Corpus Linguistics and the Dream of Objectivity

Evan C. Zoldan[*]

A growing number of scholars and judges have embraced corpus linguistics as a way to interpret legal texts. Their stated goal—to make legal interpretation more objective—is an admirable one. But, is their claim that corpus linguistics reduces the subjectivity associated with judicial intuition and biased data more than just a dream? After analyzing the way that corpus linguistics is used to interpret statutes, this Article concludes that corpus linguistics does not live up to its promise to make legal interpretation more objective. Instead, the use of corpus linguistics to interpret statutes results in interpretations that are radically acontextual, disrupting its proponents' dream of objectivity.

I. INTRODUCTION

The notion of objectivity in legal interpretation carries great rhetorical force. It evokes a world in which judges are neutral, dispassionate, and all but invisible in the contests they judge.[1] In a well-known statement made during his Senate confirmation hearing, then-Judge John Roberts invoked this vision of objectivity when he declared that "Judges are like umpires. Umpires don't make the rules, they apply them . . . .   [A]nd I will remember that it's my job to call balls and strikes, and not to pitch or bat."[2] Not all legal thinkers express such optimism about the prospect that a legal interpreter can remain separated from an interpretation. More modestly,

---

[1]   *See* Charles Fried, *Balls and Strikes*, 61 EMORY L.J. 641 (2012).

[2]   *Confirmation Hearing on the Nomination of John G. Roberts, Jr. to be Chief Justice of the United States: Hearing Before the S. Comm. on the Judiciary,* 109th Cong. 55–56 (2005) (statement of John G. Roberts).

Professor Kent Greenawalt suggests, it is inevitable that the interpreter's perspective will be incorporated into an interpretation; nevertheless, a community's shared language and culture means that "questions about meaning can often be answered with confidence, that an answer can be objectively right or wrong."[3] For Professor Owen Fiss, objectivity connotes standards, implying that "an interpretation can be measured against a set of norms that transcend the particular vantage point of the person offering the interpretation."[4]

These visions of objectivity differ; but in each one, interpretive objectivity means not merely signposts for legal interpreters, but also fences to corral their interpretations. It is the desire to *constrain* the interpreter that is the dream of objectivity: an objective interpretation reduces the role of the legal interpreter in the interpretive process by limiting the range of permissible interpretations.[5]

It is with this dream of objectivity that some legal interpreters have embraced corpus linguistics methods to interpret legal texts. Corpus linguistics, a methodology or set of tools for studying language data in bodies of text,[6] is not a new discipline. What is new is the use of corpus linguistics to interpret legal texts, like statutes and the Constitution. Asserting that statutory interpretation often calls on the interpreter to find the "ordinary meaning" of a text, users of corpus linguistics methods seek meaning in bodies of text, called corpora, that they claim reflect the ordinary usage of those words. Although some corpus users acknowledge the role that the interpreter's judgment plays in interpretation, the dream of objectivity is the rallying cry of corpus linguistics for legal interpretation. Indeed, corpus users frame the utility of corpus linguistics techniques in terms of a critique of other methods of interpretation, which are often derided as "simple cherry-picking,"[7] "subjective,"[8] or "idiosyncratic."[9] In

[3]   KENT GREENAWALT, LAW AND OBJECTIVITY 74 (1992).

[4]   Owen Fiss, *Objectivity and Interpretation*, 34 STAN. L. REV. 739, 744–45 (1982). *See also* RONALD DWORKIN, LAW'S EMPIRE vii–ix (1986) (insisting that "in most hard cases there are right answers," even if those right answers cannot be proved to be correct); *id.* at 255–56 (arguing that a community's history and political principles are objective facts that provide outer bounds on the subjectivity of legal interpretation); Ronald Dworkin, *No Right Answer?*, 53 N.Y.U. L. REV. 1, 30–31 (1978).

[5]   Ofer Raban, *Between Formalism and Conservatism: The Resurgent Legal Formalism of the Roberts Court*, 8 N.Y.U. J.L. & LIBERTY 343, 368 (2014).

[6]   GRAEME KENNEDY, AN INTRODUCTION TO CORPUS LINGUISTICS 1 (2014). Krzysztof Kredens & Malcolm Coulthard, *Corpus Linguistics in Authorship Identification*, *in* THE OXFORD HANDBOOK OF LANGUAGE AND LAW 504–05 (Peter M. Tiersma & Lawrence M. Solan eds., 2016) (noting different definitions of "corpus" and "corpus linguistics").

[7]   Thomas R. Lee & Stephen C. Mouritsen, *Judging Ordinary Meaning*, 127 YALE L.J. 807 (2018).

[8]   James C. Phillips, Daniel M. Ortner & Thomas R. Lee, *Corpus Linguistics & Original Public Meaning: A New Tool to Make Originalism More Empirical*, 126 Yale

contrast, some corpus users claim that corpus linguistics can "help us deliver on the promise of an objective inquiry."[10]

Corpus users' criticisms of other methods of interpretation have been persuasive to a small but growing group of judges who have adopted, to various degrees, corpus linguistics methods in their opinions. These methods have been used most consistently, and explained most thoroughly, by the Supreme Court of Utah's Justice Lee, an early judicial adopter of corpus methods.[11] In the first adoption by a state high court, the Supreme Court of Michigan relied explicitly on corpus linguistics methods to interpret a state statute.[12] Importantly, corpus linguistics methods have spread from state courts to federal courts. A federal district court relied in part on corpus methods when interpreting a term of the Federal Credit Union Act, using data from a corpus search alongside a Westlaw search, dictionaries, and canons of construction.[13] In a pair of Court of Appeals opinions interpreting provisions of ERISA, judges of the Third Circuit[14] and the Sixth Circuit[15] have relied on corpus data as well. And most notably, Justice Thomas cited data returned by a corpus search in his dissent in *Carpenter v. United States*.[16] So, too, has the scholarly community begun to take notice. Building on the recent burst of judicial uses of corpus linguistics, legal scholarship has begun working out the possibilities of this new methodology.[17] All of this suggests that corpus

L.J.F. 21, 23–24 (2016) (arguing that corpus linguistics sometimes can "rescue" originalism from subjectivity). *See also* Stephen C. Mouritsen, *Hard Cases and Hard Data*, 13 COLUM. SCI. & TECH. L. REV. 156, 202 (2011) ("Thus, the corpus method embodies the lexicographer's proud ideal of descriptive objectivity; his citations (and interpretations of them) are publicly verifiable.") (internal citations omitted).

[9] Stefan Th. Gries & Brian G. Slocum, *Ordinary Meaning and Corpus Linguistics*, 2017 BYU L. REV. 1417, 1441 (2017).

[10] Lee & Mouritsen, *supra* note 7, at 796.

[11] State v. Rasabout, 356 P.3d 1258, 1274 (Utah 2015) (Lee, J., concurring); *In re* Adoption of Baby E.Z., 266 P.3d 702, 725–27 (Utah 2011) (Lee, J., concurring).

[12] People v. Harris, 885 N.W.2d 832, 838–39 (Mich. 2016).

[13] Am. Bankers Ass'n v. Nat'l Credit Union Admin., 306 F. Supp. 3d 44, 68 (D.D.C. 2018).

[14] Caesars Entm't Corp. v. Int'l Union of Operating Eng'rs, 932 F.3d 91, 95 (3d Cir. 2019).

[15] Wilson v. Safelite Grp., Inc., 930 F.3d 429, 439, 440 (6th Cir. 2019) (Thapar, J., concurring).

[16] Carpenter v. United States, 138 S. Ct. 2206, 2238 (2018) (Thomas, J., dissenting).

[17] *E.g.*, Anya Bernstein, *Democratizing Interpretation*, 60 WM. & MARY L. REV. 435 (2018); Clark D. Cunningham & Jesse Egbert, *Scientific Methods for Analyzing Original Meaning*: *Corpus Linguistics and the Emoluments Clause* (Ga. St. Univ. Coll. Law, Legal Studies Research Paper No. 2019-02, 2019), https://papers.ssrn.com/sol3/papers.cfm?abstract id=3321438; J.S. Ehrett, *Against Corpus Linguistics*, 108 GEO. L. J. ONLINE (2019); Gries & Slocum, *supra* note 9, at 1434; Ethan Herenstein, *The Faulty Frequency Hypothesis*, 70 STAN. L. REV. ONLINE 112, 116-119 (2017); Carissa Byrne Hessick, *Corpus Linguistics and*

linguistics methods are increasingly relevant to interpretive theory. As a result, the time has come to examine whether corpus linguistics can deliver on its promise of interpretive constraints.

After examining the theory and practice of corpus use for the purposes of statutory interpretation, this Article concludes that, despite its allure, corpus linguistics does not bring legal interpretation closer to fulfilling the dream of objectivity. Part II introduces the discipline of corpus linguistics and describes how it is used by legal scholars and judges to interpret statutes. Part III sets out the case for objectivity made by proponents of corpus linguistics for statutory interpretation. It first describes corpus users' criticisms of the subjectivity of other methods of interpretation, specifically, the subjectivity that flows from the interpreter's reliance on her own intuition and on biased data. Next, it describes the ways in which some corpus users believe that corpus linguistics techniques reduce the subjectivity associated with intuition and biased data.

Part IV analyzes corpus linguistics methods to determine whether they do, in fact, reduce the subjectivity of the interpretive process. Despite the claims of its proponents, corpus users introduce subjectivity into the interpretive process at the moment they choose a corpus to search. This subjective decision determines the final interpretive outcome, disrupting corpus users' dream of objectivity. Part V compares the subjectivity of corpus linguistics methods to the subjectivity corpus users criticize in other methods of interpretation. Strikingly, the subjectivity of corpus use rests also on the interpreter's intuition and on biased data, undermining the claim that corpus linguistics is superior to other methods of interpretation.

This Article makes three main contributions to the literature on corpus linguistics and legal interpretation. First, this Article contests the central claim made by users of corpus linguistics for statutory interpretation,

---

the Criminal Law, 2017 BYU L. REV. 1503, 1508–09 (2017); Jennifer L. Mascott, *The Dictionary as a Specialized Corpus*, 2017 BYU L. REV. 1557 (2017) [hereinafter Mascott, *Dictionary*]; Jennifer L. Mascott, *Who Are Officers of the United States?*, 70 STAN. L. REV. 443, 496 (2018) [hereinafter Mascott, *Officers*]; Phillips et al., *supra* note 8, at 24–26; James Cleith Phillips & Sara White, *The Meaning of the Three Emoluments Clauses in the U.S. Constitution*, 59 S. TEX. L. REV. 181 (2019); Brian G. Slocum, *Ordinary Meaning and Empiricism*, 40 STAT. L. Rev. 13 (2019); Lawrence M. Solan, *Legal Linguistics in the U.S.: Looking Back, Looking Ahead* (Brook. L. Sch., Legal Studies Paper No. 609, 2019), https://papers.ssrn.com/sol3/papers/cfm?abstract_id=3428489; Lawrence M. Solan & Tammy Gales, *Corpus Linguistics as a Tool in Legal Interpretation*, 2017 BYU L. REV. 1311, 1332; Lawrence B. Solum, *Intellectual History as Constitutional Theory*, 101 VA. L. REV. 1111, 1135–36 (2015) (arguing that corpus linguistics is appropriate for the investigation of semantic meaning of constitutional text because the Constitution, for historical reasons, is "directed to the public at large"); Lee J. Strang, *How Big Data Increases Originalism's Methodological Rigor: Using Corpus Linguistics to Recover Original Language Conventions*, 50 U.C. DAVIS L. REV. 1181 (2017) (suggesting extra-textual analysis to determine parameters of corpus search).

namely, that it constrains legal interpretation in a way that reduces subjectivity in the interpretive process. By focusing on the subjective choice of corpus that users of corpus methods must make, this Article concludes that corpus use is no more objective than the methods of interpretation criticized by corpus users themselves.

Second, previous work on corpus linguistics methods generally do not distinguish between the use of these methods for statutory interpretation and their use in constitutional interpretation. This Article focuses on statutory interpretation alone, analyzing corpus methods in light of the legislative process and the unique nature of statutory language. In particular, this Article describes the significant differences between statutory language and nonlegal language. It concludes that searching for legal meaning in a corpus of nonlegal language is particularly inappropriate for statutory interpretation. Rather than simply serving as another "tool in the toolbox" of statutory interpretation, corpus linguistics is different from traditional tools of statutory interpretation because it leads to interpretations that are radically acontextual.[18]

Third, previous work on corpus linguistics and interpretation has suggested that the construction of a legal corpus might cure the defects that flow from the mismatch between statutory language and nonlegal texts.

---

[18] This Article does not address the use of corpus linguistics techniques in general, nor does it suggest that corpus linguistics techniques are inappropriate in disciplines outside of the context of statutory interpretation. Moreover, this Article does not address the use of corpus linguistics for constitutional interpretation. Nevertheless, much of the work here might be applicable to the question of constitutional interpretation as well. McGinnis and Rappaport have argued that the Constitution is written in legal language, much as I argue here that statutes are written in legal language. John O. McGinnis & Michael B. Rappaport, *The Constitution and the Language of the Law*, 59 WM. & MARY L. REV. 1321 (2018). By contrast, other scholars have suggested that at least some constitutional terms are written in ordinary language. Solum, *supra* note 17, at 1136. In her work interpreting the term "Officers of the United States," for example, Jennifer Mascott defended the decision to search for the term "officer" as an ordinary term after first conducting an extensive extra-textual analysis to determine whether the constitutional term was an ordinary term rather than a term of art. Mascott, *Officers*, *supra* note 17, at 471–72. Mascott looked not only to other clauses of the Constitution, but also to the Constitution's drafting history, founding-era debates, and preratification history. Mascott, *Officers*, *supra* note 17, at 471–79. Mascott's process follows the elaborate approach set out by Lee Strang, who suggested that corpus research should normally be done only after parameters for the search are established through the study of sources external to the text under consideration. Strang, *supra* note 17, at 1208–09. Strang suggests, for example, that a "stable of possible language conventions" can be gathered by searching case law, scholarship, and primary and secondary sources. Strang, *supra* note 17, at 1207. *See also* John Mikhail, *The Definition of Emolument in English Language and Legal Dictionaries, 1523–1806* (forthcoming) (on file with Georgetown Univ. Law Center). I take no position at this time whether any part of the Constitution can be interpreted as if it were ordinary language. But, to the extent that constitutional language is legal language, the arguments made in this Article also suggest the inappropriateness of searching in a corpus of nonlegal language for constitutional meaning.

This Article anticipates different types of legal corpora that might be constructed and demonstrates practical and theoretical difficulties that make the objective use of a legal corpus for statutory interpretation unlikely.

II. CORPUS LINGUISTICS AND STATUTORY INTERPRETATION

Corpus linguistics is a methodology or set of tools for studying language data in bodies of text.[19]   Although the study of linguistic information found in collections of text is not a new activity,[20] corpus linguistics has taken on new importance as electronic storage and retrieval systems allow users to search bodies of text, or corpora, that cannot feasibly be searched manually.[21] This wealth of electronic data has opened up new possibilities for researchers to find evidence of patterns of language use.[22] Corpus linguistics techniques are often used to determine how frequently words are used, in what context they are used, and, when a word has multiple shades of meaning, or senses, what other words tend to collocate with each of a word's senses.[23] The word "deal," for example, has multiple senses, including to solve a problem ("deal with the situation"), to cope ("deal with the tragedy"), and to engage in business interactions ("deal with the supplier").[24] Analyzing how "deal" collocates with the words around it can help an interpreter determine which sense of deal is meant in a particular text.[25]

Corpus linguistics methods have proved enticing to legal interpreters, some of whom have touted the ability of corpus analysis to minimize subjectivity in textual interpretation.[26] A growing number of scholars and

---

[19]   KENNEDY, *supra* note 6, at 1; Kredens & Coulthard, *supra* note 6, at 504–05 (noting competing and overlapping definitions of corpus linguistics); TONY MCENERY & ANDREW HARDIE, CORPUS LINGUISTICS: METHOD, THEORY, AND PRACTICE 1–2 (2012) (noting that corpus linguistics is still a rapidly developing field and that, as a result, its methods and definitions are contested).

[20]   KENNEDY, *supra* note 6, at 1; DOUGLAS BIBER ET AL., CORPUS LINGUISTICS: INVESTIGATING LANGUAGE STRUCTURE AND USE 21–22 (1998).

[21]   BIBER ET AL., *supra* note 20, at 22–23; KENNEDY, *supra* note 6, at 1; Peter M. Tiersma & Lawrence M. Solan, *Introduction in* HANDBOOK OF LANGUAGE AND LAW 3 (2012).

[22]   BIBER ET AL., *supra* note 20, at 1–2.

[23]   *Id*. at at 23–25.

[24]   *Id*. at 42.

[25]   *Id*. at 42–43. Importantly, however, linguists recognize that there is not a perfect correlation between collocation and usage of a word. Rather, a "single collocation can be used with a range of senses." BIBER ET AL., *supra* note 20, at 43. *See also* George Miller, *Contextuality*, *in* MENTAL MODELS IN COGNITIVE SCIENCE 2–3 (1996) (discussing polysemy).

[26]   State v. Rasabout, 356 P.3d 1258, 1274–75 (Utah 2015) (Lee, J., concurring) (arguing that corpus linguistics is intended to ameliorate judicial intuition); Utah v. J.M.S.,

judges have turned to corpus linguistics techniques when searching for the original public meaning of the text of the Constitution[27] or the ordinary meaning of statutes.[28]   Although there are a number of corpora with different sets of texts, users of corpus linguistics methods for statutory interpretation rely exclusively, or virtually so, on general corpora, that is, corpora that contain a balance of different types of texts that are meant to approximate ordinary speech.[29] For example, the Corpus of Contemporary American English (COCA) contains more than 560 million words that come from a variety of sources, including transcriptions of spoken language, fiction, popular magazines, newspaper articles, and academic works.[30]

     A recent case from the Supreme Court of Michigan is typical of the way legal interpreters have relied on corpus linguistics techniques.   In *Harris*, a statute prohibited the state from using a law enforcement officer's disclosure of "any information" in a subsequent criminal proceeding against the officer.[31] Three officers made false statements in the course of an internal investigation into their conduct. The state used these false statements in a criminal proceeding against the officers on the theory that "information" pertains only to accurate statements and, as a result, false statements do not fall within the ambit of the statute.[32] The court rejected this argument, relying, in part, on data derived from the COCA.

---

280 P.3d 410, 419 n.3 (Utah 2011) (Lee, J., concurring) (arguing that corpus linguistics is used to check imperfect judicial intuition); Mouritsen, *supra* note 8, at 175–78 (arguing that judges tend to be idiosyncratic in their intuition about ordinary usage and therefore cannot determine ordinary usage intuitively); D. Carolina Nuñez, *War of the Words: Aliens, Immigrants, Citizens, and the Language of Exclusion*, 2014 BYU L. REV. 1517, 1521 (arguing that corpus linguistics is a more reliable guide to language use than native speaker intuition); Phillips & White, *supra* note 17, at 182–83.

    [27]   Carpenter v. United States, 138 S. Ct. 2206, 2238 (Thomas, J., dissenting); Mascott, *Dictionary*, *supra* note 17, at 1557; Mascott, *Officers*, *supra* note 17, at 496; Phillips, Ortner & Lee, *supra* note 8, at 24–26; Phillips & White, *supra* note 17, at 183; Strang, *supra* note 17, at 1181.

    [28]   Caesars Entm't Corp. v. Int'l Union of Operating Eng'rs, 932 F.3d 91, 95 (3d Cir. 2019); Wilson v. Safelite Grp., Inc., 930 F.3d 429, 440 (6th Cir. 2019) (Thapar, J., concurring); Am. Bankers Ass'n v. Nat'l Credit Union Admin., 306 F. Supp. 3d 44, 68 (D.D.C. 2018); People v. Harris, 885 N.W.2d 832 (Mich. 2016); *Rasabout*, 356 P.3d at 1274 (Lee, J., concurring); *In re:* Adoption of Baby E.Z., 266 P.3d 702, 724 n.23 (Utah 2011) (Lee, J., concurring); Stephen Mouritsen, *The Dictionary Is Not a Fortress*, 2010 BYU L. REV. 1915, 1956–57 (2010); Daniel Ortner, *The Merciful Corpus: The Rule of Lenity, Ambiguity, and Corpus Linguistics*, 25 B.U. PUB. INT. L.J. 101, 123 (2016).

    [29]   *See e.g.*, Craig v. Provo City, 389 P.3d 423, 428 n.3 (Utah 2016); *Baby E.Z.*, 266 P.3d at 724 n.23 (Lee, J. concurring in part); Mouritsen, *supra* note 28, at 1956; Ortner, *supra* note 28, at 128–29.

    [30]   CORPUS CONTEMPORARY AM. ENG., https://www.english-corpora.org/coca/ (last visited Sept. 21, 2019).

    [31]   *Harris*, 885 N.W.2d at 833.

    [32]   *Id.* at 835.

Examining the COCA to determine whether the word "information" is commonly collocated with words denoting truth or falsity, the court found that the word information is often used in close proximity with words that denote both truth and falsity, like "accurate," and "inaccurate."[33] The court concluded that the ordinary meaning of information, as used in the statute, includes both true and false information.[34] As a result, the court interpreted the statute to prohibit the use of even false statements against the officers.

III. CORPUS LINGUISTICS AND THE DREAM OF OBJECTIVITY

Although many users of corpus linguistics acknowledge that the elimination of the judge from the interpretive process is impossible,[35] its strongest proponents see corpus linguistics as a way to help achieve the dream of objectivity. Indeed, most often, they frame the utility of corpus linguistics techniques in terms of a critique of other methods of interpretation, which are often derided as "simple cherry-picking," [36] "subjective,"[37] or "idiosyncratic."[38] By contrast, corpus users have argued, corpus linguistics techniques can "help us deliver on the promise of an objective inquiry"[39] by reducing the subjectivity that plagues the process of legal interpretation.[40]   Corpus users' criticisms of other methods of statutory interpretation fall roughly into one of two categories. Some interpretations, they argue, rely on the *intuition* of the interpreter rather than data of language use.[41]   Other interpretations, they claim, rely on sources of language data that are *biased* and, as a result, fail to reflect

---

[33]   *Id.* at 839.

[34]   *Id. See also id.* at 839 n.33. In a revealing passage, the dissent demonstrated that the methodology employed by the court could support the opposite inference about the meaning of "information." *Id.* at 850 n.14 (Markman, J., dissenting in part).

[35]   *See* Gries & Slocum, *supra* note 9, at 1447; Mouritsen, *supra* note 8, at 203.

[36]   Lee & Mouritsen, *supra* note 7, at 807.

[37]   Phillips, Ortner & Lee, *supra* note 8, at 23–24 (arguing that corpus linguistics sometimes can "rescue" originalism from subjectivity). *See also* Mouritsen, supra note 8, at 202 ("Thus, the corpus method embodies the lexicographer's proud ideal of descriptive objectivity; his citations (and interpretations of them) are publicly verifiable.") (internal citations omitted).

[38]   Gries & Slocum, *supra* note 9, at 1441.

[39]   Lee & Mouritsen, *supra* note 7, at 796.

[40]   *See generally* State v. Rasabout, 356 P.3d 1258, 1274 (Utah 2015) (Lee, J., concurring); State v. J.M.S., 280 P.3d 410, 419 n.3 (Utah 2011); *In re* Adoption of Baby E.Z., 266 P.3d 702, 728 (Utah 2011) (Lee, J., concurring); Gries & Slocum, *supra* note 9, at 1441; Lee & Mouritsen, *supra* note 7, at 867; Mouritsen, *supra* note 8, at 175–78, 180; Nuñez, *supra* note 26, at 1521; Phillips & White, *supra* note 17, at 187.

[41]   *Rasabout*, 356 P.3d at 1274 (Lee, J., concurring); *J.M.S.*, 280 P.3d at 419 n.3; Gries & Slocum, *supra* note 9, at 1441; Lee & Mouritsen, *supra* note 7, at 806; Mouritsen, *supra* note 8, at 175–78; Nuñez, *supra* note 26, at 1521.

actual language use.[42] This Part identifies the two categories of subjectivity criticized by corpus users and describes the ways in which corpus users believe that corpus linguistics techniques reduce subjectivity.

Evaluating the claim that corpus techniques can reduce subjectivity and help achieve objectivity is complicated by two facts. First, although they use the language of "objectivity" and "subjectivity," proponents of corpus linguistics methods for statutory interpretation have so far declined to define these terms consistently. Second, objectivity and subjectivity have a number of different definitions. Without denying the diversity of views on the topic, for the purposes of this Article, an interpretation is objective if the mechanisms for arriving at it are free of bias and other distorting factors that obscure the interpretation.[43] As a corollary, an interpretation is objective if the process for reaching it is "reliably constrained by effective rules, procedures, or goals."[44] Conversely, an interpretation is subjective if it is not objective; that is, if bias or other distorting factors obscure the interpretation or if the interpretation is unconstrained by effective rules, procedures, or goals. Although there are other ways to define these terms, I use these definitions because they capture neatly both corpus users' criticisms of other methods of interpretation and their claim that corpus linguistics is a superior method of interpretation. Using these definitions, criticism by corpus users of interpretations that rely on the interpreter's intuition or biased data can be framed as a critique of the subjectivity of these methods compared with the relative objectivity of corpus methods. These definitions also capture the criticisms of corpus linguistics methods that I raise in Part V, below.

---

[42] Craig v. Provo City, 389 P.3d, 423, 428 n.3 (Utah 2016); *Baby E.Z.*, 266 P.3d at 728 (Lee, J., concurring); Ellen P. Aprill, *The Law of the Word: Dictionary Shopping in the Supreme Court*, 30 ARIZ. ST. L.J. 275, 297 (1998); James J. Brudney, *Faithful Agency Versus Ordinary Meaning Advocacy*, 57 ST. LOUIS U. L.J. 975, 981 (2013); Mouritsen, *supra* note 28, at 1921; Lawrence M. Solan, *Can Corpus Linguistics Help Make Originalism Scientific*, 127 YALE L.J.F. 57, 59 (2016).

[43] *Objective*, THE SAGE DICTIONARY OF SOCIOLOGY (1984) ("[W]e are being objective when we see things accurately, without our perception being distorted by our preferences, biases, and prejudices."); Fiss, *supra* note 4, at 744–45 ("[A]n interpretation can be measured against a set of norms that transcend the particular vantage point of the person offering the interpretation."). *See* William Eskridge, *The New Textualism and Normative Canons*, 113 COLUM. L. REV. 531, 576 (2013) (a canon is objective if it "impels judges to read statutes without regard to their own political preferences"). *See also* WILLARD QUINE, WORD AND OBJECT 7–8 (1960).

[44] Frank Lovett, *What Counts as Arbitrary Power?*, 5 J. POL. Power 137, 139 (2012) (defining arbitrary power). *Cf.* DOUGLAS E. EDLIN, COMMON LAW JUDGING 59 (2016) ("the opposite of validity is arbitrariness").

A. *Intuition and Statutory Interpretation*

The resort to intuition in statutory interpretation is neither new nor particularly controversial. As a number of commentators have noted, a native speaker's intuitive understanding of the meaning of a word is, in an important sense, its ordinary meaning.[45] In ordinary speech, speakers do not refer to "definitions, rules, or reasons" governing language use. Rather, speakers "simply use words," which are understood "because use and meaning are constituted by the life and practices of a community."[46] In this view, interpretation should largely be a non-technical exercise: a "competent user of ordinary language" normally should be able to determine a word's meaning.[47] This view of interpretation is supported by an important strain of linguistic theory. Noam Chomsky explains that native speakers of a language know (or perhaps better, "cognize"[48]) whether language use is ordinary or atypical, grammatical or ungrammatical.[49] In line with this view, it is common for judges to interpret statutes based on their intuitive understanding of the meaning of statutory text.[50]

This optimistic view of intuition, and the judicial practice that flows from it, has been criticized by corpus users,[51] among many others.[52]

---

[45] Robert Summers, *The Argument from Ordinary Meaning in Statutory Interpretation, in* ESSAYS IN LEGAL THEORY 216 (2000); Gries & Slocum, *supra* note 9, at 1434; Lawrence M. Solan, *The New Textualists New Text*, 38 LOY. U. L.A. L. REV. 2027, 2053–54 (2005) ("simple introspection is generally adequate to discover" the "most ordinary sense" of statutory language); Solan & Gales, *supra* note 17, at 1332.

[46] Paul E. McGreal, *Slighting Context: On the Illogic of Ordinary Speech in Statutory Interpretation*, 52 U. KAN. L. REV. 325, 334 (2004).

[47] Summers, *supra* note 45, at 228.

[48] NOAM CHOMSKY, RULES AND REPRESENTATIONS 70 (1980).

[49] *See* NOAM CHOMSKY, LANGUAGE AND THE PROBLEMS OF KNOWLEDGE 3 (1988); Alani Golanski, *Linguistics in Law*, 66 ALB. L. REV. 61 (2002); Lawrence M. Solan, *Can the Legal System Use Experts on Meaning*, 66 TENN. L. REV. 1167, 1181 (1999).

[50] Indeed, Justice Lee, a principal judicial proponent of corpus linguistics for interpretive purposes, acknowledges as much. State v. Rasabout, 356 P.3d 1258, 1274 (Utah 2015) (Lee, J., concurring) ("That leaves a third explanation for the majority's conclusion: The court's sense of *discharge* as *shoot* may simply be an expression of the majority's linguistic intuition."). *See* RICHARD POSNER, THE PROBLEMS OF JURISPRUDENCE 72–73 (1990) (noting that practical reason, including intuition, is our "principal set of tools for answering questions large and small"); *id.* at 124–25 ("Intuition, itself a method of practical reason, has its claims, and establishes presumptions that the other methods of practical reason may not always be able to overcome.").

[51] Gries & Slocum, *supra* note 9, at 1441; Lee & Mouritsen, *supra* note 7, at 806–07; Mouritsen, *supra* note 8, at 175–78; Nuñez, *supra* note 26, at 1520–21; Joseph Scott Miller, *Reasonable Certainty & Corpus Linguistics*, 66 U. KAN. L. REV. 39, 45–46 (2017); Lawrence B. Solum, *Originalist Methodology*, 84 U. CHI. L. REV. 269, 283–84 (2017).

[52] Outside the context of corpus linguistics, and indeed outside the context of statutory interpretation, judicial intuition has been both defended and criticized. *See, e.g.*, Arthur L. Corbin, *The Interpretation of Words and the Parole Evidence Rule*, 50 CORNELL L. REV.

Intuition, they argue, is a poor indicator of word meaning. Although intuition may be able to tell an interpreter whether a usage is ungrammatical, it cannot reliably be used to measure "the statistical *frequency* of words and word senses."[53] In other words, proponents of corpus linguistics argue, to the extent that ordinariness of meaning is linked to frequency of use (an assumption certainly open to challenge),[54] intuition cannot tell a judge which of two grammatical usages is the more ordinary one.[55] Instead, when judges rely on their intuition, their interpretive decisions can obscure why legislative drafters chose to use one term rather than another[56] or may simply be the result of a particular judge's idiosyncratic lexicon.[57] Moreover, even if reliance on intuition is up to the task of getting us through our daily lives without great difficulty,[58] statutory interpretation is a specialized activity that requires a more precise source of language data than intuition can provide. As Professors Lawrence Solan and Tammy Gales have described, the fine distinctions among word senses that judicial decisions require do not lend themselves to resolution by intuition alone.[59] Thomas Lee, Associate Justice of the Utah Supreme Court, and an outspoken proponent of corpus linguistics, has described judicial intuition as "less-than-perfect,"[60] "fallible,"[61] and "unreliable."[62]

Proponents of corpus linguistics usage point to Judge Richard Posner's use of Google as an example of a well-intentioned, but ultimately insufficient, attempt to cure problems associated with intuition.[63] In *Costello*, Judge Posner used Google to determine the meaning of the

---

161, 164 (1965) (arguing that, in the context of contract interpretation, "when a judge refuses to consider relevant extrinsic evidence on the ground that the meaning of written words is to him plain and clear, his decision is formed by and wholly based upon the completely extrinsic evidence of his own personal education and experience").

[53] Lee & Mouritsen, *supra* note 7, at 831.

[54] *See generally* Ehrett, *supra* note 17, at 62–64; Herenstein, *supra* note 17, at 116–19; Hessick, *supra* note 17, at 1508–09.

[55] Lee & Mouritsen, *supra* note 7, at 831; Mouritsen, *supra* note 8, at 175. Moreover, Chomsky's views on the innateness of language have been criticized as unsupported by biological research and the variety of extant grammatical structures. *See generally* Wolfram Hinzan, *The Philosophical Significance of Universal Grammar*, 34 LANG. SCI. 635 (2012).

[56] *Rasabout*, 356 P.3d at 1274.

[57] Mouritsen, *supra* note 8, at 175. *See* Gries & Slocum, *supra* note 9, at 1441.

[58] *See* Miller, *supra* note 25, at 2 (noting that polysemy "seldom causes any problems" in "everyday life").

[59] *See generally* Solan & Gales, *supra* note 45, at 1311.

[60] State v. J.M.S., 280 P.3d 410, 419 (Utah 2011) (Lee, J., concurring).

[61] *Rasabout*, 356 P.3d at 1275 (Lee, J., concurring).

[62] *In re* Adoption of Baby E.Z., 266 P.3d 702, 727 (Utah 2011) (Lee, J., concurring).

[63] Lee & Mouritsen, *supra* note 7, at 812–13; Phillips, Ortner & Lee, *supra* note 8, at 28–29.

statutory term "harbor."[64] Posner first formulated phrases, like "harboring fugitives," and searched for them on Google. He then noted how many search results matched the phrases he formulated and drew conclusions about the scope of "harbor" in the statute from the frequency with which his phrases appeared in the search results.[65] As proponents of corpus linguistics techniques have argued, Posner's use of Google reflects his intuition about which phrases containing "harbor" were likely to return results that could be compared profitably.[66] In other words, he searched only for those phrases that he thought were representative of potential meanings of harbor in the statute. His intuition about the possible statutory meanings of harbor may have been correct, and his choice of phrases to search well-considered, but the results his search returned were limited by his intuition about what searches to perform.[67]

Corpus users view reliance on corpus data to be a partial antidote to the subjectivity they attribute to reliance on intuition. Professor Lawrence Solum has argued that by relying on an individual's recollection, intuition provides only secondary evidence of language usage; corpus linguistics, by contrast, provides primary evidence of language usage.[68] Other commentators have argued that, in contrast with reliance on intuition, corpus linguistics is transparent; that is, by relying on data of language usage, a corpus user's conclusions about language usage can be challenged by other users.[69] Justice Lee made a detailed defense of corpus usage as compared with reliance on intuition in his concurring opinion in *Rasabout*.[70] In that case, the court interpreted the word "discharge" in a statute that prohibited unlawful discharge of a firearm. The question before the court was whether the defendant's action, firing twelve shots in rapid succession at the same target, was a single "discharge" or twelve separate discharges. The majority held that each shot was a separate discharge, justifying the conviction of twelve counts of violating the unlawful discharge statute.[71]

In concurrence, Lee opined that the court's conclusion was based on its equation of "discharge" with "shoot." Lee argued that this conclusion, while not necessarily wrong, was based on the court's intuition that

---

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

[65]   *Id.*

[66]   Lee & Mouritsen, *supra* note 7, at 812–13.

[67]   *Id.*

[68]   Solum, *supra* note 51, at 283–84.

[69]   Mouritsen, *supra* note 8, at 202–03; John D. Ramer, *Corpus Linguistics: Misfire or More Ammo for the Ordinary Meaning Canon*, 116 MICH. L. REV. 303, 326 (2017) (arguing that "transparency" is the COCA's greatest strength).

[70]   State v. Rasabout, 356 P.3d 1258, 1274 (Utah 2015) (Lee, J., concurring).

[71]   *Id*. at 1262–64.

"discharge" and "shoot" are roughly synonymous. Lee objected not to the conclusion, but to the fact that the court did not test its intuition against alternatives. When judges rely on intuition, he argued, they often fail to acknowledge that there are "alternative senses of the operative terms."[72] Because a judge's intuition is based on her particular experiences and recollections of a word's meaning, the parties are deprived of the ability to challenge the basis of judge's conclusion. Although a judge's intuition may be representative of general usage, it may be idiosyncratic instead.[73] Corpus linguistics methods, Lee opined, cure this defect in intuition-based textual analysis by allowing an interpreter to demonstrate how she determined a word's meaning. Lee tested his intuition that "discharge" means "shoot" by searching a corpus for information about how discharge is used when referring to a firearm. By relying on publicly available data rather than intuition, Lee argued, his assumptions can be challenged, his methods replicated, and conclusions falsified or validated.[74] Other corpus users have acknowledged that corpus usage does not completely eliminate subjectivity due to reliance on intuition. Gries and Slocum noted that "a certain degree of subjective intuition is virtually unavoidable" in corpus analysis.[75] And Mouritsen acknowledged that the "human beings at both ends of the corpus—the architect and the user" are both "subject to . . . errors and biases."[76] But, corpus users tend to agree with Lee that, by making the decision-making process more transparent, corpus linguistics provides a standard against which one user's conclusions can be measured by another.[77]

B. *Biased Reference Data and Statutory Interpretation*

When they do not rely on their intuition about the meaning of a text, legal interpreters refer to materials outside the text being interpreted to determine its meaning. This, too, is common and uncontroversial, at least in some forms. When an interpreter searches Westlaw to learn how other judges have interpreted a statutory term—for example "knowingly and willfully"—she is seeking word meaning outside the text of the statute itself.[78] Seeking meaning from reference data becomes more controversial,

---

[72] *Id*. at 1274.

[73] *Id*. at 1274–75. *See also id.* at 1275 n.10; Gries & Slocum, *supra* note 9, at 1441.

[74] *Rasabout*, 356 P.3d at 1282.

[75] Gries & Slocum, *supra* note 9, at 1447.

[76] Mouritsen, *supra* note 8, at 203.

[77] *Id*.; Ramer, *supra* note 69, at 326 (arguing that "transparency" is the COCA's greatest strength).

[78] *See, e.g.*, United States v. Wheeler, 540 F.3d 683, 689 (7th Cir. 2008) (interpreting "knowingly and willfully" in embezzlement statute). For just a few of the hundreds of federal and state statutes that use the phrase "knowingly and willfully," both in the criminal

however, when it appears the data consulted have been compiled or selected arbitrarily, opportunistically, or otherwise without justification.[79] The much-criticized[80] *Muscarello* case illustrates how courts sometimes interpret statutory language by referring to data that appears to be selected arbitrarily.[81] In *Muscarello*, the Supreme Court considered whether a person "carries" a firearm when he drives with it in his locked glove compartment or trunk. The Court weighed two possible interpretations: if carry means only to "bear on one's person," then the conduct was not prohibited; by contrast, if carry means "transport," then it covered the defendants' conduct. The Court interpreted the word carry by referring to the use of that word in a wide variety of sources, including *Robinson Crusoe*, *Moby Dick*, the *King James Bible*, newspaper articles, and a series of dictionaries.[82] Although *Muscarello* has been widely criticized, it is not unique. Courts, with regularity, consult materials outside the text to determine the meaning of words in statutes.[83] In *Whitfield*, the Supreme Court interpreted the word "accompany" by reference to the use of that word in *David Copperfield*, *Pride and Prejudice*, and a newspaper marriage announcement.[84] And as noted above, in *Costello*, Judge Posner performed a Google search to determine the meaning of the word "harbor" in a statute that prohibited harboring an alien.[85]

Users of corpus linguistics techniques criticize these uses of extra-textual materials, but *not* because they object to searching for meaning outside the statutory text. Indeed, corpus users support the search for meaning outside the text of the statute being interpreted.[86] They do criticize, however, what they perceive to be subjective and unprincipled references to materials outside the text—what Lee and Mouritsen have

---

and civil context, *see, e.g.*, 5 U.S.C. § 552a (2018); 6 U.S.C. § 625 (2018); 15 U.S.C. § 1196 (2018); 18 U.S.C. § 669 (2018); 18 U.S.C. § 1001 (2018); 21 U.S.C. § 331(yy) (2018); 2 U.S.C. § 1966 (2018); 42 U.S.C. § 1383 (2018); 49 U.S.C. § 60123 (2018); LA. STAT. ANN. § 18:1505.5 (2006); WIS. STAT. ANN. § 146.84. (West 2006).

[79] 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, 490, 566 (2013).

[80] Lee & Mouritsen, *supra* note 7, at 807–10; Mouritsen, *supra* note 28, at 1931–32; Solan, *supra* note 45, at 2052–53.

[81] Muscarello v. United States, 524 U.S 125, 126–127 (1998).

[82] *Id.* at 128–31.

[83] Magone v. Heller, 150 U.S. 70, 74–75 (1893) (citing one of Shakespeare's plays for the meaning of statutory text); Bok v. McCaughn, 42 F.2d 616, 618–19 (3d Cir. 1930) (citing a version of the Bible for the meaning of statutory text).

[84] Whitfield v. United States, 135 S. Ct. 785, 788 (2015).

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

[86] *E.g.*, Gries & Slocum, *supra* note 9, at 1441; Lee and Mouritsen, *supra* note 7, at 807.

called "simple cherry-picking."[87]   As Solan has noted, the *Muscarello* Court chose reference materials without offering a reason why they were relevant to the interpretive question before the Court.[88]   For example, the Court referred to the Latin and Old French roots of "carry" and to the use of the word in a translation of the Bible to derive its meaning. It is not obvious why etymological information or use of a word in translation is relevant to the meaning of statutory text and the Court provided no explanation.[89]

Corpus users have criticized the practice of resorting to dictionaries for interpreting statutes especially harshly.[90]   Although dictionaries are routinely consulted by not only the Supreme Court,[91] but lower federal courts[92] and state courts[93] as well, the practice has long been criticized by scholars.[94]   Among the many shortcomings that have been described, dictionaries are detached from ordinary meaning and legislative intent,[95] they are often deliberately devoid of context,[96] they do not purport to describe all semantically acceptable word meanings,[97] they contain

---

[87]   Lee & Mouritsen, *supra* note 7, at 807.

[88]   Solan, *supra* note 45, at 2052.

[89]   *Id.* (criticizing citation to etymology and the Bible); Mouritsen, *supra* note 28, at 1939–40 (criticizing reference to etymology). The Court hinted at an explanation when it noted that the "greatest of writers" have used the term "carry" in the way it suggested. Muscarello v. United States, 524 U.S 125, 129 (1998). The Court did not explain the relevance to statutory meaning of the way these writers used the phrase. The Court, perhaps, was implying that these great writers used the word in a typical or ordinary way. Typical or ordinary usage in a novel or the Bible, however, is not the same as typical or ordinary usage in a statute.

[90]   Mouritsen, *supra* note 28, at 1939; BRIAN G. SLOCUM, ORDINARY MEANING: A THEORY OF THE MOST FUNDAMENTAL PRINCIPLE OF LEGAL INTERPRETATION 217 (2015).

[91]   *E.g.*, Yates v. United States, 135 S. Ct. 1074, 1081 (2015); Taniguchi v. Kan Pac. Saipan, Ltd., 566 U.S. 560, 566–69 (2012); MCI Telecomm. Corp. v. Am. Tel. & Tel. Co., 512 U.S. 218, 225 (1994); Smith v. United States, 508 U.S. 223, 228–29 (1993).

[92]   *E.g.*, Phillips v. AWH Corp., 415 F.3d 1303, 1318 (Fed. Cir. 2005) ("dictionaries . . . have been properly recognized as among the many tools that can assist the court in determining the meaning of particular terminology"); Zatarains, Inc. v. Oak Grove Smokehouse, Inc., 698 F.2d 786, 792 (5th Cir. 1983).

[93]   *E.g.*, Wasatch Prop. Mgmt. v. Degrate, 112 P.3d 647, 653 (Cal. 2005); Koontz v. Ameritech Servs., Inc., 645 N.W.2d 34, 39 (Mich. 2002); Madison Constr. Co. v. Harleysville Mut. Ins. Co., 735 A.2d 100, 107–08 (Pa. 1999).

[94]   Brudney & Baum, *supra* note 79, at 490, 566; Pamela Hobbs, *Defining the Law: (Mis)using the Dictionary to Decide Cases*, 13 DISCOURSE STUD. 327, 330-31 (2011). *See also* United States v. Costello, 666 F.3d 1040, 1043–44 (7th Cir. 2012); Cabell v. Markham, 148 F.2d 737, 739 (2d Cir. 1945).

[95]   Alice A. Wang, *Googling for Meaning: Statutory Interpretation in the Digital Age*, 125 YALE L.J.F. 267, 278 (2016).

[96]   *Costello*, 666 F.3d at 1044; *see also* Craig v. Provo City, 389 P.3d 423, 428 n.3 (Utah 2016).

[97]   Aprill, *supra* note 42, at 297.

definitions that support multiple readings of the statute,[98] and they are often used opportunistically by legal interpreters.[99] To take just one recent example of the dubious value of judicial use of dictionaries, the Court of Appeals of Utah[100] relied, in part, on a definition in the Urban Dictionary,[101] a crowdsourced, online compilation of user-approved phrases and their definitions, often including "ad-hoc neologism[s], invented just for this dictionary."[102] Similarly, proponents of corpus linguistics have criticized Judge Posner's use of Google, in part, because of the secrecy in the way that Google returns search results. This secrecy, they argue, detracts from the ability to replicate the results of a Google search.[103]

Proponents of corpus linguistics techniques argue that corpus research is not prone to the same biases that affect legal interpreters using reference materials like dictionaries or Google. Using a general corpus, like the COCA, they argue, allows a legal interpreter to search for the meaning of a word in the context of how words are ordinarily used in spoken and written language.[104] This process mitigates the bias inherent in the choice of a word's meaning from a list of dictionary definitions, which is necessarily acontextual.[105] Moreover, they argue, using corpus data broadly representative of written and spoken language diminishes the bias associated with the interpreter's choice of a particular dictionary.[106] Compared with Google, proponents of corpus linguistics techniques argue, corpora like the COCA are more transparent about the methodologies they use to organize search results. As a result, the use of corpora like the COCA is less likely to reflect biases that are unknown and unknowable to the average user not privy to Google's search algorithm.[107]

In conclusion, this Part described the dream of objectivity pursued by users of corpus linguistics. It is a dream rooted in perceived deficiencies of

---

[98] *See Craig*, 389 P.3d at 428 n.3.

[99] Brudney & Baum, *supra* note 79, at 490, 566; Brudney, *supra* note 42, at 975, 981. This is, strictly speaking, a criticism of dictionary use, not of dictionaries themselves. But, to the extent that dictionaries are especially susceptible to opportunistic use, it is worth including this flaw among the dictionary's other shortcomings.

[100] Utah is, incidentally, the intellectual home of the use of corpus linguistics methods for legal interpretation.

[101] O'Hearon v. Hansen, 409 P.3d 85, 93 n.8 (Utah Ct. App. 2017).

[102] Virginia Heffernan, *Street Smart: The Unruly, Unlexicographical but Surprisingly Useful Offerings of Urban Dictionary*, N.Y. TIMES, July 5, 2009, at SM16.

[103] Lee & Mouritsen, *supra* note 7, at 812.

[104] *Id*. at 831–32.

[105] Gries & Slocum, *supra* note 9, at 1441.

[106] *Id*. at 1438.

[107] *See* Eric Goldman, *Search Engine Bias and the Demise of Search Engine Utopianism*, 8 YALE J.L. & TECH. 188, 192–94 (2005-2006). *See also* A. Diaz, *Through the Google Goggles: Sociopolitical Bias in Search Engine Design*, *in* WEB SEARCH: MULTIDISCIPLINARY PERSPECTIVES 29–30 (Spink and Zimmer, eds., 2008).

common methods of statutory interpretation, like reliance on judicial intuition and reliance on certain reference materials extrinsic to the text. As corpus users describe these methods of interpretation, they are subjective within the definition given earlier. An interpretation relying on intuition is subjective because it is not constrained by effective procedures and may be colored by the idiosyncratic knowledge and disposition of the interpreter. Reliance on sources like dictionaries and Google searches is subjective because these sources contain data that is biased in favor of atypical uses; reliance on these biased data will obscure accurate interpretations. By contrast, they argue, corpus linguistics results are objective within the definition described above because they reflect data that is broadly representative of ordinary language usage. Because they reflect ordinary language, corpus users argue, they are not susceptible to biases that distort the result. Part IV, below, assesses whether the dream of objectivity held by proponents of corpus linguistics withstands scrutiny.

IV. CORPUS LINGUISTICS AND THE REALITY OF SUBJECTIVITY

Proponents of corpus linguistics argue that, by using corpus linguistics techniques, interpreters can reduce the subjectivity endemic to statutory interpretation.[108] Although some corpus users acknowledge that the elimination of subjectivity is impossible,[109] they argue that corpus linguistics methods are not as subjective as other methods of interpretation.[110] Perhaps most explicitly, Lee and Mouritsen argue that corpus linguistics can "help us deliver on the promise of an objective inquiry"[111] and, in particular, help textualism deliver on its "promise of determinacy."[112] Indeed, as Professor Carissa Byrne Hessick explained in her critique of using corpus linguistics techniques to interpret criminal statutes, much of the appeal of corpus linguistics is "that it promises us right answers."[113] Because users of corpus linguistics for statutory interpretation place such weight on its ability to reduce subjectivity, it is appropriate to assess the depth and contours of the subjectivity involved in using corpus linguistics itself. Only then can potential users of corpus linguistics techniques for statutory interpretation—including judges and scholars—make an informed decision about whether corpus linguistics

---

[108]   State v. Rasabout, 356 P.3d 1258, 1274–75 (Utah 2015) (Lee, J., concurring); State v. J.M.S., 280 P.3d 410, 419 (Utah 2011) (Lee, J., concurring); Gries & Slocum, *supra* note 9, at 1441; Lee and Mouritsen, *supra* note 7, at 867; Mouritsen, *supra* note 8, at 175–78; Nuñez, *supra* note 26, at 1521; Phillips & White, *supra* note 26, at 186–87.

[109]   Gries & Slocum, *supra* note 9, at 1447; Mouritsen, *supra* note 8, at 203.

[110]   Gries & Slocum, *supra* note 9, at 1441; Lee and Mouritsen, *supra* note 7, at 807.

[111]   Lee & Mouritsen, *supra* note 7, at 796.

[112]   *Id*. at 876.

[113]   Hessick, *supra* note 17, at 1519.

techniques offer any advantage over other methods of interpretation.

Despite the emphasis that users of corpus linguistics place on its subjectivity-reducing capabilities, corpus linguistics techniques involve significant subjective interpretive choices. These choices disrupt the dream of objectivity held by proponents of corpus linguistics as a method of statutory interpretation. There are a number of distinct points during the interpretive process at which a user of corpus techniques must make a subjective decision that influences the interpretive outcome. First, at the beginning of the interpretive process, a user of corpus linguistics techniques must choose a particular corpus to search. Just as a legal interpreter resorting to a dictionary must choose one or more dictionaries to consult,[114] users of corpus linguistics techniques must choose a particular corpus to search. There are many different corpora. Each contains a different mix of texts and reliance on one does not lead to the same results on reliance on another.[115] Second, the user must choose search parameters. If a statute makes it a crime to "carry a firearm," for example, the corpus user must decide whether to search for the word "carry," the phrase "carry a firearm," or some other term.[116] The corpus user also must decide whether and how to customize the search to return results indicating only certain parts of speech,[117] or results reflecting certain geographic locations,[118] speech communities,[119] or time periods.[120] Third, a corpus search will often return results that the user believes are not germane to the statutory inquiry. The user of corpus linguistics techniques must make a subjective decision about which search results to evaluate and which results to exclude from evaluation.[121]

Because each of these three sources of subjectivity is significant enough to warrant its own separate treatment, this Article will explore just the first source of subjectivity identified above: the choice of corpus. This Part first examines the act of choosing a corpus and concludes that, rather than leading to an objective interpretation, the choice of corpus introduces subjectivity into the interpretive process. It next assesses the choice of corpus consistently made by users of corpus techniques—a general corpus populated by nonlegal language—and concludes that it is the wrong choice

---

[114] Aprill, *supra* note 42, at 296–97; Gries & Slocum, *supra* note 9, at 1421.
[115] Solan & Gales, *supra* note 17, at 1311. *Compare* NOW CORPUS (NEWS ON THE WEB), https://corpus.byu.edu/now (last visited Sept. 16, 2019) *with* CORPUS HISTORICAL AM. ENG., https://corpus.byu.edu/coha (last visited Sept. 16, 2019).
[116] Solan & Gales, *supra* note 17, at 1346.
[117] Ramer, *supra* note 69, at 327. *See also* Gries & Slocum, *supra* note 9, at 1448.
[118] *E.g.*, Lee & Mouritsen, *supra* note 7, at 857.
[119] Solan, *supra* note 45, at 2059.
[120] Lee & Mouritsen, *supra* note 7, at 857.
[121] *E.g.*, Lee & Mouritsen, *supra* note 7, at 850–51.

for statutory interpretation. Finally, it considers the possibility of using a still-hypothetical "legal corpus;" it concludes that this hypothetical corpus would not ameliorate the subjectivity problems that plague the choice of corpus.

A. *The Choice of Corpus is Subjective*

Just as a legal interpreter resorting to a dictionary must choose a particular dictionary to use, so too must the user of corpus linguistics techniques choose a corpus to search. The choice of corpus is subjective because it is not constrained by any principle that suggests why one corpus rather than another should be chosen. As Professor Solan has explained, there is nothing internal to a particular corpus that requires its use in certain circumstances.[122] Likewise, there is nothing about a particular term or phrase that tells the interpreter which corpus to use when searching for its meaning. As a result, simply by opting for a corpus search, the user of corpus linguistics techniques introduces a subjective element into the interpretive process.

Corpus usage confirms that the choice of corpus is subjective: corpus users rely on multiple or different corpora without articulating a standard for determining when one corpus would be appropriate and another would not be appropriate. Take Lee and Mouritsen's searches for the terms "vehicle," "carry," and "interpreter" in their work advocating the adoption of corpus techniques.[123] Lee and Mouritsen rely on searches in the News on the Web (NOW) Corpus and the Corpus of Historical American English (COHA) without describing why either or both of these corpora are appropriate for their searches and despite the significant differences between the texts found in these corpora.[124] The NOW Corpus, for example, contains not only news sources, but also online magazines with subjects as diverse as video games, cricket, and fashion. And the origin of these web sources? The NOW Corpus includes texts that come not only from the United States, but, unless specifically excluded by the researcher, texts from markedly different linguistic communities, like India, Nigeria, Singapore, Kenya, Pakistan, and the Philippines, among others.[125] The

---

[122] Solan, *supra* note 45, at 2059–60. *See also* Solan & Gales, *supra* note 17, at 1314–15.
[123] Lee & Mouritsen, *supra* note 7, at 836–51.
[124] *Id. See* Solan, *supra* note 42, at 60–61 (arguing that there is a substantive choice involved in searching for ordinary meaning rather than an expansive meaning or specialized meaning).
[125] NOW CORPUS (NEWS ON THE WEB), *supra* note 115. The NOW Corpus contains "8.5 billion words of data from web-based newspapers and magazines from 2010 to the present time. More importantly, the corpus grows by about 140–160 million words of data each month (from about 300,000 new articles), or about 1.8 billion words each year." It is

COHA, by contrast, includes different kinds of texts, including movie scripts and poetry.[126]

Some proponents of corpus linguistics techniques acknowledge that they must choose a corpus, but minimize the significance of the choice by suggesting that it is driven by a distinction between "ordinary" words and legal terms of art.[127] If the word under consideration is an "ordinary" one, they search for it in a general corpus, like the COHA, the COCA, or the NOW Corpus; by contrast, if it is a legal term of art, some intimate that interpreters should use a still-hypothetical specialized legal corpus.[128] However, framing the choice of corpus as a choice between an ordinary term and a legal term of art does not eliminate its subjectivity; it merely substitutes one subjective decision for another. The determination that a word is ordinary itself reflects a subjective decision because there is not an objective way to distinguish between ordinary words and legal terms of art.[129] As linguists have noted, the line between legal terms of art and ordinary words is indistinct at best. David Mellinkoff notes that not every word "that has the sound of the law is a term of art."[130] Conversely, many words that sound ordinary, because they are used in nonlegal settings, also have specialized legal meanings.[131] For these reasons, the "difference

---

possible to limit NOW's results by country.

[126]   CORPUS HISTORICAL AM. ENG., *supra* note 115.

[127]   *See* Phillips, Ortner & Lee, *supra* note 8, at 23 (asserting without explanation that "corruption of blood" is a term of art while "commerce" is an ordinary word).

[128]   Phillips, Ortner & Lee, *supra* note 8, at 24–25. *See also* Solan, *supra* note 42, at 59–60 (noting that searching a corpus designed to reflect ordinary meaning is not very useful for determining the meaning of terms of art). *Cf.* James A. Heilpern, *Dialects of Art: A Corpus-Based Approach to Technical Term of Art Determinations in Statutes*, 58 JURIMETRICS 4, 380 (2018) (suggesting that technical terms of art (but not legal terms of art) should be interpreted according to meaning found in technical documents).

[129]   *E.g.*, Heikki E. S. Mattila, *Legal Vocabulary*, *in* THE OXFORD HANDBOOK OF LANGUAGE AND LAW, *supra* note 6, at 29, 31 ("The difference between legal terms and words of ordinary language is relative and hard to define."); DAVID MELLINKOFF, THE LANGUAGE OF THE LAW 17–18 (2004) (describing that some, but not every, legal-sounding term is a term of art and that some legal words are intended for both lawyers and non-lawyers); PETER M. TIERSMA, LEGAL LANGUAGE 107–08 (1999); Isabel Richard, *Is Legal Lexis Characteristic of Legal Language*, 11 J. LEGAL LEXICOLOGY 1, 9 (2018) ("Firstly, legal lexis is used by law, but not exclusively, and may have legal meaning, but not necessarily.").

[130]   MELLINKOFF, *supra* note 129, at 17.

[131]   Mattila, *supra* note 129, at 31 ("[T]he use of ordinary words in a technical legal sense is particularly widespread."); MELLINKOFF, *supra* note 129, at 11–12; Ralf Poscher, *Ambiguity and Vagueness in Legal Interpretation*, *in* THE OXFORD HANDBOOK OF LANGUAGE AND LAW, *supra* note 6, at 132; Frederick Schauer, *Hohfeld on Legal Language*, *in* THE LEGACY OF WESLEY HOHFELD: EDITED MAJOR WORKS, SELECT PERSONAL PAPERS, AND ORIGINAL COMMENTARIES 7–8 (Shyam Balganesh et al. eds., forthcoming), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3183858 (noting that words like search and seizure have legal meanings that are both over- and under-inclusive of their ordinary

between legal terms and words of ordinary language is relative and hard to define."[132] Even linguists who are more optimistic about the possibility of identifying legal terms of art recognize the significant disagreement over what constitutes a legal term.[133] Because choosing to designate a statutory term ordinary rather than legal does not appear to be "reliably constrained,"[134] the choice between a general corpus and a still-hypothetical specialized legal corpus is subjective.[135]

Practice confirms the subjectivity of designating a term ordinary or legal.   Rather than announcing and adhering to a standard for distinguishing between ordinary terms and legal terms, corpus users treat terms as ordinary without analysis and without adhering to any discernable principle. For example, corpus users have searched for the statutory terms "results in,"[136] "information,"[137] and "discharge"[138] in the COCA. In none of these cases did the interpreters demonstrate that they applied some rule or principle to determine whether these words were ordinary rather than legal terms. The recent *American Bankers* case is illustrative of the lack of standards applied by corpus users choosing a general corpus. In that case, the court relied on searches in the COHA and databases of newspaper articles[139] without acknowledging that these corpora differ in essential ways. Similarly, in the *Harris* case, the Michigan Supreme Court relied on a search in the COCA to uncover the meaning of the term "information."[140] The court asserted that it was searching for the ordinary meaning of the term, but did not justify this assertion. Curiously, the court purported to rely on a statute that governs statutory interpretation in that state. The statute the court relied on, however, provides no support for the proposition

---

meaning); David A. Strauss, *Why Plain Meaning?*, 72 NOTRE DAME L. REV. 1565, 1568 (1997) ("Moreover, the so-called 'ordinary meaning' is not so ordinary. It is the ordinary *legal* meaning . . . . Terms like witness, zoning, even speed limit, when used in a legal context, can mean something quite different from what they might mean when used in other contexts.").

[132]   Mattila, *supra* note 129, at 31. *See also* TIERSMA, *supra* note 129, at 108 (the distinction between terms of art and legal jargon "is mainly one of degree").

[133]   TIERSMA, *supra* note 129, at 108.

[134]   Lovett, *supra* note 44, at 139 (defining arbitrary power).   In the context of "intersubjectivity," "the opposite of validity is arbitrariness." EDLIN, *supra* note 44, at 59.

[135]   As will be discussed in Part IV.C, *infra*, even if it were possible to determine objectively that a statutory term is a legal term of art rather than an ordinary term, a single term of art can have multiple meanings because there is not a single "legal English."

[136]   Wilson v. Safelite, Grp., Inc., 930 F.3d 429, 439–40 (6th Cir. 2019) (Thapar, J., concurring).

[137]   People v. Harris, 885 N.W.2d 832, 839, 839 n.33 (Mich. 2016).

[138]   State v. Rasabout, 356 P.3d 1258, 1274–75 (Utah 2015) (Lee, J., concurring).

[139]   Am. Bankers Ass'n v. Nat'l Credit Union Admin, 306 F. Supp. 2d 44, 68 n.5 (D.D.C. 2018).

[140]   *Harris*, 885 N.W.2d at 839.

that the word at issue should be interpreted as an ordinary term; rather, it simply provides that an ordinary word should be given its ordinary meaning while a term of art should be interpreted according to its appropriate technical meaning.[141]   Nevertheless, the court searched a general corpus without indicating why the word "information" in the statute was an ordinary one rather than a legal term of art.[142]

Moreover, even when corpus users acknowledge that ordinary terms and specialized terms should be treated differently, the reason for their choice of a general corpus is obscure. Take, for example, the Utah case, *In re: Baby E.Z.* In his dissent, Justice Lee considered the interpretation of the word "custody" in the federal Parental Kidnapping Prevention Act. Lee acknowledged that statutory terms of art should be read according to their *legal* meaning rather than their ordinary meaning.[143]   He found that the statutory term was a legal term that should be interpreted according to its legal meaning.[144]   Nevertheless, Justice Lee proceeded to search for "custody" in the COCA,[145] a *general* corpus that includes transcriptions of spoken language, fiction, popular magazines, newspaper articles, and academic works but, crucially, no statutory text.[146]   Lee's reliance on the COCA in *Baby E.Z.* suggests that corpus users are not relying on a distinction between ordinary terms and legal terms when choosing a corpus. That is, it appears that no matter whether a word is a legal term or an ordinary one, the corpus user will search a general corpus for its meaning. If Lee's use of the COCA in *Baby E.Z.* is the correct way to use corpus techniques to interpret statutes, it is difficult to imagine the existence of a test that can be used to choose a corpus objectively.

B. *A Corpus of Nonlegal Language is the Wrong Choice for Statutory Interpretation*

Users of corpus linguistics techniques for statutory interpretation rely—virtually exclusively—on searches in general corpora, like the COCA or the COHA. The justification for interpreting statutory language in accordance with the meaning of words in a general corpus rests on the assumption that the meaning of words in a general corpus is the same, in a

---

[141]   MICH. COMP. LAWS § 8.3(a) (2017).
[142]   *See also Rasabout*, 356 P.3d at 1274–75 (Lee, J., concurring) (rejecting without explanation the argument that a statutory term was used as a legal term).
[143]   *In re* Adoption of Baby E.Z., 266 P.3d 702, 723-24 (Utah 2011) (Lee, J., concurring).
[144]   *Id.* ("Instead, the omission of a definition for the term 'custody' and its repeated use in the [Act] suggest that we ought to interpret the term with reference to its ordinary legal meaning.").
[145]   *Id.* at 724–25 n.23.
[146]   CORPUS CONTEMPORARY AM. ENG., *supra* note 30.

relevant way, as the meaning of those same words in statutes. Users of corpus linguistics for statutory interpretation justify the equation of statutory language and general corpus language by suggesting that statutory language and general corpus language are both "ordinary" uses of language. That is, they argue that the law often requires them to look for the "ordinary meaning" of statutory language and that they can find this meaning by looking at the way language is used in a general corpus.[147]

Embedded in this argument is the premise that the ordinary meaning of statutory language is its nonlegal meaning.[148] A number of scholars, including scholars of corpus linguistics, have suggested that it would be normatively attractive if this were true. Professor Slocum explains that if "one assumes that successful communication is the goal in most cases," then statutes "should be understood by different people in the same way . . . . [Therefore,] legal texts should be understandable to the general public, as well as to judges and sophisticated practitioners."[149]

There is, of course, ample authority for the proposition that the law *ought* to provide notice to those who are governed by it. Famously, if unrealistically, Jeremy Bentham argued that laws ought to put into the mind of the citizen "an exact idea of the will of the legislator."[150] More

---

[147] *E.g.* Wilson v. Safelite Grp., Inc., 930 F.3d 429, 439 (6th Cir. 2019) (Thapar, J., concurring).

[148] This argument also assumes that courts are in fact attempting to interpret statutory language to conform with ordinary meaning, however defined. It is true that courts sometimes claim that they are searching for a term's ordinary meaning. It would be a mistake, however, to read too much into judicial statements that courts are in fact attempting to interpret statutory language according to its ordinary meaning. For one reason, courts typically do not confine their interpretations to ordinary meaning. SLOCUM, *supra* note 90, at 172–174. As Miranda McGowan noted, the "ordinary meaning rule," if it can be called a rule, is "riddled with exceptions." Miranda McGowan, *Do as I Do, Not as I Say: An Empirical Investigation of Justice Scalia's Ordinary Meaning Method of Statutory Interpretation*, 78 MISS. L.J. 129, 140, 157 (2008). *See also* Linda D. Jellum, *On Reading the Language of Statutes*, 8 U. MASS. L. Rev. 184, 204 (2013). For another reason, just as often as courts claim that they are interpreting legal language according to its ordinary meaning, they suggest that they are relying on the "plain meaning" of the text, a phrase that linguists do not take to mean the same thing as "ordinary." SLOCUM, *supra* note 90, at 22, 24–26. *See, e.g.* Kingdomware Techs., Inc. v. United States, 136 S. Ct. 1969, 1978 (2016) (applying the plain meaning of statutory language despite the statute's prefatory language announcing a different objective); Lockhart v. United States, 546 U.S. 142, 146 (2005) (giving effect to plain meaning although Congress may not have intended it); Lamie v. U.S. Tr., 540 U.S. 526, 535 (2004) (applying plain meaning although the sentence is "awkward"); I.N.S. v. Phinpathya, 464 U.S. 183, 192 (1984) (applying the "plain meaning of the statute" "however severe the consequences"). Because courts do not always purport to apply a statute's ordinary meaning, and because, even when they do, they do not always apply the ordinary meaning in fact, it is misplaced to rely on judicial statements about their search for ordinary meaning to conclude that it is appropriate to interpret statutory language according to its ordinary meaning.

[149] Slocum, *supra* note 17, at 14.

[150] *E.g.*, Jeremy Bentham, *A General View of a Complete Code of Laws, in* 3 THE

modestly, Professor Richard Fallon suggests that law "too far divorced from its ordinary understanding would not be legitimate."[151]   And the Supreme Court has reiterated that laws must "give the person of ordinary intelligence a reasonable opportunity to know what is prohibited, so that he may act accordingly."[152]   But, the normative judgment that legislatures *should* speak in language that is intelligible to the general public is not the same as the descriptive claim that legislatures do in fact speak in the type of language found in a general corpus—that is—*nonlegal language*. Instead, the validity of searching a general corpus for the meaning of statutory language depends on the descriptive claim that nonlegal language is relevantly similar to statutory language.[153]

There is significant scholarly debate about whether legal language is the same as nonlegal language in a way that is relevant to statutory interpretation.[154]   Here, I will identify and discuss the differences between nonlegal language and statutory language that bear directly on the question of searching for statutory meaning in a general corpus.[155]   I demonstrate

---

WORKS OF JEREMY BENTHAM 207–08 (1816).
    [151]   Richard H. Fallon, Jr., *The Statutory Interpretation Muddle*, 114 NW. L. REV. 269, 331 (2019).
    [152]   Grayned v. City of Rockford, 408 U.S. 104, 108 (1972).
    [153]   *See* Peter M. Tiersma, *Some Myths About Legal Language*, *in* SPEAKING OF LANGUAGE AND LAW 27, 32 (Lawrence M. Solan et al. eds., Oxford Univ. Press, 2015) (noting the longstanding hope that law can be expressed in a way that is understood by ordinary people, but concluding that this is unlikely). Moreover, the normative argument that legislatures should speak in nonlegal language is doubtful. Because of the complex tasks that modern legislation is designed to accomplish, any attempt by legislatures to write in nonlegal language is apt to be ineffective.
    [154]   *E.g.*, Andrei Marmor, *The Pragmatics of Legal Language*, 21 RATIO JURIS 423, 429 (2008); Frederick Schauer, *On the Relationship Between Legal and Ordinary Language, in* SPEAKING OF LANGUAGE AND LAW, *supra* note 153, at 35; Frederick Schauer, *Is Law a Technical Language*, 52 SAN DIEGO L. Rev. 501, 501–02 (2015); Schauer, *supra* note 131, at 18-21; SLOCUM, *supra* note 90, at 5–14; William Baude & Stephen Sachs, *The Law of Interpretation*, 130 HARV. L. REV. 1082, 1085–87 (2017); Summers, *supra* note 45, at 229; David A. Strauss, *Why Plain Meaning?*, 72 NOTRE DAME L. REV. 1565, 1568 (1997); Paul E. McGreal, *Slighting Context: On the Illogic of Ordinary Speech in Statutory Interpretation*, 52 U. KAN. L. REV. 325, 334–35 (2004); Marc Poirier, *On Whose Authority? Linguists' Claim of Authority to Interpret Statutes*, 73 WASH. U. L. Q. 1025, 1033–34, 1057 (1995); Victoria Nourse, *Two Kinds of Plain Meaning*, 76 BROOK. L. REV. 997, 1000–05 (2011); Mark Greenberg, *Legislation as Communication? Legal Interpretation and the Study of Linguistic Communication, in* PHILOSOPHICAL FOUNDATIONS OF LANGUAGE IN THE LAW 217, 221 (2011); Stephen F. Ross, *The Limited Relevance of Plain Meaning*, 73 WASH. U. L. Q. 1057, 1064–65 (1995); John O. McGinnis & Michael B. Rappaport, *Original Methods Originalism*: *A New Theory of Interpretation and the Case Against Construction*, 103 NW. U. L. REV. 751, 765 (2009); McGinnis & Rappaport, *supra* note 18, at 1356–60; *see also* Jim Chen, *Law as a Species of Language Acquisition*, WASH. U. L. Q. 1263, 1286–87 (1995).
    [155]   Although these same arguments can also be made about some conceptions of ordinary meaning outside of the context of corpus linguistics, they apply with special force

SETON HALL LAW REVIEW   [Vol. 50:401

that statutory language and the language of texts found in a general corpus have different purposes, audiences, and other linguistic characteristics, like word choice and syntax. In other words, statutory language and nonlegal language do not share the same context. Because of their different contexts, interpreting statutory language according to the meaning of those same words in nonlegal texts fails to capture meaning attributable to the fact that the words are, in fact, found in statutes. As a result, the meaning of statutory text cannot be equated with the meaning of nonlegal texts for the purpose of interpreting statutes. It is therefore a mistake to interpret terms in a statute according to the meaning of those same words found in a general corpus.

1. Statutes and the Texts of a General Corpus Have Different Purposes

There are many ways, at many different levels of abstraction, to describe the purposes of statutory language.[156] But, it is not necessary to decide on the legitimacy of these purposes of statutory language, or prioritize them, to conclude that statutory language serves different purposes than the language found in the texts of a general corpus.[157]

Statutory language is *authored* for a different purpose than the type of language found in a general corpus. Most obviously, statutory language is written to prescribe behavior.[158] That is, statutes are written to control

---

to the context of searches in general corpora to find meaning in statutory texts. Because a search in a general corpus deliberately seeks meaning wholly outside of the legal context, the meaning returned by such a search will fail to reflect the unique features of statutory language.

[156] Jeremy Bentham described legislation's purpose as "the happiness of the body politic," which includes subsistence, abundance, equality, and security. JEREMY BENTHAM, BENTHAM'S THEORY OF LEGISLATION 123 (Etienne Dumont ed. & trans., 1914). This is the model of legislation that courts have in mind when they routinely uphold statutes addressing "broad and general social or economic problem[s]." Energy Reserves Grp. v. Kan. Power and Light, 459 U.S. 400, 411–12 (1983). Certainly, this public-regarding model of legislation is subject to challenge: while "some legislation serves legitimate public-regarding" goals, other legislation can "only be described as amorally redistributive." Jonathan R. Macey, *Promoting Public-Regarding Legislation Through Statutory Interpretation: An Interest Group Model*, 86 COLUM. L. REV. 223, 228 (1986) (citation omitted). For example, rather than providing a rule of general applicability, special legislation provides a rule that applies only to a particular individual, often providing special benefits to named individuals. Evan C. Zoldan, *Reviving Legislative Generality*, 98 MARQ. L. REV. 625 (2014); Evan C. Zoldan, *Legislative Design and the Controllable Costs of Special Legislation,* 78 MD. L. REV. 415 (2019) [hereinafter Zoldan, *Legislative Design*].

[157] Greenberg, *supra* note 154, at 233–40; *cf.* SLOCUM, *supra* note 90, at 5–6 (arguing that communicative content of statutory language is presumptively its meaning, although this presumption is defeasible).

[158] Fiss, *supra* note 4, at 751; Marmor, *supra* note 154, at 425 ("Legal norms prescribe modes of conduct, grant rights, [and] impose obligations . . . .").

conduct by providing rules of decision for individuals, administrative agencies, and courts to follow.[159] Because it is written to control conduct, statutory language, unlike the language found in a general corpus, uses constructions that are effective at requiring or prohibiting action or granting authority.[160] For example, statutes often prohibit conduct in the way that the Food, Drug, and Cosmetic Act (FDCA) begins its long list of prohibited acts: "The following acts and the causing thereof are prohibited . . . ."[161] Similarly, statutes vesting authority in agencies often begin the way that the FDCA vests rulemaking authority in the Secretary of Health and Human Services: "[t]he authority to promulgate regulations for the efficient enforcement of this chapter . . . is vested in the Secretary."[162]

Furthermore, statutory language is written for the purpose of making some change to the law. As a result, it is phrased to conform with legislative drafting conventions for lawmaking.[163] In Congress, for example, the legislature must use the following language to enact a statute: "Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled."[164] Moreover, unlike nonlegal language, legislation is written to classify future conduct or objects. As a result, it is written in general, prospective, impersonal language in order to encompass both conduct that is known and conduct that is unknown.[165] Again, the FDCA provides typical phrasing, prohibiting in general and impersonal terms "the adulteration or misbranding of any food, drug, device, tobacco product, or cosmetic . . . ."[166]

Because legislative drafters write language with the purpose of accomplishing some important real-world goal, they use language to achieve their desired results and minimize the damage of unintended consequences.[167] Sometimes, this purpose leads legislators to "seek to achieve a high level of explicitness and thus to minimize or perhaps even

---

[159] Marmor, *supra* note 154, at 425; Maurizio Gotti, *Text and Genre*, *in* THE OXFORD HANDBOOK OF LANGUAGE AND LAW, *supra* note 6, at 52, 63 (noting that function of legislative language is to "impose obligations and or confer rights").

[160] *See* Yon Malley, *The Language of Legislation*, 16 LANGUAGE & SOC'Y 28, 40 (1987).

[161] 21 U.S.C. § 331 (2018).

[162] *Id.* § 371.

[163] *See* Malley, *supra* note 160, at 30.

[164] 1 U.S.C. § 101 (2018).

[165] Malley, *supra* note 160, at 40. Rarely, legislation is purposefully written in language that is not general and prospective. For example, special legislation singles out a particular individual for special treatment. *See e.g.* Zoldan, *Legislative Design*, *supra* note 156 at 422.

[166] 21 U.S.C. § 331(b).

[167] Nicholas Allott & Benjamin Shaer, *Inference and Intention in Legal Interpretation*, *in* THE PRAGMATIC TURN IN LAW 116 (Jane Giltrow et al. eds., 2017).

eliminate implicated content . . . ."[168] Conversely, statutory language can also be unusually vague or ambiguous compared with ordinary language.[169] This general principle is borne out by empirical work on Congress. In their interviews of Capitol Hill staffers, Professors Victoria Nourse and Jane Schachter confirmed that ambiguity or vagueness is often a feature of legislative language, not a result of error in its drafting: where legislators harbor different policy opinions on a key point, they often agree to use ambiguous or vague language, each hoping that an agency or court will later resolve the uncertainty in his favor.[170]

Whether unusually explicit or unusually vague, legal language is often complex precisely because it has the purpose of addressing a complex social issue that cannot be reduced to simple language.[171] Consider the Affordable Care Act,[172] the Social Security Act,[173] or countless other transformative modern statutes. These statutes did not merely tinker with well-known common law concepts; they completely reorganized the relationship between the citizen and the state within their subject matters. Complex concepts, addressed by modern legislation, cannot be expressed in language that is simple enough for untrained people to understand while still accomplishing what it is supposed to accomplish.[174]

Conversely, attempting to render statutory language into words that can be readily grasped by a person without legal training may make a statute ineffective.[175] An example from Australia provides evidence of the challenges legislative drafters face when trying to draft statutory language as if it were nonlegal speech. In order to make the statute easily understood, Australian legislative drafters wrote the *Coroners Bill* in simple, nonlegal language. It did not take long, however, for lawyers and judges to determine that there were large, unintended gaps in the law. Moreover, many of the bill's ramifications were not obvious from the statute's language, including the consequences for failing to comply with

---

[168] *Id.*

[169] SLOCUM, *supra* note 90, at 196–97.

[170] Victoria F. Nourse & Jane S. Schachter, *The Politics of Legislative Drafting: A Congressional Case Study*, 77 N.Y.U. L. REV. 575, 595–96 (2002). *See also* Summers, *supra* note 45, at 243.

[171] Duncan Berry, *Legislative Drafting: Could Our Statutes Be Simpler*, 1987 STAT. L. REV. 92, 93 (1987).

[172] The Patient Protection and Affordable Care Act, Pub. L. No. 111–148, 124 Stat. 119 (2010).

[173] Social Security Act, Pub. L. No. 74–271, 49 Stat. 620 (1935).

[174] Rabeea Assay, *Can the Law Speak Directly to Its Subjects*, 38 J. L. & SOC'Y 376, 399–401 (2011); I.M.L. Turnbull, *Problems of Legislative Drafting*, 1986 STAT. L. REV. 67, 68 (1986).

[175] *See* Christopher Williams, *Legal English and Plain Language: An Introduction*, *in* ESP ACROSS CULTURES 111, 122 (2004).

the law.[176]

Not only is statutory language authored for a different purpose than nonlegal language, it is also *read* for a different purpose than nonlegal language. Because a reader of legislative language often has the goal of learning what conduct is prohibited or permitted,[177] an interpreter reads statutory text for a statement of a rule, the outer limits of the application of the rule, exceptions to the rule, similarities or differences in language between different parts of the text, and other features that are uniquely important for the purpose of learning what conduct is permitted or prohibited. By contrast, a person reading a play, poem, or other nonlegal text will not be reading it for these same purposes. As a result, a reader will interpret a word in a nonlegal text differently than she would interpret the very same word in a statute. As Professors McGinnis and Rappaport described this phenomenon, legal texts are read against background understandings about how the text should be read and interpreted.[178] Moreover, this point has been demonstrated experimentally: in one study, subjects were given a text and told that their purpose in reading it was "studying." Researchers determined that the subjects "employed stringent standards focused on intratextual relations, striving for deep understanding and coherence in their representation of the texts."[179] By contrast, subjects tasked to read the very same texts for the purpose of "entertainment" were "much less concerned with constructing a coherent representation of the text itself but instead focused more on connecting text events to their own personal experiences."[180] The study authors concluded that "when the text genre, reading task, and/ or reader motivation varies, readers systematically alter their criteria for comprehension and, hence, generate predictably different patterns of inferences."[181] In other words, simply having a different purpose leads subjects to think about, and ultimately interpret, a text differently. Because people read statutory language for different purposes than ordinary texts, the very same person is likely to systematically interpret the words of statutes differently than she would

---

[176]   Berry, *supra* note 171, at 101.

[177]   Gustavo Arosemena, *Human Rights*, *in* INTRODUCTION TO LAW 261 (2014) ("One natural way to look at the law is to see it as a collection of rules laid down by a competent authority that tell us in more or less concrete terms what we should do, what is required, prohibited and permitted.").

[178]   McGinnis & Rappaport, *supra* note 18, at 1340–41.

[179]   Paul Van Den Broek, Robert F. Lorch, Jr., Tracy Linderholm & Mary Gustafson, *The Effects of Readers' Goals on Inference Generation and Memory for Texts*, 29 MEMORY & COGNITION 1081, 1085 (2001).

[180]   *Id.*

[181]   *Id.* at 1082. *See also* Gregory Schraw & Rayne Sperling Dennison, *The Effect of Reader Purpose on Interest and Recall*, 26 J. READING BEHAV. 1, 14–15 (1994) (showing that differences in memory and interest follow from different reading goals).

interpret those same words when they are found in the texts of a general corpus.

2. Statutes and the Texts of a General Corpus Have Different Audiences

Texts, including statutory texts, are addressed to specific audiences.[182] The texts of a general corpus, like the COCA, include transcripts of "unscripted conversation from more than 150 different TV and radio programs," "[s]hort stories and plays from literary magazines, children's magazines," popular magazines covering subjects as diverse as "health, home and gardening, women, financial, religion, [and] sports," newspapers, and academic journals.[183] The diversity of these texts suggest that, if there is a single audience for the texts in the COCA, it is a general audience (perhaps an audience of hypothetical reasonable people) without a single, shared set of norms for interpreting language.[184] By contrast, the audience of statutory text always includes public officials, subject-matter experts, lawyers, and judges, all of whom interpret law in light of their professional roles and obligations. As a result, it should come as no surprise that general audiences are unable to understand statutory text.

i. The Audience of Statutes Always Includes Experts Interpreting in Their Official Capacity

The audience of statutes always includes experts who interpret statutory language in their official capacity. First, many statutes are addressed exclusively to users of language who have an institutional role in the interpretation and enforcement of the statute; indeed, these statutes do not even purport to act on individuals without an official interpretive role. As Edward Rubin described, statutes addressed to public officials, like regulators who supply the content of the law, dominate lawmaking.[185] According to Rubin, "[m]odern legislation in its essence is an institutional practice by which the legislature . . . issues directives to the governmental mechanisms that implement that policy."[186] That the audience of most modern statutes consists of government officials rather than the public in general is most clearly true with respect to statutes that vest broad authority

---

[182]   Malley, *supra* note 160, at 33.

[183]   CORPUS CONTEMPORARY AM. ENG., *supra* note 30 (describing texts and registers).

[184]   Lee & Mouritsen, *supra* note 7, at 793.

[185]   Edward L. Rubin, *Law and Legislation in the Administrative State*, 89 COLUM. L. REV. 369, 381–82, 404 (1989) (arguing that the ordinary citizen is not apprised of legal rules "by their verbal formulation in the statute books"); Ross, *supra* note 154, at 1057 (noting that non-criminal statutes are directed at "a small community of lawyers, regulators, and people subject to their specific regulations").

[186]   Rubin, *supra* note 185, at 372.

in administrative agencies to interpret and enforce the law.[187] Take, for example, the Federal Land Policy and Management Act, which governs how federal lands are managed. It directs federal agencies to establish federal land leasing programs,[188] maintain an inventory of public lands and their value,[189] dispose of or acquire land, and most broadly, "promulgate rules and regulations to carry out the purposes" of the statute.[190] This statute, and countless others like it, are addressed only to institutional actors (both lawyers and nonlawyers) who are steeped in the particular missions and vernacular of their agencies and who read statutory language in light of their roles, knowledge, and professional obligations. Statutory language vesting authority in institutional actors is addressed to these actors and reflects these roles, knowledge, and professional obligations. These statutes epitomize the legal language that Mellinkoff argued is "divorced from the common speech."[191]

Second, even statutes that act directly on individuals without an official interpretive role, including statutes with criminal penalties, are often addressed primarily or exclusively to subject-matter experts. Most saliently, a number of commentators have argued that fair notice considerations are most acute in the context of criminal laws that act directly on individuals.[192] But, it would be too facile to conclude that a statute has an audience of ordinary individuals rather than expert interpreters simply because it carries criminal penalties. The FDCA, for example, provides criminal penalties[193] for the commission of a long list of prohibited acts, including the adulteration and misbranding of food, drugs, cosmetics, tobacco products, and medical devices.[194] But, even though it prescribes criminal penalties for its violation, the FDCA is addressed primarily to officials of the Food and Drug Administration (FDA), which is vested with broad authority to promulgate regulations under the statute.[195] Perhaps most importantly, the FDA is authorized to define, and does define, important statutory terms, including determining what counts as misbranding[196] or adulteration.[197] Even to the extent that the audience of the FDCA includes individuals outside of the FDA, these individuals are,

---

[187] *Id*. at 381.
[188] 43 U.S.C. § 1703 (2018) (imposing obligations on agency).
[189] *Id.* § 1711 (imposing obligations on agency).
[190] *Id.* § 1740 (setting out rulemaking obligations).
[191] MELLINKOFF, *supra* note 129, at 18.
[192] Gries & Slocum, *supra* note 9, at 1427.
[193] 21 U.S.C. § 333 (2018).
[194] *Id.* § 331.
[195] *Id.* § 371.
[196] *E.g.*, 21 C.F.R. § 201.1(a) (2019).
[197] *E.g.*, 21 C.F.R. § 106.1(a).

like government regulators, experts in the statutory subject-matter rather than hypothetical reasonable people. This group includes lawyers, lobbyists, technical experts, scientists, and compliance officers employed by the small cohort of companies that are members of the pharmaceutical industry.[198] And the language used in the FDCA, like the language used in most statutes, reflects the fact that its primary if not exclusive audience is an audience of experts knowledgeable about the subject matter regulated by the statute.

Third, even if a statute can be said to be directed to nonexpert individuals—a situation most likely in the context of a simple rather than a complex statute—its audience is never *limited* to these ordinary individuals.[199] Consider one of the simplest federal statutes, reproduced below in full:

> Whoever falsely represents himself to be an officer, agent, or employee of the United States, and in such assumed character arrests or detains any person or in any manner searches the person, buildings, or property of any person, shall be fined under this title or imprisoned not more than three years, or both.[200]

Because the conduct proscribed is intuitively wrong, and because it is concisely written, perhaps this statute can fairly be characterized as having an audience that includes ordinary people without specialized training. Nevertheless, it would misunderstand the way the law is enforced to conclude that its audience is limited to nonexperts. The audience of this impersonation statute also includes: federal agents charged with enforcing the law, lawyers in the United States Attorney's Office who will decide whether to prosecute an accused offender, the accused's counsel, and the judge who will oversee the ensuing trial. All of these actors will interpret the statutory language in light of their background knowledge and professional obligations. All of these experts will have at least some specialized knowledge, which a person without special training lacks, about the contours of what is prohibited; for example, the boundaries of what constitutes an "arrest" or a "search" are notoriously technical.[201] Moreover, the judges and lawyers interpreting this statute will be constrained to interpret this language in accordance with professional standards and

---

[198] Ross, *supra* note 154, at 1061–62.

[199] MELLINKOFF, *supra* note 129, at 17–18; William N. Eskridge & Judith N. Levi, *Regulatory Variables and Statutory Interpretation*, 73 WASH. U. L. Q. 1103, 1010–11 (1995); Victoria F. Nourse, *Misunderstanding Congress: Statutory Interpretation, the Supermajoritarian Difficulty, and the Separation of Powers*, 99 GEO. L.J. 1119, 1142 (2011); Mattila, *supra* note 129, at 31.

[200] 18 U.S.C. § 913 (2018).

[201] Schauer, *supra* note 131, at 7–8 (noting that words like search and seizure have legal meanings that are both over- and under-inclusive of their ordinary meaning).

ethical obligations that bind lawyers and judges. Finally, these expert interpreters will need some knowledge of the broader legal regime in order to cross-reference this statute with other statutes to learn what fine might be applicable. Similarly, all statutes that are addressed to individuals untrained in the law are also addressed to those with legal or subject-matter training who will interpret the statute in light of their expertise and professional obligations.[202]

ii. Nonexpert Audiences Do Not Understand Legal Language

Perhaps the best evidence that statutory language has a different audience than the type of language found in a general corpus is the enduring difficulty that writers and speakers of legal language encounter when attempting to communicate with those not versed in the law.[203] Legal language, including statutory language, has long been criticized as being unintelligible to those untrained in the law. It has been called "elitist, bloated, and filled with gobbledygook" and "too dense and clouded for laypersons to understand."[204] Even undoubtedly sophisticated readers like Swift, Dickens, Jefferson, and Bentham all have criticized legal language as unintelligible.[205]

The unintelligibility of legal English is not a new phenomenon—for peculiar historical reasons, there has "never been a time since the Norman Conquest when the English of the law has been in tune with the common usage. It has always been considered a language apart."[206] But, although the reasons for the distinctiveness of legal English are ancient, "the gap between legal discourse and everyday discourse is still very wide. Present day legal discourse retains its identity as a highly specialised and distinctive discourse type or genre of English."[207] Jurors, for example, do not understand jury instructions, even when they think that they do,[208]

---

[202] MELLINKOFF, *supra* note 129, at 17–18 (discussing words that have special meanings to lawyers trained in the law); Marmor, *supra* note 154, at 437 ("[T]he legislature deliberately speaks with several voices."). Many of the arguments about audience can be made, perhaps with greater force, about interpretation of agency regulations. Even more so than statutes, regulations are usually addressed to industry insiders, defining statutory language that is itself directed to agency and industry insiders, and which draw on complex statutory and regulatory history.

[203] JOHN GIBBONS, LANGUAGE AND THE LAW 13 (John Gibbons ed., 1994); McGinnis & Rappaport, *supra* note 18, at 1338–39.

[204] Soha Turfler, *Language Ideology and the Plain Language Movement: How Straight-Talkers Sell Linguistic Myths*, 12 LEGAL COMM. & RHETORIC 195, 196 (2015).

[205] Williams, *supra* note 175, at 116.

[206] GIBBONS, *supra* note 203, at 11–12. *See also* Williams, *supra* note 175, at 116.

[207] GIBBONS, *supra* note 203, at 13.

[208] Bradley Saxton, *How Well Do Jurors Understand Jury Instructions? A Field Test Using Real Juries and Real Trials in Wyoming*, 33 LAND & WATER L. REV. 59, 67 (1998) (collecting and describing studies about jury comprehension); Walter W. Steele Jr. &

because jury instructions use "legal phrases and concepts that are foreign to the layperson."[209] Translators have found that translating legal texts is unlike translating nonlegal language. Because legal language is "complex and highly technical," "legal translation is generally recognized as the most complex and demanding of all areas of specialized translation."[210] As a consequence, it is not sufficient for legal translators to have language proficiency; they also must have "considerable specialist knowledge of both the source and target legal systems."[211] And the persistent unintelligibility of legal language to nonlawyers has given birth to "plain language" movements, both in the United States and abroad, which are dedicated to making legal language accessible to nonlegal audiences.[212] Despite some successes in the area of consumer contracts and agency guidance materials,[213] however, these efforts have not had an impact on legislative drafting in the United States.[214]

Of all legal language, statutory language has been called the most complex and esoteric, rendering it "incomprehensible to all except the specialist reader."[215] The influential Renton Committee, convened by the British Parliament to study statutory language, concluded that statutory language was impenetrable to ordinary citizens and might "as well be written in a foreign language."[216] Finally, and most tellingly, government entities themselves have acknowledged that they cannot communicate

---

Elizabeth G. Thornburg, *Jury Instructions: A Persistent Failure to Communicate*, 67 N.C. L. REV. 77, 81 (1988).

[209] Nancy Marder, *Instructing the Jury*, *in* THE OXFORD HANDBOOK OF LANGUAGE AND LAW, *supra* note 6, at 435, 439–40.

[210] Susan Šarčević, *Challenges to the Legal Translator*, *in* THE OXFORD HANDBOOK OF LANGUAGE AND LAW, *supra* note 6, at 187, 189.

[211] *Id*. Similarly, as McGinnis and Rappaport have suggested, nonlawyers recognize when they are reading legal language and defer to expert opinions, that is, lawyers, about it. McGinnis & Rappaport, *supra* note 154, at 765.

[212] Mark Adler, *The Plain Language Movement*, *in* THE OXFORD HANDBOOK OF LANGUAGE AND LAW, *supra* note 6, at 67, 82–83. Congress also acknowledged the need for a "plain language" movement when it enacted the Plain Writing Act of 2010, Pub. L. No. 111-274, 124 Stat. 2861 (2010) (requiring agencies to issue documents in plain language).

[213] Plain Writing Act of 2010, Pub. L. No. 111–274, 124 Stat. 2861 (2010). The Plain Writing Act requires agencies to write plainly and report back to Congress with the results of their efforts to do so. It also requires the Office of Management and Budget (OMB) to provide guidance to agencies to fulfill their statutory obligations. OMB directed agencies to a set of guidelines created by an organization called PLAIN, which provides guidance on writing for the intended audience and avoiding unnecessarily complicated language or legal, foreign, or technical jargon. *See, e.g.*, Rachel Stabler, *What We've Got Here Is a Failure to Communicate: The Plain Writing Act of 2010*, 40 J. LEGIS. 280, 294–95 (2014).

[214] *See* Williams, *supra* note 175, at 117–19.

[215] GIBBONS, *supra* note 203, at 25.

[216] DAVID RENTON, PARLIAMENT OF THE UNITED KINGDOM, THE PREPARATION OF LEGISLATION: REPORT OF A COMMITTEE APPOINTED BY THE LORD PRESIDENT OF THE COUNCIL 37 (1975).

statutory obligations effectively to their citizens through statutory language. Instead, it is common for government entities to provide the public with summaries of statutes or regulations as a means of educating them about the law's requirements.[217]   These publications summarize statutory language in narrative form or as bullet points or FAQs, provide rough definitions of legal terms, give examples to explain the intended meaning of statutory language, and even demonstrate statutory meaning with charts or pictures.[218]   To take just one example that includes many of these features, the United States Occupational Safety and Health Administration publishes a workers' rights pamphlet directed to nonlegal audiences, including summaries of statutes, rough definitions, examples, and narrative language, all intended to provide guidance in nonlegal language.[219]

3. Statutes Have Different Linguistic Characteristics than the Texts of a General Corpus

Because the purpose and audience of statutory language are different than that of language found in the texts of a general corpus, it is not surprising that statutory language has different linguistic characteristics than nonlegal language.[220] First, legal texts use words in unordinary ways. They contain "word usages that have no parallel in ordinary conversation,"[221] like interplead and demurrer. Legal language also preserves words and constructions that were once common in nonlegal speech but that are no longer current,[222] like therefor, whereas, and "comes now the plaintiff." Moreover, legal English contains frequent traces of

---

[217]   A special thanks to Larry Solan for suggesting this line of inquiry.

[218]   *See* Nicholas R. Parrillo, *Federal Agency Guidance and the Power to Bind: An Empirical Study of Agencies and Industries*, 36 YALE J. ON REG. 165, 167 (2018) ("Guidance comes in an endless variety of labels and formats, depending on the agency: advisories, circulars, bulletins, memos, interpretive letters, enforcement manuals, fact sheets, FAQs, highlights, you name it.").

[219]   U.S. DEP'T LAB. OCCUPATIONAL SAFETY AND HEALTH ADMIN., WORKERS' RIGHTS OSHA 3021-06R (2017), https://www.osha.gov/Publications/osha3021.pdf. *See also* MICH. LEGISLATURE, A PRACTICAL GUIDE FOR TENANTS & LANDLORDS (2017), https://www.legisla ture.mi.gov/Publications/tenantlandlord.pdf (providing a plain language "practical guide" for tenants and landlords designed to "inform tenants and landlords about their rights and responsibilities in rental relationships").

[220]   In more technical language, linguists might describe the different genres, registers, or domains that these different types of text occupy. David Y. W. Lee, *Genres, Registers, Text Types, Domains, and Styles: Clarifying the Concepts and Navigating a Path Through the BNC Jungle*, 5 LANGUAGE, LEARNING, AND TECH. 37, 37–41 (2001). *See* Tiersma, *supra* note 153, at 27–28 (noting that law is replete with technical vocabulary, archaic, formal and unusual terminology, and unusual grammatical constructions). *See also* Williams, *supra* note 175, at 112–13 (noting presence in legal language of foreign words and phrases).

[221]   McGreal, *supra* note 46, at 326. *See also* Tiersma, *supra* note 153, at 27–28.

[222]   MELLINKOFF, *supra* note 129, at 12–13; Mattila, *supra* note 128, at 32.

Latin and law French,[223] like fee simple and *ab initio*. Harder to spot, but no more intelligible than these foreign or archaic words, is legal language's frequent use of specialized terms and legal jargon,[224] like "four corners of the complaint," "lower court," and "damages."

Most commonly, and most relevant to the purposes of this Article, legal language uses common terms, but gives them meanings different from, and sometimes even at odds with, the same words as used in nonlegal speech.[225] As Mellinkoff described, legal language is characterized by the "frequent use of common words with uncommon meanings,"[226] like claim and discovery. It is this kind of language that creates the greatest risk of confusion when nonlawyers interact with the legal system. Whether words like "seizure,"[227] "reckless,"[228] "utter,"[229] and "consideration"[230] are used in a legal sense or nonlegal sense makes all the difference to whether one's rights have been violated or whether one is liable for punitive damages, has committed a crime, or has enforceable contract rights.

The differences between legal language and nonlegal language are more than just differences in vocabulary. Indeed, it is the differences in syntax and drafting conventions that "render[s] legislative texts incomprehensible to all except the specialist reader . . . ."[231] Legal language is more complex[232] and reflects a different "structure and arrangement of principal sentence elements" than nonlegal language.[233] Sentence length[234] and clause structure[235] also differ between legal and nonlegal language. Moreover, legal language uses polysemes (words or

---

[223]   MELLINKOFF, *supra* note 129, at 13–15.

[224]   *Id.* at 17–19.

[225]   Schauer, *supra* note 154, at 35-36 (many "legal" words also have ordinary meanings that are different than legal meaning); Poscher, *supra* note 131, at 132; Mattila, *supra* note 129, at 31 ("the use of ordinary words in a technical legal sense is particularly widespread"); Schauer, *supra* note 131, at 7–8 (words like search and seizure have legal meanings that are both over and under-inclusive of their ordinary meaning).

[226]   MELLINKOFF, *supra* note 129, at 11–12.

[227]   *See* 5 U.S.C. § 5545(a)(2)(A) (2018) (setting out obligation of federal officer to possess knowledge of "court decisions concerning . . . search and seizure").

[228]   *See* 30 U.S.C. § 1235(*l*) (2018) (defining gross negligence to include "reckless" behavior).

[229]   *See* 18 U.S.C. § 331 (2018) (imposing criminal liability for a person who "utters" an altered, defaced, or mutilated coin).

[230]   *See* 31 U.S.C. § 3727(d) (2018) (providing circumstances under which a contract may be changed without "consideration").

[231]   GIBBONS, *supra* note 203, at 25.

[232]   Gotti, *supra* note 159, at 53.

[233]   Risto Hiltunen, *The Grammar and Structure of Legal Texts*, *in* THE OXFORD HANDBOOK OF LANGUAGE AND LAW, *supra* note 6, at 39, 41.

[234]   Hiltunen, *supra* note 233, at 42; Gotti, *supra* note 159, at 53–54.

[235]   Hiltunen, *supra* note 233, at 43; Gotti, *supra* note 159, at 53.

phrases with different, but related senses) with greater frequency than nonlegal language.[236] At times, legal language can be unusually precise,[237] painstakingly including contextual knowledge that would be assumed in nonlegal speech.[238] Other times, legal language is unusually vague,[239] perhaps because legislators seek agreement on language even when they cannot agree on its meaning.[240] It is these variations from nonlegal syntax that serve "to distinguish the language of the law from the common tongue."[241]

Nonlegal language also has different drafting conventions than legal language. For example, as Tiersma described, nonlegal English often replaces a repeated noun with a pronoun; writers of legal language, by contrast, tend to repeat the noun. For example, it is common in legal English to write "Buyer promises that Buyer will pay," when one means that the same person is doing the promising and buying. By contrast, if a nonlegal English speaker writes "'John kissed John's girlfriend,' we normally assume that there are two distinct people named John."[242] Statutory language also abides by the convention of placing multiple related thoughts in the same sentence. This convention results in sentences that can run hundreds of words and is responsible for the tightly-packed character of statutory language.[243]

Third, although many of the above-described differences in word choice and syntax apply equally to statutory and non-statutory legal language, statutory language is especially different from the kind of language found in a general corpus. One reason is that the texts of a general corpus include transcripts of spoken language, which is fundamentally different from written language. *One-fifth* of the COCA consists of "unscripted conversation from more than 150 different TV and radio programs."[244] But, there are "fundamental differences between the interpretation of verbal utterances and texts."[245] Among other differences, oral interlocutors share time and space, creating a great deal of shared context that does not appear in the words themselves.[246] For this reason,

---

[236] Mattila, *supra* note 129, at 30.

[237] MELLINKOFF, *supra* note 129, at 21–22.

[238] Marmor, *supra* note 154, at 425; Allott & Shaer, *supra* note 167, at 115–16.

[239] Poscher, *supra* note 131, at 134.

[240] William N. Eskridge, *Dynamic Statutory Interpretation*, 135 U. PA. L. REV. 1479, 1532 (1986–87); Nourse & Schachter, *supra* note 170, at 595–96.

[241] MELLINKOFF, *supra* note 128, at 23.

[242] Tiersma, *supra* note 153, at 30.

[243] GIBBONS, *supra* note 203, at 25.

[244] CORPUS CONTEMPORARY AM. ENG., *supra* note 30.

[245] SLOCUM, *supra* note 90, at 43.

[246] SLOCUM, *supra* note 90, at 43–50.

Professor Brian Slocum notes that "ordinary conversations are a poor model for the interpretation of legal texts because the context of interpretation of oral statements differs so greatly from the context of interpretation of legal texts."[247]

In addition, statutory language is special, even when compared with other legal language. As Tiersma notes, statutory language has "its own relatively rigid format and sometimes requires specific forms of language."[248] As a result, even when concluding that legal English as a whole is not a distinct language from nonlegal English, Tiersma notes that the two are most similar when nonlegal English is rendered in highly formal, written prose. The formality of some nonlegal English, however, stands in sharp contrast with even the written texts of a general corpus. The COCA, for example, includes not only spoken language, but informal written speech, like the text of children's magazines.[249] The NOW Corpus contains not only news sources, but also online magazines with subjects as diverse as video games, cricket, and fashion from speech communities markedly different from the United States.[250] And the COHA includes, among other texts, movie scripts and poetry.[251] Because a general corpus includes spoken and informal written language, and because of the special nature of statutes, whatever similarities there are between legal English and nonlegal English more generally, statutory language and the language of the texts of a general corpus are considerably less similar.

4. A General Corpus Should Not Be Used to Interpret Statutory Language

It is never appropriate to search for statutory meaning in a general corpus. As described above, statutory language and the language found in a general corpus differ in purpose, audience, and linguistic characteristics. In other words, statutory language and the texts of a general corpus do not share the same *context*. As a result, an interpretation of a statute according to the meaning of language in a general corpus is lacking the *statutory context* that is necessary to understand statutory meaning. An interpretation of a statute without statutory context, in other words, fails to reflect the meaning attributable to the fact that a statute is statutory language as opposed to a novel, poem, or some other nonstatutory text.[252]

---

[247] SLOCUM, *supra* note 90, at 14; *see also* Tiersma, *supra* note 153, at 27; Allott & Shaer, *supra* note 167, at 115–16; Gillian Brown, *Understanding Spoken Language*, 12 TESOL Q. 271 (1994).

[248] Tiersma, *supra* note 153, at 30.

[249] CORPUS CONTEMPORARY AM. ENG., *supra* note 30.

[250] NOW CORPUS (NEWS ON THE WEB), *supra* note 115.

[251] CORPUS HISTORICAL AM. ENG., *supra* note 115.

[252] For a contrary view, *see* Brian G. Slocum, *Pragmatics and Legal Texts: How Best to*

Moreover, the differences between statutory language and nonlegal language are pervasive rather than exceptional. That is, the differences described above suggest that statutory language is something other than nonlegal language sprinkled with occasional legal terms of art. Instead, it makes more sense to think of statutes as written in a different dialect or sublanguage—statutory language—albeit one that contains both words that differ obviously from their use in nonlegal language and also words that mean the same thing as their cognates in nonlegal language.[253]

Consider the following non-statutory text:

(1) When I lived in London, I rented a flat overlooking the Thames.

Here, the word "flat" is used differently than Americans would use that word. When reading (1), an American reader would simply substitute the word "apartment" for "flat" and read the rest of the sentence without much effort. Nevertheless, it does not follow that the sentence is written in American English with one word, flat, written in British English. Instead, depending on the context of the utterance (including the purpose, audience, and surrounding texts), it might make more sense to conclude that the whole sentence is written in British English, a separate dialect with many cognates in American English. Similarly, it would be a mistake to read a statute and conclude, because much of it can be read with little effort, that it is written in nonlegal language except for the few words that appear to be legal terms of art.[254] Better, for all the reasons described above, statutory text should be considered a dialect of natural language (or a "sublanguage," as Tiersma put it),[255] statutory language, whose meaning cannot be determined simply by importing the meaning of its words from nonlegal

---

Account for the Gap Between Literal Meaning and Communicative Meaning, *in* THE PRAGMATIC TURN IN LAW: INFERENCE AND INTERPRETATION IN LEGAL DISCOURSE 140 (Janet Giltrow & Dieter Stein eds., 2017) (disagreeing with the proposition that "the draftors of legal texts, particularly statutes, do not use language in the same ways as do others, and that these differences preclude the applicability of conversational implicatures."). *See also* Fallon, *supra* note 151, at 331–33 (arguing that legal language is not a language in the same sense as natural language); Summers, *supra* note 45, at 234 (arguing that interpreting statutory language does not require a person to "leave the world of ordinary language and enter a specialized legal world governed by some special tongue").

[253]   Schauer, *supra* note 131, at 19–20. *See also* McGinnis & Rappaport, *supra* note 18, at 1377 (arguing that the structure of a document, not simply its words, determine whether it is written in technical or ordinary language).

[254]   Tiersma, *supra* note 153, at 29 (it "would be the wrong lesson to draw" that "legal language is nothing more than ordinary English with a lot of specialized vocabulary").

[255]   Tiersma, *supra* note 153, at 31. *Cf.* Fallon, *supra* note 151, at 331–32. Fallon argues, quite reasonably, that legal language is not independent of natural language and does not share its status as an independent language. But, simply because legal English differs from spoken English less than, say, French, it does not follow that legal English is similar to spoken English in a way that is relevant to statutory interpretation.

texts. Instead, a statute should be read as statutory language, with all of the distinctive features of this language, including its unique purposes, audience, word choice, syntax, and other conventions.[256]

Interpreted in this manner, the meaning of statutory language is always a distinctively *legal* meaning. In some cases, giving the words in a statute their legal meaning yields the same interpretation as giving its words their nonlegal meaning; in other cases, recognizing the distinctiveness of statutory language leads to a different interpretation. Either way, a reader who interprets statutory language as something distinct from nonlegal language recognizes that the project of interpreting legal language is not just one of translating individual technical words and phrases and inserting them into an otherwise nonlegal text. By contrast, when users of corpus linguistics techniques search for statutory meaning in general corpora, they risk missing the meaning that is attributable to the statutory context. As a result, corpus users, in a real sense, misconstrue the language of the text they interpret.[257] Because it is an error to interpret even nontechnical statutory language according to its nonlegal meaning, it is never appropriate to search in a general corpus for statutory meaning.

The distinction between nonlegal meaning and statutory meaning leads to one final point: it is possible to reject the claim that statutory language should be interpreted according to its nonlegal meaning and still accept the claim that statutory language should be interpreted according to its ordinary meaning, properly understood.[258] This is true because ordinary meaning is not synonymous with nonlegal meaning. As Professor David Strauss has noted, the ordinary meaning of statutory language "is not so ordinary." Rather, it is better thought of as "ordinary *legal* meaning," as distinguished from ordinary (nonlegal) meaning.[259] Consider, for example, a prohibition on "uttering" in a statute concerning securities fraud. The ordinary meaning of a prohibition on "uttering" in a statute about securities

---

[256]   Schauer, *supra* note 131, at 19–20. *See also* Fiss, *supra* note 4, at 744 (arguing that an interpreter is constrained by "a set of rules that specify the relevance and weight to be assigned to the material").

[257]   To be clear, I am not suggesting that statutory language is a literally a distinct *language* from ordinary English. As Schauer sensibly noted, "[l]egal English is related to ordinary English in ways that Estonian is not." Schauer, *supra* note 154, at 36. But, the differences are clear enough to justify Tiersma's judgment that legal language is a "sublanguage" of English, falling "somewhere between a separate language and ordinary English." Tiersma, *supra* note 153, at 31.

[258]   Frederick Schauer, *Statutory Construction and the Coordinating Function of Plain Meaning*, 1990 SUP. CT. REV. 234, 234 n.6 (noting that it is "implausible" that plain meaning "must necessarily be the same as ordinary non-technical meaning").

[259]   Strauss, *supra* note 154 at 1568. *Cf.* SLOCUM, *supra* note 90, at 12–13, 179–80 (distinguishing ordinary legal meaning from ordinary meaning).

fraud includes passing off a worthless check as genuine,[260] including handing a counterfeit check to a bank cashier.[261] By contrast, an interpretation of the term "utter" according to its *nonlegal* meaning (that is, the meaning found in a general corpus), would attribute to it a meaning involving making a sound with one's voice or perhaps a meaning involving completeness or totality.[262] As this example demonstrates, the ordinary meaning of a statutory term can differ from its nonlegal meaning. Because nonlegal meaning is not coextensive with ordinary meaning, rejecting the use of corpus linguistics for statutory interpretation does not entail rejecting the common preference for interpreting texts according to their ordinary meaning.[263]

This section has described the subjective and nontrivial choice that a user of corpus linguistics techniques makes when choosing a corpus to search for statutory meaning. The choice is subjective because corpus users do not adhere to standards for choosing a corpus and perhaps cannot articulate a persuasive test distinguishing between ordinary terms and legal terms of art. Moreover, and perhaps more importantly, it is never appropriate for a user of corpus linguistics for statutory interpretation to rely on a general corpus. Because legal language is different than nonlegal language—in purpose, audience, word choice, syntax, and drafting conventions—the meaning of statutory language is not the same, in a relevant way, as the meaning of cognate words found in a general corpus.

C. *Is a "Legal Corpus" the Answer?*

Because the use of corpus linguistics techniques for statutory interpretation has focused, virtually exclusively, on searches in general corpora, the mismatch between statutory language and nonlegal language is the most pressing theoretical problem facing courts and commentators relying on corpus techniques to interpret statutes. In response to these problems, one might suggest that a "legal corpus" could be constructed— that is—a corpus populated by legal texts rather than nonlegal texts.[264]

---

[260] 18 U.S.C. § 513 (2018) (penalizing a person who "utters or possesses a counterfeited security . . . [or who] makes, utters or possesses a forged security . . . with intent to deceive another person"). United States v. Peters, 462 F.3d 953, 959 (8th Cir. 2003) (upholding conviction for "uttering" where defendant "deposited, and substantially spent, a check that he knew was counterfeit into his bank account").

[261] *Peters*, 462 F.3d at 953.

[262] A search for "utter" in the COCA, for example, returns many hits that reflect these nonlegal meanings. CORPUS CONTEMPORARY AM. ENG., *supra* note 30 (search for "utter").

[263] *E.g.*, Eskridge, *supra* note 43, at 538–39.

[264] Phillips, Ortner & Lee, *supra* note 8, at 24. *See also* Solan, *supra* note 42, at 59–60 (noting that searching a corpus designed to reflect ordinary meaning is not very useful for determining the meaning of terms of art). *Cf.* Heilpern, *supra* note 128, at 380 (suggesting that technical terms of art—but not legal terms of art—should be interpreted according to

Even if it is inappropriate to interpret statutes according to their nonlegal meaning, the argument might go, this is not a problem with corpus linguistics techniques per se, but rather with the choice of corpus. A corpus user, therefore, could search a legal corpus to interpret statutory language, eliminating the problems associated with nonlegal language. Even on its face, however, this response is inadequate to eliminate the subjectivity inherent in corpus use. Because legal terms cannot be separated neatly from nonlegal terms in the same text, the "difference between legal terms and words of ordinary language is relative and hard to define."[265] As a result, a corpus user would have no objective way to choose between a legal corpus and general corpus for many statutory terms. The construction of a legal corpus, therefore, would not eliminate the subjectivity of the choice of corpus.

Even assuming that there is a way to distinguish legal words from nonlegal words sufficient to meet this objection, there are a number of other reasons why a search in a legal corpus cannot help uncover interpretations of statutory language objectively. First, even if a corpus user determined that a statutory term should be given its legal meaning, words used in statutory language often have more than one legal meaning. A "claim" means something wholly different in the context of patent law,[266] civil procedure,[267] and government contracts.[268] And "discharge" of a firearm[269] is not the same as the "discharge"[270] of a pollutant or discharge

---

meaning found in technical documents).

[265] Mattila, *supra* note 129, at 31. *See also* TIERSMA, *supra* note 129, at 108 (the distinction between terms of art and legal jargon "is mainly one of degree"); MELLINKOFF, *supra* note 129, at 17.

[266] Markman v. Westview Instruments, Inc., 517 U.S. 370, 373–74 (1996) ("[A] patent includes one or more claims, which particularly poin[t] out and distinctly clai[m] the subject matter which the applicant regards as his invention. A claim covers and secures a process, a machine, a manufacture, a composition of matter, or a design, but never the function or result of either, nor the scientific explanation of their operation. The claim define[s] the scope of a patent grant, and functions to forbid not only exact copies of an invention, but products that go to the heart of an invention but avoids the literal language of the claim by making a noncritical change.") (internal citations and quotes omitted).

[267] A claim is a set of facts for which the law provides redress, an innovation in civil procedure designed to avoid "the unfortunate rigidity and confusion surrounding the words cause of action." 5 FED. PRAC. & PROC. CIV. § 1216 (3d ed.).

[268] 31 U.S.C. § 3729 (2018) ("[C]laim . . . means any request or demand, whether under a contract or otherwise, for money or property and whether or not the United States has title to the money or property, that is presented to an officer, employee, or agent of the United States.").

[269] State v. Rasabout, 356 P.3d 1258, 1261 (Utah 2015) (Lee, J., concurring) ("the allowable unit of prosecution for unlawful discharge of a firearm is each discrete shot").

[270] 33 U.S.C. § 1362(12) (2018) ("discharge of a pollutant" means "any addition of any pollutant to navigable waters from any point source . . . .").

from the hospital.[271] Each of these terms has multiple distinct meanings; as a result, searching for the meaning of any of these terms in a legal corpus would provide no more precision than consulting a legal dictionary and do nothing to reduce the subjectivity of the choice.[272] Take, for example, the use of the term "claim" in the Fraud Enforcement and Recovery Act (FERA). FERA amended the False Claims Act (FCA), retroactively, to include "all *claims* under the False Claims Act" pending as of a given date.[273] On one hand, the "claims" referred to in FERA might be the type of claims that are the subject of the FCA, that is, demands for money from the federal Treasury.[274] On the other hand, FERA's "claims" might refer to lawsuits *by* the government under the FCA, that is, the civil procedure meaning of the term.[275] The difference between these two interpretations has real-world significance because each interpretation reaches a different set of cases. Nevertheless, results returned by a search in a legal corpus would shed no light on which of these two distinct meanings is the meaning of "claims" in FERA.

Second, a possible response to the problem of multiple legal meanings would be to construct multiple subject-matter-specific legal corpora.[276] For example, if a lawyer was interpreting the term "discharge" in an environmental case, the argument might go, she could simply search an environmental law corpus rather than a criminal law corpus for the term. But, multiple subject-matter-specific corpora would not help a corpus user interpret a statute in an objective way. This proposed solution presupposes that there are relatively firm legal categories, like "environmental law" and "criminal law"; but, of course, legal categories are not nearly as distinct as this proposed response suggests. For example, "discharge" of a pollutant is an environmental crime.[277] There is not, therefore, a firm line between hypothetical environmental and criminal corpora. Similarly, there is not a

---

[271] United States *ex rel.* Worthy v. E. Me. Healthcare Sys., No. 2:14-CV-00184-JAW, 2017 WL 211609, at *8 (D. Me. 2017) (noting different types of hospital discharge statuses, each requiring different codes).

[272] Moreover, both common law and civil law traditions have terms of art, but their drafting conventions, resulting from their different histories, are different. Gotti, *supra* note 159, at 58.

[273] Fraud Enforcement and Recovery Act of 2009, Pub. L. 111-21, § 4(f), 123 Stat. 1617, 1625 (2009) (emphasis added).

[274] Matthew Titolo, *Retroactivity and the Fraud Enforcement and Recovery Act of 2009*, 86 IND. L.J. 258, 268–69 (2011).

[275] *Id.*; Sanders v. Allison Engine Co., 703 F.3d 930, 937 (6th Cir. 2012).

[276] *Cf.* Heilpern, *supra* note 128, at 380 (suggesting that technical terms of art—but not legal terms of art—should be interpreted according to meaning found in technical documents).

[277] 33 U.S.C. § 1311 (2018) (prohibiting discharge of effluent); *id*. § 1319 (providing criminal penalties for discharge of effluent).

firm line between other hypothetical subject-matter-specific legal corpora.

Moreover, even if an interpreter were certain about which legal corpus to choose, a search in a subject-matter-specific legal corpus provides no way to choose among closely related legal meanings of words, all of which would be found in the same subject matter corpus. Polysemes—words or phrases with different, but related senses—occur frequently in legal language.[278] Because they are closely related, legal polysemes would be found in the same specialized legal corpus. As a result, choosing a particular subject-matter-specific legal corpus would do nothing to help determine which of two related possible meanings is the meaning of a statutory term. Consider, for example, the ubiquitous legal term "jurisdiction." It can refer to the particular physical territory where a body of law governs (like Congress's exclusive jurisdiction over the District of Columbia),[279] the power of the court to exercise authority over particular types of disputes (subject matter jurisdiction),[280] the power of the court to exercise authority over a particular individual (personal jurisdiction),[281] or the power of the court to hear a case in a particular procedural posture (original vs. appellate jurisdiction).[282] Because these polysemes would all likely be in the same hypothetical subject-matter-specific corpus, the construction of subject-matter-specific corpora would not help an interpreter choose an interpretation objectively.

V. SUBJECTIVITY AND CORPUS USAGE REVISITED

Part IV identified the choice of corpus as a key point in the interpretive process at which corpus users must make a subjective choice. Because the stated goal of corpus users is to reduce subjectivity, recognizing the subjectivity of the use of corpus linguistics in statutory interpretation should give legal interpreters pause before they adopt corpus methods for an interpretive decision. This Part will make the stronger claim that corpus use for statutory interpretation, as described above, is subjective *in the same way* as the interpretive methods that corpus users criticize. This demonstration suggests that corpus methods do not add value to the interpretive process at all.

Specifically, this Part will connect the types of subjectivity identified by proponents of corpus linguistics with the types of subjectivity that attach to corpus use itself. As Part III showed, proponents of corpus linguistics techniques for statutory interpretation roundly criticize two main sources of

---

[278] Mattila, *supra* note 129, at 30.
[279] U.S. CONST. art. I, § 8, cl. 17.
[280] *E.g.*, 28 U.S.C. § 1331 (2018).
[281] *E.g.*, FED. R. CIV. P. 4(k).
[282] U.S. CONST. art. III, § 2, cl. 2.

subjectivity: reliance on the intuition of the interpreter and reliance on biased reference data. In order to evaluate corpus users' claim that corpus use is less subjective than other methods of interpretation, this Part will compare the subjectivity of corpus analysis with corpus users' critiques of other methods of interpretation. The comparison demonstrates that the subjectivity of the choice of corpus identified in Part IV is also rooted in the intuition of the interpreter and in reliance on biased reference data. This conclusion undercuts the claim that corpus linguistics is more objective than the methods of interpretation it critiques.

A. *Intuition and Corpus Linguistics*

A choice of corpus relies on an intuition about what kind of word is being interpreted—that is—whether the word is an ordinary term or, by contrast, a specialized legal term. As noted, corpus users have not articulated, and likely cannot articulate, a persuasive account describing when a word is a legal term of art rather than an ordinary term.[283] Nevertheless, users of corpus techniques regularly conclude, without analysis, that terms like "results in,"[284] "information,"[285] and "harbor"[286] are ordinary words rather than legal terms. In the absence of an explanation, and given the lack of a pattern of use that would suggest the application of a standard,[287] it appears that it is the interpreters' intuition, rather than any theory or replicable data, that is being consulted to make this ordinariness determination.

Moreover, even when corpus users acknowledge that some words are terms of art, they rely on their intuition to determine that a particular word is ordinary. Consider again the *Rasabout* case, in which Justice Lee criticized the majority for relying on intuition to determine that the statutory term "discharge" roughly means "shoot."[288] Lee argued that the equation of discharge with shoot may be correct, but it is based on intuition rather than data.[289] But, the same can be said for Lee's determination that "discharge" is an ordinary term rather than a term of art. Rather than offering a reason why discharge is an ordinary word, he gave only a reason for not considering the question. Specifically, he noted that "no one has proffered the view that *discharge* is a legal term of art subject to

---

[283] *See* Part IV.A.
[284] Wilson v. Safelite Grp., Inc., 930 F.3d 429, 439, 440 (6th Cir. 2019) (Thapar, J., concurring).
[285] People v. Harris, 885 N.W.2d 832, 839 n.33 (Mich. 2016).
[286] Lee & Mouritsen, *supra* note 7, at 812.
[287] *In re* Adoption of Baby E.Z., 266 P.3d 702, 724 (Utah 2011) (searching the COCA after identifying a word as a term of art).
[288] State v. Rasabout, 356 P.3d 1258, 1274–75 (Utah 2015) (Lee, J., concurring).
[289] *Id.* at 1274–75.

specialized meaning in the law. Everyone agrees that this term is being used in its ordinary sense."[290] In other words, Lee relied on his own intuition (along with the intuition of the majority) to conclude that "discharge" was used in its ordinary sense. But, aggregated intuitions, like anecdotes, are not data. Lee's intuition may be "correct" in the sense that it matched the unverified intuition of others, but it is no more rooted in objective data than the interpretation he criticized.[291]

B. *Biased Reference Data and Corpus Linguistics*

The choice of a general corpus for statutory interpretation entails the reliance on biased reference data. A general corpus, as described above, is designed to represent nonspecialized speech by containing texts that use language in nonspecialized circumstances.[292] The COCA's 560 million words, for example, include transcriptions of spoken language, fiction, popular magazines, newspaper articles, and the like.[293] The use of language in a general corpus is biased with respect to statutory language because it shares none of the same context—that is, purpose, audience, word choice, and syntax—as statutory language.

To make this point clear, consider corpus users' criticism of dictionary use for statutory interpretation. Corpus users criticize dictionary use for statutory interpretation purposes because dictionaries lack the context of ordinary language.[294] By providing definitions of words without accompanying context, they argue, dictionaries systematically underreport ordinary uses of words. This flaw, corpus users suggest, biases dictionaries in favor of atypical usages, leaving even the good-faith interpreter unable to rely on dictionary usage without the risk of reaching atypical results.[295]

But, if the dictionary's disease is its lack of context, the corpus cure is

---

[290] *Id*. at 1287.

[291] Similarly, in the context of constitutional interpretation, Phillips, Ortner, and Lee assert that "corruption of blood" is a legal term of art while "commerce" is an ordinary word. Phillips, Ortner & Lee, *supra* note 8, at 24. Their only explanation for why "commerce" is an ordinary word is that "it makes no sense, and completely undermines the premise of ordinary public meaning, to argue that because a word or phrase is used in a legal document it automatically has a specialized legal sense." *Id.* This explanation, even if true, indicates only that not every word in a legal document is a term of art. It does not describe why commerce is ordinary, a conclusion that appears based on intuition rather than objective data. *See also* Ilya Somin, *Originalism and Political Ignorance*, 97 MINN. L. REV. 625, 651–52 (2012) (noting that terms like Bill of Attainder, privileges and immunities, and corruption of blood are terms of art, but words like liberty, property, and commerce are ordinary words).

[292] Gries & Slocum, *supra* note 9, at 1441; Lee & Mouritsen, *supra* note 7, at 828–29.

[293] CORPUS CONTEMPORARY AM. ENG., *supra* note 30.

[294] Gries & Slocum, *supra* note 9, at 1441.

[295] S*ee, e.g.*, Craig v. Provo, 389 P.3d 423, 428 n.3.

worse. By relying on words in a general corpus, a corpus user will interpret statutory language as if it were used in the same types of contexts as spoken language, newspaper articles, poetry, screen plays, magazines, and the like. Because statutory language is written and read for different purposes, reflects different audiences, and has different linguistic characteristics than nonlegal language, interpreting statutory language by reference to the use of words in nonlegal language does not place statutory language in context. Rather, corpus use is *radically acontextual*, divorcing statutory language from its distinctly legal context and guaranteeing interpretations that do not reflect the unique characteristics of statutory language.[296]

Indeed, the radical acontextuality that comes from interpreting statutory language as nonlegal language has been expressly embraced by corpus users. In his *Rasabout* dissent, Justice Lee considered, and rejected, the suggestion that statutory language should be interpreted in the context of other statutory language. The *Rasabout* majority criticized Lee's use of a corpus containing no statutory text for excluding the "the only speaker that matters," that is, the legislature.[297] Lee responded that the text of the Utah Code was *not* an appropriate corpus for determining the meaning of a term in the Utah Code.[298] Lee's response suggests that the acontextuality of searching a general corpus for statutory language is purposeful rather than an oversight.

VI. CONCLUSION

The dream of objectivity has driven the adoption of corpus linguistics techniques by commentators and a growing number of judges. The timing of this move is not surprising. One reason is obvious: technological advances have made corpus linguistics methods accessible to legal interpreters without specialized linguistics training. But, there is another reason why courts and commentators have been quick to adopt this new methodology. The American legal profession, by many accounts, is turning self-consciously toward formalism. As legal interpreters seek to

---

[296] For more on context, corpus linguistics, and ordinary meaning, *see* Slocum, *supra* note 17, at 13. What should be most alarming to interpreters considering a general corpus for statutory interpretation is the fact that it might contain language outside the context of American English altogether. The NOW Corpus, for example, contains texts that come not only from United States, but from markedly different linguistic communities, like India, Nigeria, Singapore, Kenya, Pakistan, and the Philippines, among others. Although a corpus user can exclude this data from a search, a default search in the NOW Corpus contains language from these countries. Interpreting statutes by relying on corpus data that includes English words used in foreign countries gives rise to significant democratic accountability problems.

[297] State v. Rasabout, 356 P.3d 1258, 1266 (Utah 2015).

[298] *Id.* at 1287 (Lee, J., concurring).

minimize their role—or the appearance of their role—in the interpretive process, the search for an objective tool of interpretation becomes more attractive. Seen in this light, the problem with corpus linguistics techniques is not their subjectivity, which may well be an inevitable part of the interpretive process, but the erroneous claim that they are superior because of their objectivity. Corpus linguistics will continue to be appealing to legal interpreters seeking to demonstrate their self-restraint; but its proponents should take care not to dismiss traditional methods of statutory interpretation, with all their warts, to chase a merely evanescent dream of objectivity.