


August 2017

Corpus Linguistics as a Tool in Legal Interpretation

Lawrence M. Solan

Tammy Gales

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

 Part of the [Applied Linguistics Commons](#), [Constitutional Law Commons](#), and the [Legal Profession Commons](#)

Recommended Citation

Lawrence M. Solan and Tammy Gales, *Corpus Linguistics as a Tool in Legal Interpretation*, 2017 BYU L. Rev. 1311 (2018).
Available at: <https://digitalcommons.law.byu.edu/lawreview/vol2017/iss6/5>

This Article is brought to you for free and open access by the Brigham Young University Law Review at BYU Law Digital Commons. It has been accepted for inclusion in BYU Law Review by an authorized editor of BYU Law Digital Commons. For more information, please contact hunterlawlibrary@byu.edu.

Corpus Linguistics as a Tool in Legal Interpretation

Lawrence M. Solan* & Tammy Gales**

In this paper, we set out to explore conditions in which the use of large linguistic corpora can be optimally employed by judges and others tasked with construing authoritative legal documents. Linguistic corpora, sometimes containing billions of words, are a source of information about the distribution of language usage. Thus, corpora and the tools for using them are most likely to assist in addressing legal issues when the law considers the distribution of language usage to be legally relevant. As Thomas R. Lee and Stephen C. Mouritsen have so ably demonstrated in earlier work, corpus analysis is especially helpful when the legal standard for construction is the ordinary meaning of the document's terms. We argue here that four issues should be addressed before determining that corpus analysis is likely to be maximally convincing. First, the legal issue before the court must be about the distribution of linguistic facts. Second, the court must decide what makes an interpretation "ordinary." Third, if one wishes to search a corpus to glean the ordinary meaning of a term, one must decide in advance what to search. Fourth, there are different reasons as to why a particular meaning might present a weak showing in a corpus search, and these need to be understood. Each of these issues is described and discussed.

CONTENTS

INTRODUCTION.....	1312
-------------------	------

* Don Forchelli Professor of Law, Brooklyn Law School.

** Associate Professor of Linguistics, Hofstra University.

The authors express their gratitude to Jonah Gelbach and to other participants in the BYU Law and Corpus Linguistics Symposium for their valuable comments on an earlier draft of this paper. We also wish to thank the organizers and participants in the workshop on Historical Semantics and Legal Interpretation at the University of Chicago in May 2017, and Joshua Kaufman for his excellent work as our research assistant.

I. ORDINARY MEANING AS AN INTERPRETIVE NORM	1316
A. The Centrality of the Ordinary Meaning Principle	1316
B. The Defeasibility of the Ordinary Meaning Principle....	1320
II. FINDING ORDINARY MEANING.....	1331
A. Native Speaker Intuition.....	1332
B. Dictionaries	1334
III. TURNING TO CORPORA	1336
A. A Brief Introduction to Corpus Linguistics.....	1337
B. Four Conditions that Make Corpus Analysis Efficacious in Legal Analysis	1342
1. The legal issue in the case must concern the distribution of language usage over a particular population.....	1342
2. The court must adopt a clear notion of ordinary meaning	1342
3. Conducting the right search on the right corpus	1345
4. Caution in drawing inferences from the absence of particular usages in a corpus: the merits of double dissociation.....	1351
C. Getting it Right.....	1354
IV. CONCLUSION	1356

INTRODUCTION

In this Article, we explore conditions in which the use of large linguistic corpora can be optimally employed by judges and others tasked with construing authoritative legal documents. Linguistic corpora, sometimes containing billions of words, are a source of information about the distribution of language usage both between populations and within a single population. Thus, corpora and the tools for using them are most likely to assist in addressing legal issues when the law considers the distribution of language usage to be legally relevant. As Thomas R. Lee and Stephen C. Mouritsen have so ably demonstrated in earlier work, corpus analysis is especially

helpful when the legal standard for construction is the ordinary meaning of a law's terms.¹

We argue here that four issues should be addressed before turning to corpus linguistics as the most efficacious tool in statutory interpretation. First, the legal issue before the court must be about the distribution of linguistic facts. Surely, separating the "ordinary" sense of an expression from outlying ones meets this criterion. As we show below, however, courts do not universally adopt the ordinary meaning of a term. Instead, they often engage in detailed inquiry into the context of the legislation to determine what meaning was most likely intended.

Second, along these same lines, the court must decide, as a legal matter, what makes an interpretation "ordinary."² In one Supreme Court case, Justice Breyer held that a particular meaning of the word "carry" was ordinary because one-third of usages in a corpus of news articles conveyed the same meaning.³ In a sense it is ordinary, in that people appear to be comfortable using the word to convey that particular meaning. In another sense, it is not ordinary to the extent that a second meaning predominates over the remaining two-thirds of recorded instances of usage.⁴

Third, if one wishes to search a corpus to glean the ordinary meaning of a term, one must decide, in advance, what to search. In a case decided by the Supreme Court of Michigan, all seven justices agreed that a corpus search was a superior way to determine ordinary meaning but divided four to three on what terms to evaluate.⁵ This is especially apt to happen when a law contains a term that typically has one meaning when unmodified but is subject to modification that suggests the default meaning was not the intended one. The

1. See Thomas R. Lee & Stephen C. Mouritsen, *Judging Ordinary Meaning*, YALE L.J. (forthcoming); 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 (2011); Stephen C. Mouritsen, Note, *The Dictionary Is Not a Fortress: Definitional Fallacies and a Corpus-Based Approach to Plain Meaning*, 2010 BYU L. REV. 1915 [hereinafter Mouritsen, *Dictionary*].

2. See Lee & Mouritsen, *supra* note 1, for further discussion.

3. *Muscarello v. United States*, 524 U.S. 125, 129, 140 (1998).

4. We discuss this case in detail below. See Mouritsen, *Dictionary*, *supra* note 1, for a more refined corpus analysis of this case.

5. *People v. Harris*, 885 N.W.2d 832 (Mich. 2016).

Michigan case, for example, involved deciding whether the word *information* includes false statements or is ordinarily limited to true statements. While we can speak of *inaccurate information*, when we use the word without modification, we typically understand it as referring to actual facts.

Fourth, there are two very different reasons for a particular meaning to present a weak showing in a corpus search. In some instances, it is possible, but awkward, to use a particular expression to describe an event or a set of circumstances. For example, in *Chisom v. Roemer*, a case that dealt with the scope of the Voting Rights Act, the Court had to determine whether the law's reference to the election of *representatives* should be read in its ordinary sense, rendering the statute inapplicable to the election of judges.⁶ All nine Justices agreed that *representative* is usually applied to members of the legislative branch and not the judicial branch, but the majority construed the statute more broadly than the law's ordinary meaning based on the assumption that Congress did not intend to create a safe harbor for racially discriminatory schemes for the election of judges.⁷

In other cases, a particular usage may be absent from a corpus not because speakers are uncomfortable using the expression in that way, but because it reflects relevant circumstances that do not often arise. Consider *Smith v. United States*, the 1993 Supreme Court decision in which a majority of six Justices held that a person who attempted to trade his unloaded machine gun for illegal drugs had “‘use[d]’ . . . a firearm . . . ‘during and in relation to . . . [a] drug trafficking crime.’”⁸ The case hinged on whether he had “used the gun” in the ordinary sense of that expression. Justice Scalia’s colorful dissent argued that when we think about someone using a gun we think about that person using it as a weapon.⁹ He was right about that; however, apart from linguistic arguments, the majority argued that the structure of the statute suggested that a broader

6. *Chisom v. Roemer*, 501 U.S. 380 (1991).

7. *Id.* at 403–04.

8. *Smith v. United States*, 508 U.S. 223, 225 (1993) (second alteration in original) (quoting 18 U.S.C. § 924(c)(1) (1988)).

9. *Id.* at 242–44 (Scalia, J., dissenting).

interpretation of *use* was warranted.¹⁰ In contrast with understanding *representative* to include judges, moreover, it does not seem strange to say that Smith really did use a gun in a drug crime. He just used it in an unusual way.

Thus, there are two reasons that a meaning may be absent from a corpus. The first is a fact about language: people may not feel that the meaning fits comfortably within the concept that the word denotes and typically do not use it in those circumstances, although it may be possible to do so at the margins. The second reason that a particular meaning may be absent from a corpus concerns facts about the world, rather than facts about or knowledge of language. The blue pitta is a bird found in Asia but not North America.¹¹ It is no less a bird, and we are no less comfortable calling it a bird just because it does not appear in corpora of American English. This example further illustrates the necessity of selecting an appropriate corpus to analyze ordinary meaning.

Should the legal system concern itself with the distinct reasons for a term not appearing in a corpus? That depends on whether the legal system should examine ordinary meaning to inquire further into a statute's purpose or whether the legal system relies on ordinary meaning to capture the situations that likely triggered the legislation. In the former case, it matters *why* a particular use of a word does not appear in a corpus in a particular circumstance. Below, we argue that these two perspectives can be merged through a process of double dissociation.¹² That is, the strongest cases for using corpus analysis are ones in which not only does one meaning predominate over an alternative meaning in an appropriate corpus, but the second, less common meaning is generally expressed using language other than the language in the disputed statute. Put in linguistic terms, the question is whether the legal system should use corpora as a reflection of the distribution of language use as an *external* matter, or whether it should regard corpora as windows into the *internal* workings of the mind, to the extent that corpora reflect

10. *Id.* at 240.

11. J. Erritzoe, *Blue Pitta* (*Hydrornis cyaneus*), HANDBOOK OF THE BIRDS OF THE WORLD ALIVE (J. del Hoyo et al. eds., 2017), <http://www.hbw.com/species/blue-pitta-hydrornis-cyaneus>.

12. *See infra* Section III.B.4.

when people are actually comfortable using an expression in a particular circumstance.¹³

The remainder of this Article proceeds as follows: In Part I, we discuss ordinary meaning as a foundational principle of legal interpretation and the extent to which it is defeasible. Part II briefly outlines the ways that judges seek to determine a term's ordinary meaning—most often through their sense about word meaning but augmented by the use of dictionaries. Part III first presents an introduction to corpus linguistics, then discusses a recent trend to introduce linguistic corpora into the analysis, and, finally, evaluates the various cases according to our four proposed criteria listed above. Part IV is a brief conclusion.

I. ORDINARY MEANING AS AN INTERPRETIVE NORM

A. *The Centrality of the Ordinary Meaning Principle*

Scholars and judges from across the political spectrum routinely apply the ordinary meaning canon.¹⁴ To illustrate, consider Justice Scalia's majority opinion in the 2014 Supreme Court case *Burrage v. United States*.¹⁵ Burrage, a drug dealer, sold heroin to a user who died the next day from a drug overdose. The Controlled Substances Act calls for an enhanced sentence for a defendant who has sold illegal drugs when "death or serious bodily injury results from the use of such substance."¹⁶ At trial, medical experts testified that it was impossible to determine whether the heroin Burrage sold the user was a but-for cause of death or a contributing factor of a death that

13. Linguists and philosophers occupy both camps. For the leading internalist perspective, see NOAM CHOMSKY, KNOWLEDGE OF LANGUAGE: ITS NATURE, ORIGIN, AND USE (1986). For a good exposition of the externalist perspective, see ERNIE LEPORE & MATTHEW STONE, IMAGINATION AND CONVENTION: DISTINGUISHING GRAMMAR AND INFERENCE IN LANGUAGE (2015). For critique of Lepore and Stone, see Laurence Horn, *Conventional Wisdom Reconsidered*, 59 INQUIRY 145 (2016).

14. See, e.g., WILLIAM N. ESKRIDGE JR., INTERPRETING LAW: A PRIMER ON HOW TO READ STATUTES AND THE CONSTITUTION 33–84 (2016); ANTONIN SCALIA & BRYAN A. GARNER, READING LAW: THE INTERPRETATION OF LEGAL TEXTS 69–77 (2012); BRIAN G. SLOCUM, ORDINARY MEANING: A THEORY OF THE MOST FUNDAMENTAL PRINCIPLE OF LEGAL INTERPRETATION (2015); LAWRENCE M. SOLAN, THE LANGUAGE OF STATUTES: LAWS AND THEIR INTERPRETATION 50–81 (2010).

15. *Burrage v. United States*, 134 S. Ct. 881 (2014).

16. 21 U.S.C. § 841(b)(1)(C) (2012).

might have occurred anyway because of the various drugs ingested during the binge, including, but not limited to, the heroin Burrage had sold. This raised the interpretive issue of whether the death “resulted from the use” of the substance that Burrage provided.¹⁷ Justice Scalia resolved the issue by relying on the ordinary meaning of “resulting from,” which denotes causation:

The language Congress enacted requires death to “result from” use of the unlawfully distributed drug, not from a combination of factors to which drug use merely contributed. Congress could have written § 841(b)(1)(C) to impose a mandatory minimum when the underlying crime “contributes to” death or serious bodily injury, or adopted a modified causation test tailored to cases involving concurrent causes, as five States have done It chose instead to use language that imports but-for causality. Especially in the interpretation of a criminal statute subject to the rule of lenity, . . . we cannot give the text a meaning that is different from its ordinary, accepted meaning, and that disfavors the defendant.¹⁸

Justice Scalia provided two justifications for applying the ordinary meaning of the term. The first is that Congress presumptively intended this meaning. For, if Congress intended a different, broader interpretation, it could have said so in clear language—as other legislatures have done in enacting statutes covering the same situations. Thus, the ordinary meaning approach enhances democratic values because it presumptively captures the will of the legislature. His second argument, for which he adduced the rule of lenity, is quite different. Regardless of what the legislature intended, adhering to the ordinary meaning of statutory language enhances the rule of law because it takes into account the likelihood that citizens subject to the law—the criminal law in particular—are on adequate notice of behavioral norms that society will enforce.

The argument is reminiscent of Oliver Wendell Holmes’s 1931 opinion for a unanimous Court in *McBoyle v. United States*¹⁹ that a 1919 federal law criminalizing the movement of stolen vehicles across state lines should not be applied to airplanes since, at the time

17. *Burrage*, 134 S. Ct. at 885.

18. *Id.* at 891.

19. *McBoyle v. United States*, 283 U.S. 25 (1931).

of enactment, airplanes would not have been within the ordinary understanding of the word *vehicle*. At the conclusion of a very short opinion, Holmes remarked:

Although it is not likely that a criminal will carefully consider the text of the law before he murders or steals, it is reasonable that a fair warning should be given to the world in language that the common world will understand, of what the law intends to do if a certain line is passed. To make the warning fair, so far as possible the line should be clear. 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 simply because it may seem to us that a similar policy applies, or upon the speculation that if the legislature had thought of it, very likely broader words would have been used.²⁰

“Fair warning” to the accused is not about deference to the legislature. Rather, it is itself a very basic rule-of-law value, regardless of what the legislature had in mind. Chief Justice John Marshall articulated this principle in 1821 in *United States v. Wiltberger*.²¹ Wiltberger committed manslaughter by killing an individual while aboard an American vessel docked in the Tigris River in China.²² A federal statute made it a federal offense to commit certain crimes on American vessels abroad.²³ The provision that covered murder included murder committed “upon the high seas, or in any river, haven, basin, or bay, out of the jurisdiction of any particular State,” but the provision that covered manslaughter included only “the high seas.”²⁴

The most sensible inference to draw is that Congress simply made a mistake in omitting the other bodies of water from the manslaughter provision. Marshall recognized this but rejected it as a proper way to conduct statutory interpretation:

It has been argued, and, we admit, with great force, that in this section the legislature intended to take from a citizen offending

20. *Id.* at 27.

21. *United States v. Wiltberger*, 18 U.S. (1 Wheat.) 76 (1820).

22. *Id.* at 77.

23. *Id.* at 77–86.

24. *Id.* at 79–80.

against the United States, under colour of a commission from a foreign power, any pretence to protection from that commission; and it is almost impossible to believe that there could have been a deliberate intention to distinguish between the same offence, committed under colour of such commission, on the high seas, and on the waters of a foreign State, or of the United States, out of the jurisdiction of any particular State. This would unquestionably have been the operation of the section, had the words, ‘on the high seas,’ been omitted. Yet it would be carrying construction very far to strike out those words. Their whole effect is to limit the operation which the sentence would have without them; and it is making very free with legislative language, to declare them totally useless, when they are sensible, and are calculated to have a decided influence on the meaning of the clause.²⁵

Thus, Marshall favored an ordinary meaning approach because it gave fair warning to the defendant even though the legislature most likely intended a broader meaning.

Scholars have also espoused both rationales—carrying out the intent of the legislature and giving public notice of the law—for preferring a law’s ordinary meaning. William Eskridge, often a vocal critic of Scalia’s thinking,²⁶ makes very similar points when it comes to the centrality of ordinary meaning in statutory interpretation. Like Scalia, he sees value in the approach, both as a default interpretation most likely to reflect the will of the legislature and as a value in its own right, enhancing the law’s legitimacy because of its greater accessibility.²⁷ Brian Slocum, in his book on ordinary meaning, also emphasizes the overlapping values of employing the concept.²⁸

This consensus leaves open, however, a set of questions about which there is far less agreement. Among them are:

- What do we mean by ordinary meaning? When we speak of what a word ordinarily is used to mean, we can refer to the one most typical meaning or to a range of meanings that are

25. *Id.* at 99–100.

26. *See, e.g.*, William N. Eskridge, Jr., *The New Textualism and Normative Canons*, 113 COLUM. L. REV. 531 (2013) (reviewing SCALIA & GARNER, *supra* note 14).

27. ESKRIDGE, *supra* note 14, at 35.

28. *See* SLOCUM, *supra* note 14.

all reasonably common, regardless of which one is the most common.

- How much of a role should context play in our analysis? A word may typically be used to convey one thought in a particular context but typically be used to convey a different thought in another context.
- If one of those contexts occurs more frequently than the other, should we call the meaning associated with that context the ordinary meaning, or should we consider them both ordinary? For example, if we speak more about savings banks than we do about river banks, should we conclude that only the sense of the word that describes financial institutions is ordinary? Or should we say that both meanings are within ordinary usage?
- When a word whose meaning changes when modified is used without a modifier in a legal setting, should we assume that the legislature had the unmodified default interpretation in mind, or should we assume that it had in mind how that word might be understood with whatever modifications are likely to be relevant? Some words have one meaning when not accompanied by a modifier, such as an adjective, but can be modified to expand our range of plausible interpretations. An example from the case law discussed below is the word *information*.²⁹ We can speak of false or inaccurate information, but when we just use the word alone, we typically mean to describe an accurate account of facts.

B. The Defeasibility of the Ordinary Meaning Principle

As central as ordinary meaning is in the interpretation of laws, it is defeasible. Sometimes courts determine that the legislature intended an interpretation that is either broader or narrower than the prototypical use of a word or phrase. Narrower interpretations often occur when the context in which the law was enacted suggests that the legislature had in mind only a subset of the plausible situations in which the language of the statute might apply. By the

29. See *infra* Section III.B.3 (discussing conducting the search in the right corpora).

same token, courts sometimes construe a statute as having a broader interpretation than ordinary meaning would dictate.

To illustrate from some recent cases decided by the U.S. Supreme Court under Chief Justice Roberts, consider *Yates v. United States*.³⁰ Following corporate scandals that occurred in the first years of this century, Congress enacted the Sarbanes-Oxley Act, which placed disclosure obligations on corporate executives, accountants, and lawyers and prohibited the destruction of documents and other materials to thwart government investigations even if those materials had not yet been subpoenaed.³¹ This provision was a reaction to the behavior of Arthur Andersen LLP, the accounting firm of Enron Corporation, which had been manipulating prices in the energy sector.³² The accounting firm destroyed documents and electronically-stored information on the eve of a government investigation into its role in the Enron scandal.³³ As a result, the firm was convicted of obstructing justice.³⁴ That conviction, however, was reversed because the firm had not yet been subpoenaed and claimed that it was under no legal obligation to maintain the records.³⁵ Soon thereafter, in 2002, Congress made it clear that the destruction of documents in anticipation of a federal investigation would be a criminal act going forward. The relevant portion of the Sarbanes-Oxley Act provides:

Whoever knowingly alters, destroys, mutilates, conceals, covers up, falsifies, or makes a false entry in any record, document, or tangible object with the intent to impede, obstruct, or influence the investigation or proper administration of any matter within the jurisdiction of any department or agency of the United States . . . shall be fined under this title, imprisoned not more than 20 years, or both.³⁶

30. *Yates v. United States*, 135 S. Ct. 1074 (2015).

31. *Id.* at 1074.

32. For an overview of the facts surrounding Arthur Andersen's involvement in the Enron scandal, see *Arthur Andersen LLP v. United States*, 544 U.S. 696 (2005).

33. *Id.* at 696, 700.

34. *Id.* at 696–98.

35. *Id.* at 708.

36. 18 U.S.C. § 1519 (2012).

Fast forward a decade. John Yates was a commercial fisherman catching grouper in the Gulf of Mexico off the coast of Florida.³⁷ Some of the fish caught and crated on his fishing boat were smaller than the legal minimum size.³⁸ When law enforcement agents pulled up alongside Yates's vessel, Yates had the undersized fish thrown overboard and replaced them with larger fish in the same crates.³⁹ Yates was prosecuted and convicted by the trial court of violating the Sarbanes-Oxley Act for having destroyed a tangible object [dead fish] for the purpose of impeding a government investigation.⁴⁰

The Supreme Court reversed the conviction in a five-to-four decision, holding that the act should not be applied to this situation because, read in context, the relevant provision should be limited to the destruction of financial records, whether stored on paper or on a tangible object (such as a computer drive).⁴¹ Justice Ginsburg noted in her majority opinion:

Whether a statutory term is unambiguous, however, does not turn solely on dictionary definitions of its component words. Rather, “[t]he plainness or ambiguity of statutory language is determined [not only] by reference to the language itself, [but as well by] the specific context in which that language is used, and the broader context of the statute as a whole.”⁴²

Indeed, the *noscitur a sociis* canon says that statutory words should be construed in association with the words that surround them.⁴³ By the same token, the title of the statutory provision, “Destruction, Alteration, or Falsification of Records in Federal Investigations and Bankruptcy,” suggests that the law was intended to deal with documents and the like, not fish. Courts are reluctant to use a law's title to override the law's substantive provisions, but they

37. Yates v. United States, 135 S. Ct. 1079 (2015).

38. *Id.*

39. *Id.* at 1080.

40. *Id.* at 1080–81.

41. *Id.* at 1088–89.

42. *Id.* at 1081–82 (citations omitted).

43. See ESKRIDGE, *supra* note 14, at 76–78; SCALIA & GARNER, *supra* note 14, at 195–98.

are willing to take a statute's title into account to add sufficient context to help determine the scope of the words within the law.⁴⁴

Justice Kagan (usually an ally of Justice Ginsburg on the Court) wrote a dissenting opinion that began with the ordinary meaning approach to statutory interpretation: "When Congress has not supplied a definition, we generally give a statutory term its ordinary meaning."⁴⁵ Thus, the Court was divided over the extent to which it is appropriate to engage in individual inquiry into legislative intent in the teeth of the default rule of ordinary meaning.

Consider, similarly, *Bond v. United States*,⁴⁶ a 2014 unanimous Supreme Court decision (with three Justices concurring in the judgment on constitutional instead of statutory grounds). Dr. Bond discovered that the father of her close friend's soon-to-be-born child was Bond's own husband.⁴⁷ A microbiologist, she took from her workplace a powder that could cause a skin rash upon contact and spread it on her friend's door knob, mailbox, and other such surfaces to induce irritation.⁴⁸ Eventually it worked.⁴⁹ Bond was caught and prosecuted for violating the statute that implements the Convention on the Prohibition of the Development, Production, Stockpiling, and Use of Chemical Weapons and Their Destruction,⁵⁰ a treaty that the United States ratified in 1997. The statute defines *chemical weapon* as follows: "A toxic chemical and its precursors, except where intended for a purpose not prohibited under this chapter as long as the type and quantity is consistent with such a purpose."⁵¹ The chemicals used by Bond, if ingested in sufficient quantity, could be fatal.⁵² The statute defines *toxic chemical* as follows:

The term "toxic chemical" means any chemical which through its chemical action on life processes can cause death, temporary incapacitation or permanent harm to humans or animals. The term includes all such chemicals, regardless of their origin or of their

44. See SCALIA & GARNER, *supra* note 14, at 221.

45. *Yates*, 134 S. Ct. at 1091 (Kagan, J., dissenting).

46. *Bond v. United States*, 134 S. Ct. 2077 (2014).

47. *Id.* at 2085.

48. *Id.*

49. *Id.*

50. *Id.* at 2085.

51. 18 U.S.C. § 229F(1)(A) (2012).

52. *Bond*, 134 S. Ct. at 2085.

method of production, and regardless of whether they are produced in facilities, in munitions or elsewhere.⁵³

Thus, the substance used by Bond met the statutory definition of *toxic chemical*.

The Supreme Court reversed the conviction and its affirmance by the Third Circuit.⁵⁴ It held that the Chemical Weapons Convention was intended to respond to war crimes, not to the facts of this case, which seemed to be a simple assault and more appropriate for local prosecution.⁵⁵ Chief Justice Roberts wrote for the majority:

But even with its broadly worded definitions, we have doubts that a treaty about *chemical weapons* has anything to do with Bond's conduct. The Convention, a product of years of worldwide study, analysis, and multinational negotiation, arose in response to war crimes and acts of terrorism. There is no reason to think the sovereign nations that ratified the Convention were interested in anything like Bond's common law assault.⁵⁶

Roberts had a point—one would not ordinarily regard what Dr. Bond did as a crime that violated a ban on the use of chemical weapons. But, as Roberts comes close to conceding, if one pays attention to the statutory definition of *toxic chemical*, rather than to the everyday use of *chemical warfare*, the linguistic argument is hard to maintain.⁵⁷ More convincing is the structural argument, similar to the majority's argument in *Yates*, that the law implementing the treaty was not likely intended to extend its reach to minor crimes outside the ordinary jurisdiction of the federal government and within the constitutional power of the states.⁵⁸ Thus, ordinary meaning bowed to the statute's structure, its purpose, and principles of federalism.

53. 18 U.S.C. § 229F(8)(A).

54. *Bond*, 134 S. Ct. at 2077.

55. *Id.* at 2087.

56. *Id.* (citation omitted).

57. *Id.* at 2094.

58. *Id.* at 2093–95.

In separate concurring opinions, Justices Scalia,⁵⁹ Thomas,⁶⁰ and Alito⁶¹ would have declared the law implementing the treaty unconstitutional as a usurpation of state power. Justice Scalia made it clear that as far as he was concerned, Bond had violated the law, based on its language.⁶² Thus, he rejected the majority's reliance on the ordinary meaning of the words in the title and, implicitly, its structural argument.⁶³ Scalia wrote:

The statute parses itself. There is no opinion of ours, and none written by any court or put forward by any commentator since Aristotle, which says, or even suggests, that 'dissonance' between ordinary meaning and the unambiguous words of a definition is to be resolved in favor of ordinary meaning.⁶⁴

As for the constitutional argument, which all three dissenters embraced as a reason for striking down the statute as outside congressional authority, Justice Alito summarized the argument succinctly:

The control of true chemical weapons, as that term is customarily understood, is a matter of great international concern, and therefore the heart of the Convention clearly represents a valid exercise of the treaty power. But insofar as the Convention may be read to obligate the United States to enact domestic legislation criminalizing conduct of the sort at issue in this case, which typically is the sort of conduct regulated by the States, the Convention exceeds the scope of the treaty power.⁶⁵

The debate among the justices in *Bond* illustrates a long-standing tension in statutory analysis between construing statutes in a manner that would avoid constitutional issues and addressing constitutional issues directly. What is unusual about *Bond* is that both sides of the debate came to the same conclusion about the outcome of the case.

59. *Id.* at 2094 (Scalia, J., concurring).

60. *Id.* at 2102 (Thomas, J., concurring).

61. *Id.* at 2111 (Alito, J., concurring).

62. *Id.* at 2094 (Scalia, J., concurring).

63. *See id.* at 2096–97.

64. *Id.* at 2096.

65. *Id.* at 2111 (Alito, J., concurring).

A more typical illustration is *United States v. X-Citement Video, Inc.*,⁶⁶ a case interpreting the federal law criminalizing the transmission of child pornography. The statute was poorly drafted, so that one interpretation seemed to impose strict criminal liability for those who knew they were sending some kind of depiction (video tapes in this instance) but did not know that the depictions contained child pornography.⁶⁷ The statute read in relevant part:

- (a) Any person who—
 - (1) knowingly transports or ships using any means or facility of interstate or foreign commerce or in or affecting interstate or foreign commerce by any means including by computer or mails, any visual depiction, if—
 - (A) the producing of such visual depiction involves the use of a minor engaging in sexually explicit conduct; and
 - (B) such visual depiction is of such conduct;
 -
- shall be punished as provided in subsection (b) of this section.⁶⁸

The owner of a business called X-Citement Video was peddling videos containing child pornography.⁶⁹ He claimed that the statute was unconstitutional because it required the government to prove that he knew that he was distributing pornographic videos but not that he knew the players were underage.⁷⁰ The wording of the statute, however, does not appear to require that the defendant know that he was shipping depictions containing child pornography.⁷¹ *Knowingly* seems to modify only “transports or ships in interstate . . . commerce . . . any visual depiction.”⁷² He certainly knew he was doing that.

The justices agreed that absent a mens rea requirement for the element that the depiction contain pornography, the law would be

66. *United States v. X-Citement Video, Inc.*, 513 U.S. 64 (1994).

67. *Id.* at 68–70.

68. 18 U.S.C. § 2252 (2012).

69. *X-Citement Video*, 513 U.S. at 66.

70. *Id.* at 67.

71. *Id.* at 68.

72. *Id.*

unconstitutional.⁷³ A majority of seven, in an opinion written by Chief Justice Rehnquist, rejected “the most grammatical reading” of the language as written, judicially amending the statute to include a state of mind requirement for the element that the depiction contain child pornography.⁷⁴ Justice Stevens concurred by saying that the statute, as written, should be construed as containing such a requirement.⁷⁵ Justice Scalia dissented, saying the statute was unconstitutional,⁷⁶ presaging his concurrence in *Bond* some twenty years later.⁷⁷

Courts do, at times, give statutory terms meanings that are broader than their ordinary usage would suggest. For example, *Chisolm v. Roemer*,⁷⁸ discussed above,⁷⁹ held that state supreme court justices come within the category of *representatives* under the Voting Rights Act.⁸⁰

More recently in 2015, in *King v. Burwell*,⁸¹ the Court construed language in the Affordable Care Act (“Obamacare”) more broadly than ordinary usage would permit.⁸² The statute requires those with incomes deemed sufficient to pay the premiums⁸³ to purchase health insurance or pay a fine.⁸⁴ It further calls for federal government subsidies through the tax system for those who do not meet the income threshold and who purchase their insurance on “an Exchange established by the State.”⁸⁵ Not all of the states established exchanges on which its residents could buy health insurance. The statute called for the federal government to establish exchanges in those states that decided not to participate on their own.⁸⁶ The

73. *Id.* at 78.

74. *See id.* at 70 (discussing whether the justices should amend the statute judicially to include the state of mind requirement that would allow it to meet constitutional requirements).

75. *See id.* at 79 (Stevens, J., concurring).

76. *Id.* at 86 (Scalia, J., dissenting).

77. *Bond v. United States*, 134 S. Ct. 2077, 2094–2111 (2014).

78. *Chisolm v. Roemer*, 501 U.S. 380 (1991).

79. *See supra* Intro.

80. *Chisolm*, 501 U.S. at 399.

81. *King v. Burwell*, 135 S. Ct. 2480 (2015).

82. *Id.* at 2492.

83. 26 U.S.C. §§ 5000A(a), (e)(1)(A) (2012).

84. 26 U.S.C. § 5000A.

85. 26 U.S.C. § 36B(b)(2)(A).

86. *King*, 135 S. Ct. at 2487.

plaintiffs in *King* were residents of Virginia, a state that did not establish an exchange.⁸⁷ Rather, the federal government established the exchange on its behalf.⁸⁸ The plaintiffs claimed that, because the Virginia insurance exchange was not “established by the state,” the residents were not entitled to the subsidies through the Internal Revenue Service and, therefore, should be considered outside the mandate requiring them to buy the insurance.⁸⁹

Writing for a majority of six Justices, Chief Justice Roberts held the statute to be “ambiguous.”⁹⁰ The majority acknowledged that the ordinary interpretation of the phrase *established by the State* would suggest that the state itself had to establish the exchange for the provision to apply. However, the Court held that, in the context of the entire statute, the intended meaning of the expression included exchanges established by the federal government on behalf of the state if the state did not set up its own exchange:

Petitioners’ arguments about the plain meaning of Section 36B are strong. But while the meaning of the phrase “an Exchange established by the State . . . ” may seem plain “when viewed in isolation,” such a reading turns out to be “untenable in light of [the statute] as a whole.” In this instance, the context and structure of the Act compel us to depart from what would otherwise be the most natural reading of the pertinent statutory phrase.⁹¹

We do not agree with the Chief Justice that the law was “ambiguous.”⁹² Rather, it was *inconsistent* or contained an *incoherence*. More convincing is the Court’s structural argument—without the subsidies, the entire system of the act would collapse in a majority of the states, and it is unlikely that Congress intended that to happen.⁹³

For jurisprudential reasons, the Court was reluctant to say that Congress had erred in its wording (choosing instead to declare the

87. *Id.*

88. *Id.*

89. *Id.* at 2488.

90. *Id.* at 2492.

91. *Id.* at 2495. (quoting *Dep’t of Revenue of Ore. v. ACF Indus., Inc.*, 510 U.S. 332, 343 (1994)).

92. *Id.* at 2491.

93. *See id.* at 2496.

statute ambiguous),⁹⁴ but in all likelihood the inconsistent language among the statute's provisions was an oversight in a complicated law whose language was being negotiated up until the time of enactment. As Abbe Gluck explains in her perceptive article about this case, the language arose from the Senate bill's having drawn on language from the drafts of two different committees, and because of the procedure used to enact the law, there was no opportunity to clean it up through a conference committee that would reconcile the House and Senate versions.⁹⁵ The Chief Justice recognized this reality and resolved the law's inconsistencies to further its "overall statutory scheme"⁹⁶ as identified in the structure of the statute itself.

Moreover, the Supreme Court has, in other contexts, had little trouble inferring that *state* means *state or federal*. *Mims v. Arrow Financial Services, LLC*,⁹⁷ involved the interpretation of the Telephone Consumer Protection Act of 1991 (TCPA), which prohibits telemarketers from engaging in robocalls and other such practices. If a pattern of such conduct emerges within a state, the state Attorney General may bring an action. With respect to TCPA actions brought by State Attorneys General, the statute provides: "The district courts of the United States . . . have exclusive jurisdiction."⁹⁸ The statute also grants a private right of action. With respect to the jurisdiction of those cases, the statute says:

A person or entity may, if otherwise permitted by the laws or rules of court of a State, bring in an appropriate court of that State—

(A) an action based on a violation of this subsection or the regulations prescribed under this subsection to enjoin such violation,

94. For a discussion of the Court's reluctance to declare a statute to be erroneously drafted and then to correct it, see Ryan D. Doerfler, *The Scrivener's Error*, 110 NW. U. L. REV. 811, 813 (2016) (arguing that the Court generally requires uncontestable evidence of an error before applying the doctrine).

95. See Abbe R. Gluck, *Imperfect Statutes, Imperfect Courts: Understanding Congress's Plan in the Era of Unorthodox Lawmaking*, 129 HARV. L. REV. 62, 63, 76–77 (2015).

96. *King*, 135 S. Ct. at 2492. For a theory of legislation as plans, see SCOTT J. SHAPIRO, LEGALITY (2011).

97. *Mims v. Arrow Fin. Servs., LLC*, 565 U.S. 368, 368–70 (2012). Our thanks to Jonah Gelbach for bringing this case to our attention. This case was not part of the discussion in *King*.

98. 47 U.S.C. § 227(g)(2) (2012).

- (B) an action to recover for actual monetary loss from such a violation, or to receive \$500 in damages for each such violation, whichever is greater, or
(C) both such actions.⁹⁹

Mims believed that Arrow Financial Services' calls had violated the statute and brought a case in federal court. Arrow successfully moved to dismiss the case for lack of subject matter jurisdiction based on the provision just quoted. The dismissal was upheld on appeal. However, in a unanimous decision, the Supreme Court reversed. The Court ruled that the specific mention of state jurisdiction does not strip the court of the federal question jurisdiction it would ordinarily have as a matter of law. Comparing the exclusivity provision for Attorneys General to the grant of jurisdiction to state courts in private rights of action, the Court reasoned: "Section 227(g)(2)'s exclusivity prescription 'reinforce[s] the conclusion that [47 U.S.C. § 227(b)(3)'s] silence . . . leaves the jurisdictional grant of § 1331 untouched. For where otherwise applicable jurisdiction was meant to be excluded, it was excluded expressly."¹⁰⁰

By the same token, if Congress had intended to permit the ACA to fail by its very structure, disallowing subsidies for those who live in states in which the federal government initiated exchanges on behalf of the state, it most likely would have indicated this intent directly in one way or another.¹⁰¹

Justice Scalia would have none of this:

Any effort to understand rather than to rewrite a law must accept and apply the presumption that lawmakers use words in "their natural and ordinary signification." Ordinary connotation does not always prevail, but the more unnatural the proposed interpretation of a law, the more compelling the contextual evidence must be to show that it is correct. Today's interpretation is not merely unnatural; it is unheard of. Who would ever have dreamt that

99. 47 U.S.C. § 227(b)(3).

100. *Mims*, 565 U.S. at 380–81 (alterations in original) (quoting *Verizon Md., Inc. v. Pub. Serv. Comm'n of Md.*, 535 U.S. 635, 644 (2002)).

101. See Anita S. Krishnakumar, *The Sherlock Holmes Canon*, 84 GEO. WASH. L. REV. 1 (2016).

“Exchange established by the State” means “Exchange established by the State *or the Federal Government*”?¹⁰²

Thus, whether interpreting a statute more narrowly or more broadly than ordinary meaning would permit, it is within the court’s discretion to treat ordinary meaning as defeasible and to use contextual information—including a statute’s purpose and structure—as interpretive tools. The debate among jurists and scholars is how defeasible ordinary meaning should be in these circumstances. This issue sets a boundary condition on the efficacy of corpus analysis in statutory interpretation since corpora are of little use when a court determines that fidelity to a statute’s purpose should take priority over the ordinary meaning of its terms.

II. FINDING ORDINARY MEANING

Once a court determines that the ordinary meaning of a statute’s language should carry the day, it must determine which of the alternative interpretations proposed by the parties is the “ordinary” one. The task itself assumes that words have a single, ordinary meaning and that, presumably, other meanings are outliers, perhaps extraordinary meanings. Sometimes this presumption seems reasonable. Returning to *Chisom v. Roemer*,¹⁰³ for example, the term *representative* really *is* typically used to refer to legislators, and judges are typically not referred to as representatives. In other instances, however, it is not at all clear that one meaning is ordinary while others are not.¹⁰⁴ To make this determination requires the legal system to reach an understanding of what makes ordinary meaning ordinary. As Lee and Mouritsen¹⁰⁵ observe, the system has not yet done so.

Judges typically use three methods in determining the ordinary meaning of a term: their own knowledge of the language, dictionaries as a purportedly neutral source of information about meaning, and empirical research using a corpus of English. As an

102. *King v. Burwell*, 135 S. Ct. 2480, 2497 (2015) (Scalia, J., dissenting) (citation omitted).

103. *Chisom v. Roemer*, 501 U.S. 380 (1991).

104. See discussion of *Smith v. United States*, *supra* text accompanying notes 8–10.

105. See Lee & Mouritsen, *supra* note 1.

initial matter, even when a court decides that a statutory word should be given a meaning that is broader or narrower than ordinary usage, the court is likely to engage in ordinary meaning analysis as part of the decision to look elsewhere. Thus, whatever method of interpretation a court adopts, finding the ordinary meaning of a term is likely to be part of the analysis in which the court engages.

A. Native Speaker Intuition

All of us use knowledge of our native language as our principal interpretive tool. A colorful example came from the 1973 Watergate hearings in which Senator Sam Ervin, Chair of the Senate Select Committee, questioned President Nixon's Chief Domestic Advisor, Mr. Ehrlichman.¹⁰⁶ The questioning concerned the government's having broken into the office of the psychiatrist treating Daniel Ellsberg, the individual responsible for the public release of the Pentagon Papers.¹⁰⁷

Senator Sam Ervin: The foreign intelligence activities had nothing to do with the opinion of Ellsberg's psychiatrist about his intellectual or emotional or psychological state.

John Ehrlichman: How do you know that, Mr. Chairman?

Senator Sam Ervin: Because I can understand the English language. It's my mother tongue.¹⁰⁸

This process is good enough to get us through our lives every day. In the realm of statutory interpretation, however, the task is not limited to understanding the gist of what someone is saying but

106. For discussion of Ehrlichman's positions in the Nixon Administration, see *John D. Ehrlichman*, RICHARD NIXON PRESIDENTIAL LIBR. & MUSEUM, <https://www.nixonlibrary.gov/forresearchers/find/textual/special/smf/ehrllichman.php> (last visited Jan. 27, 2018).

107. The Pentagon Papers is the informal name given to a report of the U.S. Department of Defense entitled *United States-Vietnam Relations, 1945-1967: A Study Prepared by the Department of Defense*. It received notoriety because it contained facts and explanations at odds with the public statements of President Johnson about U.S. involvement in the Vietnam War. The Pentagon Papers are available at http://www.nytimes.com/interactive/us/2011_PENTAGON_PAPERS.html.

108. *Video: U.S. Senator Sam Ervin, Jr., Questions John Ehrlichman, an Aide to President Richard Nixon, During the Televised Senate Watergate Hearings in 1973*, BRITANNICA KIDS, <http://kids.britannica.com/comptons/art-173893/US-Senator-Sam-Ervin-Jr> (last visited Jan. 27, 2018).

requires making subtle line-drawing decisions about when one category should end and another begin. Thus, the interpretation of statutory language often involves judgments about far smaller distinctions than everyday life demands.

Moreover, while people within a speech community¹⁰⁹ seem to share intuitions about whether a sentence is grammatical or not,¹¹⁰ there seems to be far less consensus about facts concerning the relative distribution of word usage. People are subject to what psychologists have termed *false consensus bias*. We tend to assume ourselves to be normal.

Solan, Rosenblatt, and Osherson studied false consensus bias as applied to the interpretation of equivocal contract terms.¹¹¹ In one scenario, participants in an experiment were asked whether they believed that damage to property should be attributed to “earth movement” when the damage was caused by the percussive force of a neighbor’s blasting activity as part of a construction project.¹¹² Their responses varied: 40% said yes, 40% said no, and 20% said they could not tell.¹¹³ These same people were then told that 100 people just like them had been asked these questions, and were asked to estimate how many of those hundred agreed with their judgments. The percentages were grossly overestimated: 67%, 63%, and 36%, respectively.¹¹⁴ Judges who participated in the study also exhibited

109. A speech community is a group of people who both speak the same language and share norms for its usage. See Richard Norquist, *Speech Community in Sociolinguistics*, THOUGHTCO., <https://www.thoughtco.com/speech-community-sociolinguistics-1692120> (last updated Apr. 26, 2017). There is disagreement about how to define the term. For a discussion of this topic, see generally Peter L. Patrick, *The Speech Community*, in THE HANDBOOK OF LANGUAGE VARIATION AND CHANGE 573 (J.K. Chambers et al. eds., 2002).

110. For recent research comparing the intuitions of linguists in professional journals with survey data, see generally Kyle Mahowald et al., *SNAP Judgments: A Small N Acceptability Paradigm (SNAP) for Linguistic Acceptability Judgments*, 92 LANGUAGE 619 (2016); Jon Sprouse, Carson T. Schütze & Diogo Almeida, *A Comparison of Informal and Formal Acceptability Judgments Using a Random Sample from Linguistic Inquiry 2001-2010*, 134 LINGUA 219 (2013); Jon Sprouse & Diogo Almeida, *Assessing the Reliability of Textbook Data in Syntax: Adger’s Core Syntax*, 48 J. LINGUISTICS 609 (2012).

111. Lawrence Solan, Terri Rosenblatt & Daniel Osherson, *False Consensus Bias in Contract Interpretation*, 108 COLUM. L. REV. 1268 (2008).

112. *Id.* at 1287–90.

113. *Id.* at 1290.

114. *Id.*

false consensus bias.¹¹⁵ People tend to believe that their own intuitions comport with those of others. Moreover, when false consensus bias is added to the inevitable problem of confirmation bias in substantive decision making,¹¹⁶ intuitive estimates of facts about how word usage is distributed do not appear likely to be very reliable.

B. Dictionaries

We will not belabor the fact that judicial reliance on dictionaries has been the subject of sharp criticism by legal academics¹¹⁷ and, to some extent, by judges themselves.¹¹⁸ The biggest problem with dictionaries is that, because they most often list meanings without sufficient context, it is possible to select from among the reported meanings without regard to whether that usage of the word is appropriate in the context of the legal dispute at hand. Thus, there is an incentive for judges to select dictionaries—and particular definitions—in a manner that tends to support the position they are taking. This incentive negates whatever objectivity there is in using a dictionary as opposed to the judge's own sense of a word's meaning.

For example, consider *MCI Telecommunications Corp. v. AT&T*,¹¹⁹ which the Supreme Court decided in 1994. After the 1970s breakup of AT&T, which had a monopoly on telephone service in the United States, the Federal Communications Commission (FCC) began to implement regulations that would remove barriers for new entrants into the market.¹²⁰ The Federal

115. *Id.* at 1292.

116. For a good discussion of the confirmation bias, see generally Raymond S. Nickerson, *Confirmation Bias: A Ubiquitous Phenomenon in Many Guises*, 2 REV. GEN. PSYCHOL. 175 (1998).

117. For an excellent review of this literature, see James Brudney & Lawrence Baum, *Oasis or Mirage: The Supreme Court's Thirst for Dictionaries in the Rehnquist and Roberts Eras*, 55 WM. & MARY L. REV. 483 (2013).

118. In cases discussed below, Judge Posner criticizes their use in *United States v. Costello*, 666 F.3d 1040, 1044 (7th Cir. 2012) (summarizing literature critical of judicial reliance on dictionaries to ascertain ordinary meaning, focusing on the gap between the context-sensitive use of words, and the acontextual nature of dictionary definitions), as does Associate Chief Justice Lee of the Supreme Court of Utah in *State v. Rasabout*, 2015 UT 72, ¶¶ 42–46, 356 P.3d 1258, 1272–73 (Lee, J., concurring).

119. *MCI Telecomms. Corp. v. AT&T*, 512 U.S. 218 (1994).

120. *Id.* at 220.

Communications Act required telephone companies to publish a complete listing of their rates¹²¹ and not charge more than what they published. At the time, telephone rates depended on the locations of the caller and the recipient, so this requirement was a burdensome one, especially for a company intending to operate nationwide. The statute also permitted the FCC to “modify any requirement made by or under the authority of this section either in particular instances or by general order applicable to special circumstances or conditions.”¹²² As part of its efforts to reduce entry barriers to new carriers, the FCC “modified” the publication requirement by having it apply only to “dominant carriers,” as it defined the term.¹²³ Because AT&T was the only dominant carrier at the time, the change in regulation would certainly serve the purpose of making it easier for competitors to enter the telephone service market.

AT&T cried foul and sued, naming MCI as one of the non-dominant carriers that benefitted from the new regulations.¹²⁴ The case made its way to the Supreme Court, which ruled in favor of AT&T by a 5–3 vote.¹²⁵ The decision hinged on whether the FCC had exceeded its statutory authority by eliminating the publication requirement for a subset of the carriers.¹²⁶

Writing for the majority, Justice Scalia combined etymology and lexicography to argue that *modify* must describe only incremental changes, not wholesale elimination:

The word “modify”—like a number of other English words employing the root “mod-” (deriving from the Latin word for “measure”), such as “moderate,” “modulate,” “modest,” and “modicum”—has a connotation of increment or limitation.

121. Federal Communications Act of 1934, ch. 652, 48 Stat. 1064 (codified as amended in scattered sections of 47 U.S.C.); 47 U.S.C. § 203(a) (2012).

122. 47 U.S.C. § 203(b)(2). The Supreme Court describes the history of these new regulations in *MCI Telecomms. Corp.*, 512 U.S. at 221.

123. *MCI Telecomms. Corp.*, 512 U.S. at 220–21.

124. *Id.* at 221. Of course, AT&T’s dispute was with the FCC, not MCI, which merely took advantage of the regulations that had been put in place. Various provisions of the Federal Communications Act permitted MCI to be brought into the case as a party.

125. *Id.* at 213.

126. *Id.* at 225.

Virtually every dictionary we are aware of says that “to modify” means to change moderately or in minor fashion.¹²⁷

He went on to cite numerous dictionaries and to berate the dissenting justices for relying on *Webster’s Third*, which includes making “a basic or important change” in its definition.¹²⁸

We present this extensive detail because the case illustrates what is generally problematic when courts rely on dictionaries to determine a word’s ordinary meaning. Scalia framed the issue as whether *modify* ordinarily describes an incremental change, rather than a major change. Dictionaries do not typically contextualize their definitions, so it should come as no surprise that it would have an answer that supports the majority’s position.

The decision to frame the issue as such, however, is not a necessary one. Scalia characterized the FCC as having eliminated the publication requirement, thus making a major change. It is possible, however, to understand the new regulations as leaving the publication requirement as it was but adjusting the population to which it applied. To illustrate, if a high school requires all students to have lunch on campus, the principal might announce that the rule has been “modified” by permitting seniors to have their lunch off campus.¹²⁹ We recognize that one may decide that eliminating the publication requirement for all but a single actor is more than a modification. Our point here, however, is that dictionary definitions are not likely to be of much help in cases like this, in which the linguistic dispute is subtler than are the lexicographers’ decisions about how to word the definitions, or perhaps the publishing company’s policies.

III. TURNING TO CORPORA

With the risks of overstating the “ordinariness” of one’s own intuitions and keeping the limitations of dictionaries in mind, some judges have begun to turn to big data to determine ordinary meaning.¹³⁰ Below we discuss four cases in which courts turned to

127. *Id.*

128. *Id.* at 225–26.

129. See generally SOLAN, *supra* note 14.

130. See *infra* Section III.B.

corpora to determine the likely meaning intended by the legislature. For each of these cases, we determine the extent to which the four conditions set forth in our introduction for optimal corpus analysis are present. Where we discover gaps, we point out the uncertainty in meaning left unaddressed. First, though, we present a brief introduction to corpus linguistics to provide context for readers not trained in its methods.

A. A Brief Introduction to Corpus Linguistics

Corpus linguistics is an empirical approach to the study of language variation and use “resulting in research findings which have much greater generalizability and validity than would otherwise be feasible.”¹³¹ Its two primary goals are “assessing the extent to which a [linguistic] pattern is found” and “analyzing the contextual features that influence [linguistic] variability.”¹³² For the past few decades, corpus methods have been revolutionizing all branches of linguistics and, more recently, providing more scientifically grounded insights into the study of language in forensic and legal contexts.

A corpus is a large collection of naturally occurring texts that are sampled to be representative of a particular type of language variety.¹³³ There are two basic types of corpora: general and specialized. General corpora—sometimes called reference corpora—are large (frequently millions to billions of words) and usually aim to capture a range of registers that are representative of a common language variety.¹³⁴ Specialized corpora are typically smaller (frequently thousands to millions of words) and focus on a more specific or less accessible variety of language.¹³⁵

131. Douglas Biber, *Corpus-Based and Corpus-Driven Analyses of Language Variation and Use*, in THE OXFORD HANDBOOK OF LINGUISTIC ANALYSIS 193, 193 (Bernd Heine & Heiko Narrog eds., 2012).

132. DOUGLAS BIBER, SUSAN CONRAD & RANDI REPPEN, *CORPUS LINGUISTICS: INVESTIGATING LANGUAGE STRUCTURE AND USE* 3 (1998).

133. TONY MCENERY ET AL., *CORPUS-BASED LANGUAGE STUDIES: AN ADVANCED RESOURCE BOOK* 4 (2006).

134. For an excellent discussion of representativeness in corpus design in this volume, see James C. Phillips & Jesse Egbert, *A Concise How-to-Guide for Law and Corpus Linguistics: Importing Principles and Practices from Survey and Content-Analysis Methodologies to Improve Corpus Design and Analysis*, 2017 BYU L. REV. 1589.

135. MCENERY ET AL., *supra* note 133, at 15.

Examples of general corpora include Brigham Young University's Corpus of Historical American English (COHA), Corpus of Global Web-based English (GloWbE), and Corpus of Contemporary American English (COCA), the last of which is probably the best-known, publicly available reference corpus and comprises 520 million words from 1990 to 2015, balanced over five registers (one spoken and four written).¹³⁶ Examples of specialized corpora include the Blog Authorship Corpus (BAC),¹³⁷ the Communicated Threat Assessment Reference Corpus (CTARC),¹³⁸ the Enron Email Dataset/Corpus,¹³⁹ and BYU's Corpus of U.S. Supreme Court Opinions (COSCOTUS).¹⁴⁰

The creation and use of corpora are not new phenomena. Early field anthropologists and lexicographers used to collect individual words on slips of paper, documenting their origin, date of acquisition, meaning, and, occasionally, the context in which they were used.¹⁴¹ These "shoebox" corpora were the basis for some of the first English dictionaries that included contextual information about a word: Charles Richardson's *New Dictionary of the English Language* (1837) and James Murray's subsequent twelve-volume *New English Dictionary on Historical Principles* (1884–1928)—now better-known as the *Oxford English Dictionary*, or the OED.¹⁴² These early corpora gave lexicographers observable evidence to provide more information about a word's origin, sense differentiation, and examples of use in context.

136. See CORPUS.BYU.EDU, <http://corpus.byu.edu/> (last visited Jan. 27, 2018); CORPUS CONT. AM. ENG., <http://corpus.byu.edu/coca/> (last visited Jan. 27, 2018).

137. BLOG AUTHORSHIP CORPUS, <http://u.cs.biu.ac.il/~koppel/BlogCorpus.htm> (last visited Jan. 27, 2018).

138. Tammy Gales, *Ideologies of Violence: A Corpus and Discourse Analytic Approach to Stance in Threatening Communications* (June 2010) (unpublished Ph.D. dissertation, University of California, Davis).

139. ENRON EMAIL DATASET, <http://www.cs.cmu.edu/~enron/> (last visited Jan. 27, 2018).

140. CORPUS U.S. SUP. CT. OPINIONS, <http://corpus.byu.edu/scotus/> (last visited Jan. 27, 2018).

141. MCENERY ET AL., *supra* note 133, at 3; M.A.K. Halliday, *Lexicology, in* LEXICOLOGY AND CORPUS LINGUISTICS: AN INTRODUCTION 1, 14 (M.A.K. Halliday et al. eds., 2004).

142. Halliday, *supra* note 141, at 15.

In the 1960s, what is considered the first modern corpus was released: the Brown University Standard Corpus of Present-day American English, which sampled 500 chunks of 2000 words, each from texts from fifteen categories such as press reportage, religion, popular lore, mystery fiction, and humor.¹⁴³ With continuing advances in computer technology, in 1987, John Sinclair at the University of Birmingham, United Kingdom, published the first fully corpus-based dictionary, *COBUILD*, which was based on the Bank of English corpus.¹⁴⁴ And, in 1999, the first fully corpus-based grammar to highlight register variation was published: the *Longman Grammar of Spoken and Written English*, which was based entirely on the forty-million-word Longman Spoken and Written English Corpus.¹⁴⁵ Since then, the number and size of corpora have increased and represent a range of written and spoken global language varieties.

There are now many ways in which general and specialized corpora are being compiled.¹⁴⁶ Corpora can be diachronic or synchronic. Diachronic corpora include similar kinds of texts across a range of time periods, allowing for investigations of language use and change over time. Synchronic corpora, on the other hand, include texts from a single time period, allowing for closer comparisons of language varieties within that time frame. Two well-known examples of diachronic, or historical, corpora are A Representative Corpus of Historical English Registers (ARCHER)—a 1.8 million-word corpus of British and American English from a range of genres from 1650 to 1990¹⁴⁷—and the Diachronic Part of the Helsinki Corpus of English Texts (the Helsinki corpus)—a 1.5 million word collection of Old to Early Modern English from the

143. MCENERY ET AL., *supra* note 133, at 4, 61–62.

144. Wolfgang Teubert, *Language and Corpus Linguistics*, in LEXICOLOGY AND CORPUS LINGUISTICS, *supra* note 141, at 73, 111.

145. DOUGLAS BIBER ET AL., LONGMAN GRAMMAR OF SPOKEN AND WRITTEN ENGLISH 4 (1999).

146. For an up-to-date compilation of corpus tools, corpora, and corpus research, see Martin Weisser's CORPUS-BASED LINGUISTICS LINKS, http://martinweisser.org/corpora_site/CBLLinks.html (last updated Oct. 14, 2016) (formerly David Lee's *Corpus-based Resources*). The website also contains additional information for corpora in other languages and language varieties (e.g., learner, multilingual, and parallel corpora). *Id.*

147. *See id.*

eighth to eighteenth centuries.¹⁴⁸ A third example is BYU's soon-to-be released Corpus of Founding Era American English (COFEA), which will include at least 100 million words from a range of registers from 1760 to 1799.¹⁴⁹ A popular collection of synchronic corpora is the International Corpus of English (ICE), which currently consists of one-million-word corpora sampled from similar registers and time periods from twenty-three different English-speaking countries, such as Canada, Fiji, India, Ireland, Malta, New Zealand, Nigeria, and South Africa, making comparative studies across language varieties possible.¹⁵⁰

Corpora can also be described as sample (or static) corpora or monitor corpora. Sample corpora are compiled once (whether diachronic or synchronic) and then no texts are added after the completion date, i.e., their size remains the same. The Brown and Frown corpora of American English and LOB/FLOB corpora of British English are good examples of sample corpora.¹⁵¹ Each of the four corpora contains approximately one million words from a variety of registers of written English, with texts for Brown and LOB having been published in 1961 and texts for Frown and FLOB in 1991. Monitor corpora, on the other hand, continually grow in size at regular intervals, allowing researchers to monitor language change across time periods in a dynamic manner. Examples include BYU's News on the Web (NOW) corpus—with five to six million words added from web-based newspapers and magazines every day¹⁵²—and the previously mentioned COCA, with twenty million words per each of five registers added every year.¹⁵³

148. See *Helsinki Corpus of English Texts*, VARIENG, <http://www.helsinki.fi/varieng/CoRD/corpora/HelsinkiCorpus/> (last updated Oct. 14, 2011).

149. See James C. Phillips, Daniel M. Ortner & Thomas R. Lee, *Corpus Linguistics & Original Public Meaning: A New Tool to Make Originalism More Empirical*, 126 YALE L.J.F. 21, 31 (2016).

150. See *The International Corpus of English*, UCL SURV. ENG. USAGE, <http://www.ucl.ac.uk/english-usage/projects/ice.htm> (last updated Dec. 1, 2016).

151. MCENERY ET AL., *supra* note 133, at 15.

152. See NOW CORPUS (NEWS ON WEB), <https://corpus.byu.edu/now/> (last visited Jan. 27, 2018).

153. See CORPUS CONTEMP. AM. ENG., <http://corpus.byu.edu/coca/> (last visited Jan. 27, 2018).

Corpus research over the past few decades has revealed 1) the extensive variation that exists in the use of features across different genres and registers;¹⁵⁴ 2) has highlighted the contextually-dependent nature of meaning, demonstrating how, instead of being located within individual words in a text, it is distributed within and across texts through prosodic collocations of features;¹⁵⁵ and 3) has emphasized the importance of testing intuitions about patterns of language structure and use within corpora. This is because:

each of us has only a partial knowledge of the language, we have prejudices and preferences, our memory is weak, our imagination is powerful (so we can conceive of possible contexts for the most implausible utterances), and we tend to notice unusual words or structures but often overlook ordinary ones.¹⁵⁶

This has led to two distinct but complementary approaches to the study of language structure and use: corpus-based and corpus-driven.¹⁵⁷ While both approaches are empirical and based on observations of large quantities of language data, corpus-based studies start with individual hypotheses about specific linguistic features and then examine how those features are distributed and function within a corpus; corpus-driven studies allow the linguistic features of interest to emerge from the corpus itself.¹⁵⁸ Thus, posing appropriate questions of the corpora, examining empirical evidence arising from the results, and integrating both complementary approaches results in a more well-rounded examination of language data that may otherwise be overlooked.¹⁵⁹ These issues are discussed in the case descriptions below.

154. See BIBER, CONRAD & REPPEN, *supra* note 132.

155. See Teubert, *supra* note 144, at 73.

156. Ramesh Krishnamurthy, *Collocation: From "Silly Ass" to Lexical Sets*, in WORDS IN CONTEXT: A TRIBUTE TO JOHN SINCLAIR ON HIS RETIREMENT 32–33 (Chris Heffer et al. eds., 2000).

157. See generally ELENA TOGNINI-BONELLI, CORPUS LINGUISTICS AT WORK (2001). For a critique of the division of these approaches, see MCENERY ET AL., *supra* note 133.

158. See Biber, *supra* note 131.

159. See Richard Xiao, *Theory-Driven Corpus Research: Using Corpora to Inform Aspect Theory*, in CORPUS LINGUISTICS: AN INTERNATIONAL HANDBOOK 987 (Anke Lüdeling & Merja Kytö eds., 2008).

B. Four Conditions that Make Corpus Analysis Efficacious in Legal Analysis

As we noted in the Introduction to this Article, the use of corpus analysis in legal argumentation is most efficacious when certain conditions hold.¹⁶⁰ In this section, we discuss those conditions in greater detail.

1. The legal issue in the case must concern the distribution of language usage over a particular population

We have already discussed the fact that ordinary meaning is not the universal standard in cases of statutory interpretation.¹⁶¹ Rather, it is a default rule that can be overridden when the particular circumstances of a case suggest the legislature had a broader or narrower meaning in mind. Different judges have different thresholds when it comes to determining when the ordinary meaning of a term should not predominate. As the sharp debates among the Justices reveal, the central issue is the extent to which a court investigates the circumstances surrounding a law's enactment, draws inferences as to what the law was intended to accomplish, and decides whether the "ordinary meaning" of the language in the law is either too narrow or too broad to accomplish those goals. The cases discussed above in section I.B illustrate the lack of consensus on these important methodological issues.

2. The court must adopt a clear notion of ordinary meaning

By the same token, when ordinary meaning does prevail, the court must decide what ordinary meaning means.¹⁶² There are many ways to address this question. Preliminarily, we wish to draw a distinction between two concepts of what makes meaning "ordinary."

Ordinary Meaning 1 ("OMI"): The ordinary meaning of a term is a description of the circumstances in which the term is most likely to be used.

160. See *supra* Intro.

161. See *supra* Section I.B.

162. See Lee & Mouritsen, *supra* note 1.

Ordinary Meaning 2 (“OM2”): The ordinary meaning of a term is a description of the circumstances in which members of a relevant speech community would express comfort in using the term to describe the circumstances. More than one meaning may be ordinary for a term under this theory.

We believe that each of these meanings of ordinary meaning may have its place in legal analysis. When a term has distinct senses, it makes sense to ask whether one of them predominates in ordinary usage, and an appropriate corpus provides a good tool for performing this comparison. OM1 invites this inquiry.

In contrast, how should courts view ordinary meaning when the issue in a case is whether a single sense of a word can be used in its ordinary sense in two distinct but related circumstances? When a term can be used in distinct but highly related circumstances, OM2 may tell us that both circumstances may be considered ordinary.

A disagreement between Justices Breyer and Ginsburg over these two notions of ordinary meaning took place in *Muscarello v. United States*.¹⁶³ The law in dispute called for an enhanced sentence for the perpetrator who “uses or carries a firearm” “during and in relation to” “a drug trafficking crime.”¹⁶⁴ Muscarello had drugs in one part of his car and a gun in another, as he drove to a drug deal he was planning to make.¹⁶⁵ The nine justices agreed that the issue in the case was whether he had “carried” a firearm in the ordinary sense,¹⁶⁶ but they disagreed by five to four on which meaning was the ordinary one. Writing for the majority, Justice Breyer held that Muscarello had done so.¹⁶⁷ Much of the decision¹⁶⁸ (and Justice Ginsburg’s dissent¹⁶⁹) focused on a battle over which dictionaries should be considered authoritative. Additional arguments were adduced based on etymology,¹⁷⁰ the use of *carry* one way or the other in eighteenth- and nineteenth-century literature¹⁷¹ and in the

163. *Muscarello v. United States*, 524 U.S. 125 (1998).

164. 18 U.S.C. § 924(c)(1) (2012).

165. *Muscarello*, 524 U.S. at 127.

166. *Id.* at 127–28, 139.

167. *Id.* at 125.

168. *Id.* at 128.

169. *Id.* at 142–43 (Ginsburg, J., dissenting).

170. *Id.* at 128 (majority opinion).

171. *Id.* at 129.

Bible,¹⁷² and the order of definitions in dictionaries, even the *Oxford English Dictionary*¹⁷³ in which the order of definitions is historical and not a matter of most common usage.¹⁷⁴ Based on this process, Breyer determined that the ordinary meaning of *carry a firearm* includes carrying the firearm in a vehicle and that carrying a gun on one's person is a "special" use of the term.¹⁷⁵

In addition, Breyer conducted a search of items found in Lexis and Westlaw news libraries, using the words *carry* and *weapon* in close proximity:

We looked for sentences in which the words "carry," "vehicle," and "weapon" (or variations thereof) all appear. We found thousands of such sentences, and random sampling suggests that many, perhaps more than one-third, are sentences used to convey the meaning at issue here, *i.e.*, the carrying of guns in a car.¹⁷⁶

Not surprisingly, Justice Ginsburg retorted in dissent: "One is left to wonder what meaning showed up some two-thirds of the time."¹⁷⁷

This exchange, short as it is, illustrates the lack of consensus about what makes ordinary meaning ordinary. Although Breyer's analysis may suggest that both meanings are acceptable, he remained committed to an OMI analysis, not acknowledging that the one-third rate of occurrence may suggest to others that more than one meaning is available. Ginsburg's response, without actually taking the position that the meaning most commonly instantiated should prevail, at least suggests that looking at the data more closely is a reasonable approach.¹⁷⁸

172. *Id.*

173. *Id.* at 128; see Mouritsen, *Dictionary*, *supra* note 1, at 1934.

174. See John Simpson, *Preface to the Third Edition of the OED*, OXFORD ENG. DICTIONARY (March 2000), <http://public.oed.com/the-oed-today/preface-to-the-third-edition-of-the-oed/>.

175. *Muscarello*, 524 U.S. at 128.

176. *Id.* at 129.

177. *Id.* at 143 (Ginsburg, J., dissenting).

178. *Id.*

3. *Conducting the right search on the right corpus*

There is a second problem with the use of the corpus to determine ordinary meaning in *Muscarello*. As Stephen Mouritsen points out, Breyer conducted the wrong search.¹⁷⁹ Ginsburg's quip in dissent misses this point as well. What matters is not the relative frequency with which *carry a weapon* suggests carrying it in a vehicle when the word *vehicle* is mentioned but rather the relative frequency with which *carry a weapon* suggests carrying it in a vehicle whether or not the word *vehicle* is mentioned. After all, the statute itself says nothing about vehicles.

Mouritsen conducted two sets of searches in COCA.¹⁸⁰ The first set of searches picked out only uses of the verb *to carry*.¹⁸¹ Mouritsen analyzed a sample of 500 of them.¹⁸² Many had nothing to do with the issues in *Muscarello*, with uses such as "my cable service doesn't carry CNBC."¹⁸³ Of those that involved a person carrying an object, 85% involved carrying something on one's person; the remaining 15% involved carrying something in some kind of vehicle or on an animal.¹⁸⁴

More pointedly, Mouritsen also looked at instances in which *carry* is used in proximity to words denoting various weapons: *firearm, gun, handgun, rifle, and pistol*.¹⁸⁵ He found that *carry* was clearly meant to describe carrying a weapon on one's person 64% of the time, clearly used to describe carrying the weapon in a vehicle 3% of the time, and either ambiguous between the two or referring to neither the rest of the time.¹⁸⁶ From these results, Mouritsen infers that Breyer was incorrect in asserting that carrying a gun on one's person embraces an unusual or "special" meaning of the term and that a broader meaning, at the very least, is included in the norm.¹⁸⁷

179. See Mouritsen, *Dictionary*, *supra* note 1, at 1948.

180. *Id.* at 1958.

181. *Id.*

182. *Id.*

183. See *id.*

184. *Id.* at 1961.

185. *Id.* at 1963.

186. *Id.* at 1965.

187. *Id.* at 1927–28.

We do not believe that this ends the matter, however. For just as law contains defeasible default rules, so does language. Linguists speak of “the unmarked case” as the default case.¹⁸⁸ This may reflect syntactic choices, as illustrated below.

“Bill likes to sip iced tea after a long day of construction work.”

“After a long day of construction work, Bill likes to sip iced tea.”

Although we are comfortable with both sentences, it makes sense to conclude that the default phrasal order in English has the temporal phrase coming after the main clause because the temporal phrase modifies the verb phrase semantically.

When it comes to meaning, we can say that the unmarked meaning of an expression is the way it would ordinarily be understood without further modification. Consider:

“In my state, it is legal to carry a weapon openly on campus.”

We believe that most people would understand this sentence as referring to carrying a weapon on one’s person in plain sight. Of course, one can modify the sentence:

“In my state, it is legal to carry a weapon in one’s automobile while driving on campus.”

Without this modification, we assume the default interpretation. For this reason, we share Mouritsen’s surprise that Justice Breyer’s search yielded only about one-third instances in which a weapon was being carried in a vehicle. We would have expected the modification of *carry* by *vehicle* to have yielded higher numbers.

This leads us to a final point: The ordinary way to carry a pistol may well be on one’s person, but the ordinary way to carry an army tank is, perhaps, in a cargo plane. And the ordinary way to carry a canon is in some other vehicle, the nature of which has changed over the centuries. All of this suggests a different analysis. Perhaps there are not two separate senses of *to carry* when used to describe people moving something from one place to another. Perhaps there is only a single sense, but as a default we assume that a sentence using the

188. See EDWARD FINEGAN, LANGUAGE: ITS STRUCTURE AND USE 196–97 (6th ed. 2012).

verb describes the most natural way of carrying whatever the direct object is unless a different mode is specified.

If this indeed is how we use the word *carry* in everyday speech, what does it say about corpus analysis as a means for resolving a legal dispute? If six members of a gang drive to a drug deal in a van loaded with large firearms and come out shooting, we would probably say that this statute applies to them. Muscarello's activities, in contrast, are less clear. We can certainly say that he carried the gun in his car. But if we simply say that he "carried a gun to a drug deal," we are most likely to assume that he had the gun on his person. The legislature could have clarified the statute by adding "on his person, in a vehicle, or otherwise," to *carry*, but it did not do so.

When it comes to weapons whose sizes permit carrying them either in a vehicle or on one's person, in most instances, the word *carry* is used to describe carrying the weapon on one's person. This is the result of an OM1 analysis,¹⁸⁹ although it comes out exactly the opposite of Justice Breyer's OM1 analysis. An OM2 analysis, in contrast, asks only if there are enough instances of both meanings to demonstrate that speakers are relatively comfortable using the term both ways, regardless of distributional facts between them. In about one-third of the occurrences in Mouritsen's analysis it is not possible to tell how the weapon was being carried.¹⁹⁰ Thus, it appears to us that people are generally comfortable using the verb *to carry* to denote carrying an object in a vehicle. As noted, however, for small objects, the default interpretation is that it is being carried on one's person.

This leaves a sufficient measure of uncertainty about meaning to trigger the rule of lenity, in our opinions. That rule resolves uncertainty in the interpretation of criminal statutes in favor of the defendant. From the language of the statute alone, it is not possible to determine whether Congress had in mind enhancing the sentence of those who carried a firearm with them into the actual drug transaction, or whether the law was intended to apply more broadly to those who brought firearms with them in their vehicles but left

189. See *supra* text accompanying note 152.

190. Mouritsen, *Dictionary*, *supra* note 1.

the weapons there during the transaction itself. Ginsburg relies on this principle, in part, in her dissent.¹⁹¹

In 2016, the Supreme Court of Michigan decided *People v. Harris*.¹⁹² The question in *Harris* was whether police officers who had testified falsely in a disciplinary hearing had provided “information” in that hearing. A Michigan law, the Disclosures by Law Enforcement Officers Act (DLEOA), immunizes information testified to in such contexts from use in a subsequent criminal prosecution.¹⁹³ The purpose of the law is to enable the state to compel the testimony of law enforcement officers in disciplinary proceedings without violating their constitutional right to refrain from providing self-incriminating testimony that can be used against them in a criminal case.¹⁹⁴

Three officers were present at a traffic stop; one of them assaulted the driver without adequate cause while the others watched.¹⁹⁵ The officers testified falsely in their disciplinary hearings, not knowing that someone had made a video recording of the entire incident.¹⁹⁶ The officers were subsequently prosecuted for obstruction of justice.¹⁹⁷ All seven justices on the court were comfortable using COCA to ascertain the ordinary meaning of *information*.¹⁹⁸ However, they divided four to three on the outcome of the case.¹⁹⁹ The disagreement among the justices arose from deciding which corpus analysis should be conducted. The majority correctly pointed out that *information* can be used with modifiers such as *false* and *inaccurate* to denote false statements.²⁰⁰ It held that the word *information* can be used to describe both truthful and false statements, making the officers’ false testimony inadmissible.²⁰¹ Objecting to the dissenters’ position that *information* when not

191. *Muscarello v. United States*, 524 U.S. 125, 150 (1998) (Ginsburg, J., dissenting).

192. *People v. Harris*, 885 N.W.2d 832 (Mich. 2016).

193. MICH. COMP. LAWS § 15.393 (2006).

194. *Harris*, 885 N.W.2d at 843.

195. *Id.* at 834.

196. *Id.* at 835.

197. *Id.*

198. *Id.* at 838–39.

199. *Id.* at 861 (Markman, J., concurring & dissenting).

200. *Id.* at 839 (majority opinion).

201. *Id.* at 843–44.

modified is almost always used to describe truthful statements, the majority remarked:

Empirical data from the COCA, however, demonstrates the opposite. In common usage, “information” is regularly used in conjunction with adjectives suggesting it may be both true and false. This strongly suggests that the unmodified word “information,” can describe *either* true or false statements. Moreover, by reading each identified use of the word “information” in its surrounding context, it is clear that “information” is often used to describe false statements. Quite simply, “information” in common parlance describes perceptions conveyed about the world around us, which may be true or false.²⁰²

But as the dissenting justices pointed out, a COCA search revealed that without modification, *information* is generally used to denote accurate information, rejecting the majority’s conclusion that the presence of veracity adjectives in both directions indicates that the unmodified form of the noun can be understood equally both ways.²⁰³

Making matters more uncertain, we do not know whether the legislature intended to use the term to cover the word regardless of modification, or whether the bare word, as it appears, was meant to be understood that way. In this sense, *Harris* resembles *Muscarello* in the statute’s lack of clarity.

In *United States v. Costello*,²⁰⁴ following his criticism of excessive reliance on dictionaries, Judge Posner used a Google News search to determine that a woman did not “harbor” her boyfriend, an undocumented immigrant, in violation of a federal statute, by allowing him to live with her, at least not if the court relied upon the ordinary meaning of the verb *to harbor*.²⁰⁵

Posner reported the results of a search he conducted, showing what words most frequently co-occur with *harbor* (called “collocates” in the language of corpus linguistics and

202. *Id.* at 839.

203. *Id.* at 850 n.14 (Markman, J., concurring & dissenting).

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

205. *Id.* at 1044, 1050.

lexicography).²⁰⁶ There were thousands of hits for harboring Jews, enemies, Quakers, refugees and fugitives, but few or none for harboring victims or guests.²⁰⁷ Posner concluded:

It is apparent from these results that “harboring,” as the word is actually used, has a connotation—which “sheltering,” and *a fortiori* “giving a person a place to stay”—does not, of deliberately safeguarding members of a specified group from the authorities, whether through concealment, movement to a safe location, or physical protection. This connotation enables one to see that the emergency staff at the hospital may not be “harboring” an alien when it renders emergency treatment even if he stays in the emergency room overnight, that giving a lift to a gas station to an alien with a flat tire may not be harboring, that driving an alien to the local office of the Department of Homeland Security to apply for an adjustment of status to that of lawful resident may not be harboring, that inviting an alien for a “one night stand” may not be attempted harboring, that placing an illegal alien in a school may not be harboring, and finally that allowing your boyfriend to live with you may not be harboring, even if you know he shouldn’t be in the United States.²⁰⁸

Lee and Mouritsen criticize Posner’s use of Google in part because Google does not publish its algorithm for obtaining hits and is therefore not transparent.²⁰⁹ Moreover, Google searches only contemporaneous web pages and there is no basis for concluding that counting appearances on the web accurately reflects ordinary meaning in broader contexts.²¹⁰ On the plus side, Posner’s search comports with linguistic literature showing that some expressions have semantic prosody, that is positive or negative connotations that do not typically make their way into dictionary definitions.²¹¹ In our opinions, while the criticism of Posner’s choice of corpus is valid, his

206. *Id.* at 1044.

207. *Id.*

208. *Id.* at 1044–45 (citations omitted).

209. *See* Lee and Mouritsen, *supra* note 1.

210. *See id.*

211. Michael Stubbs, *Quantitative Data on Multi-Word Sequences in English: The Case of the Word World*, in *TEXT, DISCOURSE AND CORPORA: THEORY AND ANALYSIS* 163 (Michael Hoey et al. eds., 2007).

analysis is sufficiently robust as to put the burden on critics to argue that a better corpus analysis would yield different results.

4. Caution in drawing inferences from the absence of particular usages in a corpus: the merits of double dissociation

When a particular usage appears frequently in a corpus, it is reasonable to infer that the speakers or writers whose language has been recorded are comfortable using the term in the way the corpus instantiates. In contrast, as noted in the introduction to this Article, the absence of data in the corpus may reflect a linguistic fact—a usage is absent because people do not use the term that way—or a fact about the world—people may express a concept in a way not present in the corpus but just do not talk about such a thing much, and for that reason, it does not show up in the corpus.

Consider first the classic 1892 case, *Church of the Holy Trinity v. United States*.²¹² A law made it illegal to:

[I]n any manner whatsoever, to prepay the transportation, or in any way assist or encourage the importation or migration, of any alien or aliens, any foreigner or foreigners, into the United States, . . . under contract or agreement, . . . made previous to the importation or migration of such alien or aliens, foreigner or foreigners, to perform labor or service of any kind in the United States²¹³

A wealthy church in Manhattan had paid for the transportation of its new rector from England to New York and was prosecuted for having violated the law.²¹⁴ The Supreme Court ruled unanimously that the statute did not apply to these facts.²¹⁵ This case has remained controversial for 125 years since its publication in part because of its reliance on legislative history to determine the intent of the legislature²¹⁶ and in part because of its reliance on “the spirit” of the

212. *Church of the Holy Trinity v. United States*, 143 U.S. 457 (1892).

213. *Id.* at 458.

214. *Id.* at 457–58.

215. *Id.* at 472.

216. Carol Chomsky, *Unlocking the Mysteries of Holy Trinity: Spirit, Letter, and History in Statutory Interpretation*, 100 COLUM. L. REV. 901, 905 (2000).

statute.²¹⁷ What often goes unnoticed in discussions of this case is that, in many respects, it is a paradigmatic case about ordinary meaning. The title of the act was: “An act to prohibit the importation and migration of foreigners and aliens under contract or agreement to perform labor in the United States, its Territories and the District of Columbia.”²¹⁸ Taking advantage of the fact that the title does not mention performing service, whereas the body of the statute does, Justice Brewer remarked in his opinion:

Obviously the thought expressed in this reaches only to the work of the manual laborer, as distinguished from that of the professional man. No one reading such a title would suppose that Congress had in its mind any purpose of staying the coming into this country of ministers of the gospel, or, indeed, of any class whose toil is that of the brain. The common understanding of the terms labor and laborers does not include preaching and preachers; and it is to be assumed that words and phrases are used in their ordinary meaning. So whatever of light is thrown upon the statute by the language of the title indicates an exclusion from its penal provisions of all contracts for the employment of ministers, rectors, and pastors.²¹⁹

Thus, putting the spirit of the law aside, Brewer argued that it is unnatural to consider members of the clergy as performing “labor,” even though it is possible to acknowledge that they do so by using an expansive concept of what constitutes “labor.” In many ways, Brewer’s argument in *Holy Trinity Church* presages Scalia’s ordinary meaning argument in his dissenting opinion in *Chisom v. Roemer*,²²⁰ the Voting Rights Act case discussed above.

Now let us return to *Smith v. United States*,²²¹ the case in which a majority of six Supreme Court justices ruled that a person who traded an unloaded machine gun for cocaine was subject to a harsher

217. See ANTONIN SCALIA, A MATTER OF INTERPRETATION 19–20 (1998). The offending sentence in the opinion is “It is a familiar rule that a thing may be within the letter of the statute and yet not within the statute because not within its spirit nor within the intention of its makers.” *Church of the Holy Trinity*, 143 U.S. at 459.

218. *Church of the Holy Trinity*, 143 U.S. at 463.

219. *Id.*

220. *Chisom v. Roemer*, 501 U.S. 380 (1991).

221. *Smith v. United States*, 508 U.S. 223 (1993).

sentence by virtue of having “used a firearm during and in relation to a drug trafficking crime.”²²² Scalia’s colorful dissent ridiculed the majority for confusing *ordinary* meaning with some *possible* meaning. Scalia was right that when we think of using a gun we think of using it as a weapon.²²³ No doubt a COCA search would show many more instances of *use* and *gun* collocating when the gun was being used as a weapon than as an object of value to be bartered.

But the majority was also right to say that, given what Smith did, it does not sound strange to describe his activity as using a gun even if it is an unusual way to use a gun. That is, Scalia engaged in an OM1 analysis while the majority engaged in an OM2 analysis. In cases like *Chisolm* and *Church of the Holy Trinity*, they will produce the same result. In both of those cases, the reason that *representative* and *labor* are used principally, as the Court describes, is because the relevant terms are not generally used to convey the alternative meaning (judge as *representative* or clergyman as *laborer*).

We believe that the most efficient way to demonstrate that a corpus analysis is uncovering a linguistic regularity, rather than the mere fact that some things are spoken about less frequently than are others, is to demonstrate that the circumstances described by the infrequently used term are present in the corpus but spoken about differently. For example, if there had been a dispute about whether judicial elections are within the ordinary meaning of elections of “representatives” in *Chisolm*, it would be possible to show not only that *election*, *judge*, and *representative* do not show up together in the corpus with any regularity but also that judicial elections are described using different language when discussed.

The argument is akin to the double dissociation sometimes used in brain studies involving language deficits.²²⁴ For example, we know that damage to the Broca’s area of the brain affects the ability to

222. 18 U.S.C. § 924(c)(1) (2012). The same statute was at issue in *Muscarello*. The law makes it a crime to “use or carry” a firearm in the context of a drug trafficking crime.

223. See *Chisolm*, 501 U.S. at 404 (Scalia, J., dissenting). Scalia’s dissent was also more like the unanimous decision in *Church of the Holy Trinity* than he would have wished. Both employed strong arguments based on ordinary usage. For further discussion, see SOLAN, *supra* note 14.

224. For seminal work on double dissociation, see A.R. LURIA, RESTORATION OF FUNCTION AFTER BRAIN INJURY (O.L. Zangwill ed., Basil Haigh trans., 1963).

speaking fluidly but does not affect comprehension. Damage to the Wernicke's area of the brain does just the opposite.²²⁵ Correspondingly, the strongest corpus arguments occur when not only one usage predominates over another in the corpus but when, in addition, the circumstances describable in the infrequently instantiated case are present in the corpus but described using other language. Using a combination of corpus-based and corpus-driven approaches will help reveal such variation in use.

We similarly caution against inferring that a prototypical use of an expression is necessarily the most frequently occurring usage.²²⁶ The giraffe is the prototypical tall animal, not because we speak more of giraffes in that regard than we do, say, of camels, but because giraffes have the essential feature of tallness to an unusual extent among animals. Double dissociation can serve to reduce the likelihood of this false inference as well.

C. Getting it Right

Let us now turn to a case in which the four conditions are met. Here, we focus on Associate Chief Justice Lee's concurring opinion in the Utah Supreme Court case, *State v. Rasabout*,²²⁷ which, like *Harris*, relied on a corpus analysis using COCA. Justice Lee had been a full-time law professor at BYU before his appointment to the bench and is familiar with how to conduct a corpus analysis. The legal issue, as the justices taking different positions agreed, was the ordinary meaning of the word *discharge*.²²⁸

Andy Rasabout was a gang member. He fired twelve rounds from a semiautomatic weapon at a house and car as he drove by.²²⁹ A jury convicted him of twelve separate offenses; however, the trial court merged them into one count of "discharg[ing] a dangerous

225. LANGUAGE FILES: MATERIALS FOR AN INTRODUCTION TO LANGUAGE AND LINGUISTICS 360 (Anouschka Bergmann et al. eds., 10th ed. 2007).

226. Carissa Hessick also makes this point in her contribution to this volume. Carissa Byrne Hessick, *Corpus Linguistics and the Criminal Law*, 2017 BYU L. REV. ____; see Elizabeth B. Lynch, John D. Coley & Douglas L. Medin, *Tall is Typical: Central Tendency, Ideal Dimensions, and Graded Category Structure Among Tree Experts*, 28 MEMORY & COGNITION 41 (2000).

227. *State v. Rasabout*, 2015 UT 72, 356 P.3d 1258.

228. *Id.* ¶¶ 5–6, 356 P.3d at 1261.

229. *Id.*

weapon or firearm from an automobile . . . without written permission, within 600 feet of a house, dwelling or other building.”²³⁰ The appellate court reversed, reinstating the jury verdict,²³¹ and the Supreme Court of Utah affirmed the appellate court’s decision.

The majority opinion, written by Justice Parrish, held that “the allowable unit of prosecution for the unlawful-discharge-of-a-firearm statute . . . is each discrete shot.”²³² In so holding, the court relied on the morphology of the word discharge (dis + charge), various dictionaries, the statutory definition of *firearm* (“any device . . . from which is expelled a projectile by action of *an* explosive”),²³³ the purpose of the statute, and the whole act rule.²³⁴

Associate Chief Justice Thomas Lee concurred, applying corpus linguistic methodology to the analysis.²³⁵ He remarked:

I would interpret the terms of the statute by looking for real-world examples of its key words in actual written language in its native context. This sort of analysis has a fancy name—*corpus linguistics*. But it is hardly unusual. We often resolve problems of ambiguity by thinking of examples of the use of a given word or phrase in a particular linguistic context. I propose to do that (as I have in a couple of prior opinions) on a systematic scale—by computer-aided searches of online databases in an effort to assemble a greater number of examples than I can summon by memory on my own.²³⁶

Justice Lee’s COCA search found a substantial predominance of examples in which *discharge* was used to mean “to shoot” in the sense of the majority opinion.²³⁷ Justice Lee therefore concluded that the ordinary meaning of the verb is “to shoot,” even though it has other meanings, and concurred in the judgment based largely on his corpus analysis. In doing so, he noted that the majority had ignored

230. UTAH CODE ANN. § 76-10-508 (West 2016).

231. *Rasabout*, 2015 UT 72, ¶ 1, 356 P.3d at 1261.

232. *Id.* ¶ 15, 356 P.3d at 1263.

233. *Id.* ¶¶ 12–13, 356 P.3d at 1263–64.

234. *Id.*

235. *See id.* ¶ 40, 356 P.3d at 1271 (Lee, J., concurring in part and concurring in judgment).

236. *Id.* ¶ 41, 356 P.3d at 1271.

237. *Id.* ¶¶ 81–93, 356 P.3d at 1281–82.

the second definition (which would have assisted the defendant's case).²³⁸ Rather, the majority had relied on its own intuitions about which meaning is ordinary, without adequate inquiry. Instead, he argued, the court should take on the burden of acknowledging both definitions and coming to a reasoned decision based on empirical investigation as to which one reflects the verb's most ordinary sense.²³⁹

Our only quibble with Justice Lee's analysis is that his argument would have been even stronger if he had double dissociated by showing that COCA contains examples of the verb *to empty* to express the meaning of *discharge* proposed by the defendant. Our own search shows this to be the case. COCA contains five instances of emptying a weapon and thirty-three of emptying a gun. Almost all of these were used to express the meaning of *discharge* that means emptying a gun completely.²⁴⁰

IV. CONCLUSION

We have attempted in this Article to demonstrate the utility of corpus linguistics in the interpretation of laws and to establish criteria for the efficacy of this practice. We recognize that this Article shifts between having a supportive tone and having a cautionary tone to this endeavor. At all times, it should be kept in mind that linguistic corpora provide a tool for those engaged in statutory interpretation, but they say nothing about when this tool is most useful. That judgment must come from a combination of legal decisions about whether the distribution of word usage should determine the outcome of a case, a further legal decision about what "ordinary meaning" should mean in a legal context, and appropriate

238. *Id.* ¶ 88, 356 P.3d at 1282.

239. Justice Lee's analysis provoked a response from the majority, criticizing him for engaging in expert analysis outside the bounds of the adversarial system. The issue was not the legitimacy of his argument, but the legitimacy of judges taking the initiative to engage in scientific investigation without the parties being afforded the opportunity to cross-examine or otherwise challenge the analysis. *Id.* ¶¶ 16–21, 356 P.3d at 1264–66 (majority opinion). Because COCA is publicly available and the results of searches transparent, we are not as bothered by Justice Lee's foray as was the majority.

240. Search conducted on August 4, 2017. For example: "The teenager collapses. Van Dyke continues shooting, emptying his weapon. In all, McDonald was shot 16 times." "And then while she was lying dead on the floor, he emptied the gun into her."

use of a corpus when one is used. This last point includes asking the right questions, conducting the right searches, and drawing valid inferences from both the presence and absence of data reflecting one or another specific usage.

