __, 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 __, at 39 ("Not only is legal drafting sometimes imperfect, but often the imperfection is the consequence of a compromise that is not the function of the courts to upset—or to make impossible for the future by disregarding the words adopted.").

[232] Muscarello v. United States, 524 U.S. 125 at 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.").

[233] *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.").

So *Muscarello* is an example of a case in which pragmatic judgments about likely legislative intention are likely to be difficult. If we lack a reliable basis to decide the case on either semantic or likely intended meaning, 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 doesn't pretend to be based on communicative content (because we haven't been able to find it). 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 kicks in in the face of genuine ambiguity about statutory meaning. Such a move, moreover, would be more open and transparent than a false assertion about communicative content. And for that reason we would favor it even though it might not obviously vindicate the principles motivating the law's threshold devotion to ordinary meaning.

c. Public meaning

Much of what we said above about intended meaning applies also to public meaning. Ideally the two senses of meaning overlap completely. They overlap, for example, if the groundskeeper or ambulance driver accurately anticipates the intent of the park owner in the above hypotheticals, or if would-be drug dealers perfectly assess the sort of firearm carrying that Congress has in mind in enacting 18 U.S.C. § 924(c)(1). If that happens then "what the speaker means and what the hearers take the speakers to mean" are the same thing.[234] And in that event the pragmatic meaning question implicates many of the same issues highlighted above.

Yet the overlap will not always be complete. 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.

Data from a general, balanced corpus could tell us something about the way 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

[234] Soames, *supra* note __, at 598; *see also* Doerfler, *supra* note __, at 983–84 (getting at a similar conclusion through his premise of a "conversation" model of meaning and "fictionalized" legislative intent).

"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 it would to a mere statement of description. Thus, we may form one understanding when listening to a 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 seems impossible to suss out with corpus analysis. To get at this kind of nuance we may only be left with psycholinguistic study. (More on that below.)

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. Again that is a problem for legal theory—and essentially for 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 given 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. Corpus Data and Our Test Cases

Now we can assess the above data in light of the principles of semantic, intended, and public meaning developed above. 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.

On each of the test cases we start with a premise of ordinary meaning that is susceptible of 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—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. Where the data identify a sense that is most frequent in this setting we conclude that there is at least a prima facie basis for concluding that that sense is the ordinary one that would protect against unfair notice and vindicate the intent of the lawmaker.

Yet we also recognize some possible grounds for questioning the 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 established we think the conclusion that the most frequent sense of a term is the ordinary one may be in doubt. And in that event we identify other considerations that may be brought to bear, such as alternative means of empirical analysis of the standard picture (through a psycholinguistic experiment) or, ultimately, considerations that go to legal content rather than communicative content.

a. Vehicles

Our data yield helpful grounds for assessing the meaning of *vehicle*. We can use corpus analysis to make conclusions about whether the ordinary sense of this term extends to the various examples discussed in the literature—automobiles, bicycles, toy cars, airplanes, ambulances, and golf carts.

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. And we can make a strong case for crediting that sense as the ordinary one, in that it will best avoid unfair surprise and vindicate the presumed intent of the lawmaker.

Our review of the corpus data also lead us to conclude that a number of other potential candidates for the ordinary meaning of *vehicle* are out of the running. *Airplane* (and its synonym *plane*) are unattested in the collocation data, and a review of the concordance data didn't turn up any instances in which the word *vehicle* was used to describe what we might think of as a conventional *airplane*. The corpus data did show a few well-attested uses of *vehicle* in reference to *unmanned aerial drones.* This was not the *most common* or even a comparatively *common* use of *vehicle*, but the data demonstrated that it was an attested and *possible* use of *vehicle*. We can accordingly conclude that the ordinary meaning of *vehicle* is *automobile*—at least to the extent that we accept the *most common* use of a word and exclude other *possible* uses.

A similar story holds for *bicycle. Bicycle* (and its synonym *bike*) are unattested in the collocation data and the *bicycle* use of *vehicle* didn't show up in our randomized collection of concordance data. Yet a specific

search for *bicycle* (and *bike*) turned up a number of instances in which *vehicle* appears to be used to describe a *bicycle*. Here again, we can conclude that the ordinary meaning of *vehicle* is *automobile* if we accept the most *common use* of the word as the ordinary meaning, and exclude the rare, but *possible* use of *bicycle*.

There may be very good reasons to exclude rare, but possible uses of a word from coverage under the ordinary meaning rule. A decision to extend the law to these terms could upset reliance interests of those who—according to the data—are likely to think of automobile 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.

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. In our view this weighs against treating these examples as falling under the ordinary sense of vehicle. But again 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. And ambulance also easily fits within the ordinary (automobile) sense of vehicle. (It shares a variety of features with other automobiles and appears in many similar contexts. It is simply a less common example.) 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*. It has four wheels, a motor, carries passengers and baggage, and may even be subject to licensure by a state department of motor vehicles.[235] On the other hand, licensed or not, 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 in the *automotive* use of *vehicle*) is accordingly a difficult one. It turns on the viability of the sense divisions at work—on whether the golf cart is an unusual example or a distinct psycholinguistic 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.

We see two ways of answering that question. One would be through further corpus analysis. With sufficient corpus data we could assemble a

[235] *See* Nevada Department of Motor Vehicles, Low Speed Vehicles & Golf Cars, http://www.dmvnv.com/lowspeed.htm (last visited March 23, 2017).

list of criteria for things we speak of as an automobile. And then we could ask whether a golf cart has those criteria.[236]

An alternative would be to construct a psycholinguistic study. Psycholinguists study the human acquisition, use, and comprehension of language.[237] They may do so by means of a human subject survey or experiment. One possible experiment would be aimed at identifying the sufficient conditions for the automobile sense of vehicle—to test whether the golf cart is an unusual example of the automobile sense or a distinct psycholinguistic construct.[238]

We may ultimately find no determinate answer in the above kind of analysis for the golf cart question—or for related questions about go-carts or four-wheelers. But psycholinguistics could yet yield empirical data of some relevance. An alternative approach would be to construct a survey aimed at assessing not just the first sort of *vehicle* that comes to mind but also the range of meanings encompassed within a prohibition on vehicles in the park. Survey data could give us quantitative information about these notions of ordinary meaning—and whether, for example, a golf cart would be understood to come within the ordinary sense of *vehicle*.

Yet there are barriers to psycholinguistic survey data, and reasons to doubt the viability of this sort of analysis. One problem is timing-based—to the "meaning as of when" question. If we want to find meaning as of the date of a statute's enactment we will never be able to measure it psycholinguistically. And the older the statute the more concerning that problem might seem.

Another problem concerns the artificiality of the psycholinguistic study and its susceptibility to context effects and response bias.[239]

---

[236] 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.

[237] JOSEPH F. KESS, CURRENT ISSUES IN LINGUISTIC THEORY: PSYCHOLINGUISTICS 6 (1992) (describing the "domain of psycholinguistics" as "defined by the activities of comprehension, production, and acquisition").

[238] *See* Eleanor Rosch, *Cognitive Representation of Semantic Categories*, 104 J. EXPERIMENTAL PSYCHOL. 192, 193 (1975) (explaining the basis for this kind of study—that 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 of prototype (the clearest cases, best examples) of the category").

[239] There are very few ways to (ethically) conduct a linguistic experiment without letting the test subjects know about it. And once test subjects know they are being evaluated there is no way to prevent them from self-correcting and self-evaluating while they are participating in the experiment. People naturally, and unconsciously, attempt to give the answer they think the experiment administrator wants or the answer that they perceive to be the most "socially desirable." They try to figure out the "right" answer. This presents a problem when using psycholinguistic experiments in evaluating ordinary meaning.

Psycholinguists do their best to test real human perception that is unbiased by flaws in the study design. But there will always be flaws in study design. As to *vehicle*, for example, we would need to figure out what kind of question to ask study recipients. Do we ask an open-ended question or provide a list of possibilities to choose from? If we ask an open-ended question we may end up with complicated coding problems that could make the analysis less determinate. But if we provide a list the study design may bias the answers.[240]

These limitations 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. 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 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 the NOW Corpus we found 109 concordance lines involving carry and firearm (or a synonym) in which we could discern one of the two senses at issue. Most all of those (104) involved the bearing on your person sense of carry. The COHA concordance lines were more limited but supported the same conclusion: in 18 out of 20 lines in which we could discern the relevant sense of carry the verb was used in the packing on your person sense of the verb.

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. And that provides a *prima facie* basis for concluding that the ordinary communicative content of the mandatory minimum sentencing provision in 18 U.S.C. §924(c)(1) is limited to the personally *bearing* notion of *carry*.

Solan and Gales might observe that the data may merely be an artifact of the greater commonality of personally *bearing* carrying in the real world. 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

[240] This is hard. Perhaps that's why Larry Alexander spoke (hypothetically) of a telepathic basis for finding meaning. *See* Alexander, *supra* note __, at 142. That would be easier. And presumably this is why Fallon and Posner, and to a lesser extent Baude and Sachs, choose to fall back on a non-standard picture—on answers to questions other than the ordinary communicative content of the law. We concede the difficulty. But again suggest that we should try harder to conduct this threshold inquiry better before we fall back to other premises.

about personally *bearing* it, then the first sense of carrying to come to mind is likely to be that sense. And extending the statute to the *transporting* in a car sense may jeopardize significant reliance interests.

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

Perhaps we won't 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 by the competing opinions in *Muscarello*, and a sufficient basis for turning to other means of assessment.

One such basis could be to make 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 (as in one rooted in a dictionary or etymology).

c. Interpreter

The data seem to provide support for Justice Alito's majority view in *Taniguchi*. Of the 188 concordance lines we reviewed from the NOW corpus, we found not a single instance of interpreter in the context of text-to-text written translation. That seems strongly to indicate, at least, that this is the kind of interpreter that first comes to mind when we use this term.

That leaves the same question highlighted above in the other examples—whether the *written translator* sense is psycholinguistically separate from the *oral translator* notion.[241] Here we see reason to suspect that these are just alternative examples of a single psycholinguistic 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. And again that question could be tested empirically through a psycholinguistic study.

[241] *See Taniguchi*, 132 S.Ct. at 2010 (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").

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 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 seems inconclusive.[242] For a number of concordance lines, it simply was not possible to determine from context whether a *conceal* sense or only the *shelter* sense was applicable. Yet we found a number of instances of both the *conceal* sense (23) and the *shelter* sense (32) of this term.

That suggests that both senses are common and attested. To the extent we regard the ordinary meaning as the most common sense of a word, however, the data indicates 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 the sense 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 simply conclude that we cannot determine the ordinary communicative content of the law—and thus that we need a "fake" answer, or an answer rooted in legal content. One such answer may be found in the rule of lenity. Yet even here the data will have proven useful. We have not fallen back on the law of interpretation (legal content) just because the ordinary meaning question seemed difficult, or that our intuition told us there was an ambiguity. We have identified data to support our determination of ambiguity. The standard picture here yields to the law of interpretation, but only after the necessary work has been done.

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

[242] *See infra* CONCLUSION.

interpretation. In a few recent cases judges have proffered corpus linguistic analysis in support of their assessment of the ordinary meaning of statutory terms.[243] Some of these attempts have prompted doubt and criticism from fellow judges.[244] And even the judges who have advocated for this approach (present company included) have acknowledged cause for concern and care in this endeavor.[245]

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

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—to expert witness reports or testimony. A majority of the Utah Supreme Court has so concluded in a couple of recent cases.[246]

The "proficiency" critique has some bite to it. For reasons noted above we must 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. And a judge who proceeds willy-nilly may, either consciously or unwittingly, be proffering data that has only the appearance of careful empiricism.[247] For

---

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

[244] *See, e.g.*, Rasabout, 2015 UT 72 (majority opinion); Baby E.Z., 2011 UT 38 (majority opinion).

[245] *See, e.g.*, Rasabout, 2015 UT 72 at ¶¶ 94–101 (Lee, J., concurring in part and concurring in the judgment)

[246] Rasabout, 2015 UT 72, at ¶ 18 (majority opinion) ("Linguistics is a scientific field of study that uses empirical research to draw findings. And just as with other fields of scientific study, simply trying harder will not lead us to a better answer. 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, 179 P3d. 754, 755–56 (Utah 2008)); Baby E.Z., 2011 UT 38, at ¶ 19 n.2 ("Unless this linguistic 'context' is placed in its proper context, it is of little analytical or persuasive value.").

[247] *See* 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/ ("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."); Michael Stubbs, *Corpus Semantics* 107, *in* ROUTLEDGE HANDBOOK OF SEMANTICS (Nick Riemer ed., 2015)

these and other reasons we whole-heartedly 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.[248]

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 inevitable fact of the matter is that judges and lawyers *are linguists*. We may not be trained in linguistic methodology. But the enterprise of our work puts us consistently and inevitably in the business of resolving ambiguities in legal language.

Judges are experts, in other words, in interpreting the law.[249] 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 the aid of—instead of in open ignorance of our rebellion to—modern tools developed to facilitate that analysis."[250]

Judges are likewise not historians. And it may rightly be said that many lawyers and judges are even "bad historians" with a tendency to "make up an imaginary history and use curiously unhistorical methods."[251] "Yet judges of all stripes engage in historical analysis, particularly in their interpretation of the constitution."[252] "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."[253]

(noting that "a constant background question is whether a corpus can ever, strictly speaking, provide semantic data, since intuition is always required to interpret data," but concluding that while "[s]emantic analysis can never be entirely objective . . . corpora allow us to study language 'with a degree of objectivity [. . .] where before we could only speculate'"); *id.* at 110 (asserting that while "software can automatically extract data from large corpora, . . . intuition is necessary to interpret these data"); *Rasabout*, 2015 UT 72, ¶ 21 (noting that a potentially significant portion of corpus data "require[s] an interpretive assumption" or retains some level of ambiguity).

[248] *Rasabout*, 2015 UT 72 at ¶ 97 (Lee, 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").

[249] *Id.* at ¶¶ 107–08 (explaining that although they do not have "the kind of training possessed by 'linguistics experts' . . . judges *are* experts on one thing—interpreting the law").

[250] *Id.* at ¶ 111.

[251] MAX RADIN, LAW AS LOGIC AND EXPERIENCE 138 (1940).

[252] *Rasabout*, 2015 UT 72 at ¶ 109 (Lee, J., concurring in part and concurring in the judgment).

[253] *Id.*

The same goes for linguistic analysis. "We could continue to judge the ordinary meaning of words based on intuition, aided by the dictionary."[254] "But those tools are problematic, for reasons noted above." "So it is our *current* methodology and tools that involve bad linguistics produced by unscientific methods." For that reason, "[i]f the concern is reliability, the proper response is to embrace—and not abandon—corpus-based analysis."[255]

The potential for subjectivity and arbitrariness is not heighted but reduced by the use of corpus linguistics.[256] 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, moreover, is neither transparent, nor replicable. Nothing is statistically worse than one data point—especially a biased one. The potential for motivated reasoning is evident. A common critique of the use of legislative history comes to mind. The judge who considers only the linguistic evidence that she can summon to mind may be looking over a crowd of language to pick out her friends.[257]

Corpus linguistics, by contrast, facilitates transparency and scrutiny.[258] It is "an empirical check on our (imperfect) linguistic intuition."[259] And it is not, ultimately, a terribly complex or difficult endeavor. "Corpus analysis is like math"—everyone can do it as some basic level; at more advanced levels it becomes too complicated for all but the experts.[260] We're advocating rudimentary linguistic analysis that most anyone can

---

[254] *Id.* at ¶ 112.

[255] *Id.*

[256] As one of us has noted, while a judge may "go looking for supporting evidence in a corpus" it is possible that "after reviewing hundreds of concordance lines, a salient meaning contrary to the judge's initial conclusion becomes harder to ignore." 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 (2012).

[257] *See* Patricia M. Wald, *Some Observations on the Use of Legislative History in the 1981 Supreme Court Term*, 68 IOWA L. REV. 195, 214 (1983) ("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.'"). A parallel problem appears in cases in which judges summon examples of word usage in literary works. Whitfield v. United States, 135 S. Ct. 785 (2015) (interpreting the ordinary meaning of "to accompany" using a host of sources, including quotes from a Jane Austen as well as a Charles Dickens novel). That kind of data cherry-picking if fraught with risk of hindsight bias or motivated reasoning.

[258] *Zimmer*, *supra* note 73 ("at 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"); Mouristen, *supra* note __, at 203 ("corpus analysis brings these subconscious assumptions about language and meaning out in the open").

[259] *Rasabout*, 2015 UT 72 at ¶ 21 (Lee, J., concurring in part and concurring in the judgment).

[260] *Id.* at ¶ 115.

do.[261] We "just think we should be using a calculator instead of doing it in our heads."[262]

The path forward is for judges and lawyers to identify the sort of 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[263] will develop standards and best practices for this field. But 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 Bryan 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."[264]

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 noticed."[265] 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.[266]

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 is in Judge Posner's opinion in *Mitchell v. JCG Industries*.[267] 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. And Judge Posner's opinion included

---

[261] While the COCA 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 sound.

[262] *Rasabout*, 2015 UT 72at ¶ 115.

[263] Perhaps it is a bit ambitious—or self-aggrandizing—to call this a "movement." But we are convinced that the future of the judicial analysis of ordinary meaning must include a consideration of this linguistic tool. It is a bit embarrassing, in fact, that the law has so much to do with the construction of language but we do not have a field called law and linguistics. That will change in time. And we like to think of it as a movement.

[264] SCALIA & GARNER, *supra* note ___, at 400 (quoting KENT GREENAWALT, LEGAL INTERPRETATION: PERSPECTIVES FROM OTHER DISCIPLINES AND PRIVATE TEXTS 168 (2010)).

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

[266] *Rasabout*, 2015 UT 72 at ¶ 17 (majority opinion).

[267] 754 F.3d 837 (7th Cir. 2014).

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." "Their endeavors were videotaped." 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." Posner referred to the results of this experiment in support of "the commonsense intuition that donning and doffing a few simple pieces of clothing and equipment do not eat up half the lunch break."[268]

Judge Diane 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. Wood complained that the results of 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.[269]

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. But the analogy is inapt. A judge who considers corpus data in assessing the ordinary meaning of a statute is not investigating the *adjudicative facts* of a case; he is considering facts of relevance to the proper interpretation of the law. These are known as *legislative facts*. And the investigation of those facts is the inevitable—and quite proper—domain of the judge's "sua sponte" analysis. The governing rules of judicial ethics expressly carve this out; they allow the courts to consider "facts that may properly be judicially noticed."[270]

The point is supported by the law of evidence. Governing rules of evidence typically state that limitations on the judge's judicial notice power is addressed to "an adjudicative fact only, not a legislative fact."[271] 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."[272] Such facts "are not appropriate for a

---

[268] *Id.* at 842.

[269] *Id.* at 849 (Wood, C.J., dissenting).

[270] MODEL CODE OF JUDICIAL CONDUCT r. 2.9(C) (AM. BAR ASS'N 2011); UTAH R. J. ADMIN. CODE r. 2.9(C) (2016).

[271] 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).

[272] UTAH R. EVID. 201(a) advisory committee's 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.").

rule of evidence." They are "best left to the law-making considerations by appellate and trial courts."[273] 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. And for that reason there is no ethical or evidentiary prohibition on "sua sponte" corpus analysis by a judge.[274]

"A contrary conclusion would call into question a wide range of" inquiries routinely conducted by our courts, including the use of dictionaries.[275] "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 constitution, or from considering social science literature in resolving a difficult question under the common law."[276] And "[l]inguistic analysis is no different." "[T]o 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."[277]

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.[278] No one bats an eye when judges do their own

---

[273] UTAH R. EVID. 201(a) advisory committee's note.

[274] Another way to frame this is the distinction between adjudicative and legislative facts. Corpus-based analysis to determine the meaning of legal language falls under the latter, and is appropriate and expected for judicial inquiry. *See, e.g.*, Bulova Watch Co. v. K. Hattori & Co., 508 F. Supp. 1322, 1328 (E.D. N.Y. 1981) (concluding that "whether we explore the economic, political, or social settings to which the law must be applied explicitly, or suppress our assumptions by failing to take note of them, we cannot apply the law in a way that has any hope of making sense unless we attempt to visualize the actual world with which it interacts," and 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).

[275] *Rasabout*, 2015 UT 72 at ¶ 106 (Lee, J., concurring in part and concurring in the judgment).

[276] *Id.*

[277] *Id.*

[278] 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. ex rel. Jeffs v. West*, 2012 UT 11, 275 P.3d 228; and the question of the appropriate age at which a child may appropriately 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. No one bats an eye when judges consult additional

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 will make statutory cases more costly and time-consuming.[279]

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. And we should not do that without a good reason.

Yet we find this objection to corpus analysis a hollow one 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 *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 antecedent."[280] 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."[281] "It comes into play only if we find that the legislature is not

---

materials—such as social science studies or materials—of relevance to the policy determination implicated in these cases. That is because the inquiry involved in these cases concerns "legislative facts."

[279] *Rasabout*, 2015 UT 72 at ¶ 19 (majority opinion) (declaring that the use of corpus linguistics "would require ad hoc linguistics research that could only be reliably conducted by dueling linguistics experts," and that "[i]mposing such a significant financial burden on so many of the litigants coming through the doors of our courts would be tantamount to locking those doors for all but the most affluent").

[280] *See* Lockhart v. United States, 136 S. Ct. 958 (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 __, at107–08 ("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*.").

[281] *Rasabout*, 2015 UT 72 at ¶ 118 (Lee, J., concurring in part and concurring in the judgment).

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."[282] This yields a limited but important domain for corpus linguistics. We turn to an empirical analysis of frequency only in cases in which we have "no better way" of resolving a contest between probabilities of meaning.[283] That is a relatively rare case.[284]

Second, corpus-based analysis won't always require an expert. This "isn't rocket science."[285] 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's 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.[286] The rising generation of millennials is particularly suited to the task. They have never known life without a computer, and are constantly embracing new applications and tools for computer analysis. In time we will see competing corpus presentations as a matter

---

[282] *Id.*

[283] *Id.*

[284] *See id.* at ¶ 118 (asserting that in "five years" on the Utah Supreme Court, he 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").

[285] *Id.* at ¶ 117.

[286] 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.*, Barbara Bintliff, *From Creativity to Computerese: Thinking Like a Laywer 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, *Technocratism 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); Robert C. Berring, *Legal Research in 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* move into legal information"); Scott P. Stolley, *Shortcomings of Technology: The Corruption of Legal Research*, ABA'S APP. ADVOC. COMMITTEE NEWSL. 39 (Apr. 2004) (viewing the likes of Lexis-Nexis 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.

of course in adversary briefing.[287] 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, say, as 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.

CONCLUSION

Some points of analysis outlined herein are necessarily tentative. That seems inevitable in the course of breaking new ground on refined theories and new tools for assessing the ordinary communicative content of the language of the law. 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 are 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 identification of suitable coding methods. Scholars have begun to explore these and other related questions.[288] 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 now stands at its infancy.

---

[287] 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* In re Estate of Cliffman, 2016 WL 4480882 (Mich. 2016) (Appellate Brief) (submitting supplemental authority with exhibits "show[ing] the relative frequency with which 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").

[288] *See* James C. Phillips & Jesse A. Egbert, Improving Corpus Design and Corpus-Based Analysis for Linguists and Lawyers: Principles and Practices from Survey and Content-Analysis Methodologies, Paper presented at Brigham Young University Law School's Law and Corpus Linguistic Conference, (Feb. 3, 2017) (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."[289] That undoubtedly holds true even more for an 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. So 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]."[290] And 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."[291]

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 a potential for bias and error."[292] 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.

---

[289] GEERAERTS, *supra* note __, at178.

[290] *Id.*

[291] Glynn, *supra* note __, at 7.

[292] *See* State v. Rasabout, 2015 UT 72, ¶ 134 (Lee, J., concurring in part and concurring in the judgment).