2021

Adding Context and Constraint to Corpus Linguistics

Jeffrey W. Stempel
University of Nevada, Las Vegas -- William S. Boyd School of Law

Adding Context and Constraint to Corpus Linguistics

Jeffrey W. Stempel†©

INTRODUCTION: CORPUS LINGUISTICS AS POTENTIAL AND PITFALL

The segment of this Symposium from which this paper emerges was titled "Alternatives to Corpus Linguistics."[1] Notwithstanding the title of the panel,[2] I continue to think that a better title for the segment would have been "Supplements to Corpus Linguistics." The corpus linguistics movement provides a

---

† Doris S. & Theodore B. Lee Professor of Law, William S. Boyd School of Law, University of Nevada Las Vegas. Thanks to Bill Boyd, Dan Hamilton, Ted Lee, David McClure, Ann McGinley, Larry Solan, and Symposium participants, particularly to my colleagues on the "Alternatives to Corpus Linguistics" panel.

[1] *See* Symposium, *Data-Driven Approaches to Legal Interpretation*, 86 BROOK. L. REV. 291 (2021) (Panel II: "Alternatives to Corpus Linguistics"). Although addressing corpus linguistics in particular, this article's assessments are in my view equally applicable to other forms of textual data analysis such as Word Vector and algorithms designed to assess word meaning.

[2] *See generally* Anya Bernstein, *What Counts as Data?*, 86 BROOK. L. REV. 435 (2021) (arguing that corpus linguistics "reliably provides neither (1) evidence of ordinary understandings [of words used in statutes] nor (2) evidence of ordinary use [of the language uses in statues]" as "ordinary readers don't read statutes" in the way "they read novels and newspapers or listen to shows." Corpus linguistics thus over-emphasizes individual words and under-emphasizes the social force and effects of law); Shlomo Klapper, *Mechanical Turk Jurisprudence*, 86 Brook. L. Rev. 291 (2021) (noting the limits of corpus linguistics, in particular the "absence" problem of corpora failing to reflect particular terms that are logically related to words in a legal document and emphasizing empirical limits of survey research regarding word meaning); Kevin Tobia & John Mikhail, *Two Types of Empirical Textualism*, 86 Brook. L. Rev. 461 (2021) (using survey questions to determine lay concept of "because of" causation and implications for assessing the different versions of textualism reflected in majority and dissenting opinions in Bostock v. Clayton County, 140 S. Ct. 1731 (2020) finding Title VII against job discrimination for gay and transgender persons); Stefan Th. Gries, *Corpus Linguistics and the Law: Extending the Field from a Statistical Perspective*, 86 BROOK. L. REV. 321(2021) (noting methodological complexity of corpus linguistics and its over-simplification by judges and lawyers); Daniel Keller & Jesse Egbert, *Hypothesis Testing Ordinary Meaning*, 86 BROOK. L. REV. 489 (2021) (drawing distinction between high-confidence and low-confidence corpus findings and proposing hypothesis testing method for distinguishing); Stephen C. Mouritsen, *Natural Language and Legal Interpretation*, 86 BROOK. L. REV. 533 (2021) (comparing corpus linguistics and survey research as means of assessing language use and noting limitations of latter approach); Brian G. Slocum, *Big Data and Accuracy in Statutory Interpretation*, 86 BROOK. L. REV. 357 (2021) (distinguishing between ordinary, communicative, and legal meaning and suggesting different data analysis approaches for determining these different categories of meaning).

promising tool for legal analysis and should not be replaced per se with "alternatives." It should, however, be supplemented by consideration of other interpretative factors and indicia of meaning. In addition, the varying situational benefits of the doctrine should be better appreciated. Potentially valuable in many contexts, corpus linguistics may add little or nothing in others or may even be misleading. And as critics have noted, if embraced as an exclusive or fundamentalist tool for deciding cases, the corpus linguistics methodology could have pernicious effects.[3]

Corpus linguistics[4] standing alone is cabined by the same limits and imperfections of textualism standing alone in that it

---

[3] For criticisms of corpus linguistics as thus far applied to legal analysis, see, e.g., Anya Bernstein, *Legal Corpus Linguistics and the Half-Empirical Attitude*, 106 CORNELL L. REV. (forthcoming 2021) [hereinafter Bernstein, *Half-Empirical Attitude*]; John S. Ehrett, *Against Corpus Linguistics*, 108 GEO. L.J. ONLINE 50 (2019); Shlomo Klapper, *(Mis)Judging Ordinary Meaning?: Corpus Linguistics, the Frequency Fallacy, and the Extension-Abstraction Distinction in "Ordinary Meaning" Textualism*, 8 BRIT. J. AM. LEGAL STUD. 327 (2019); Kevin Tobia, *Testing Ordinary Meaning*, 134 HARV. L. REV. 726 (2020); Evan C. Zoldan, *Corpus Linguistics and the Dream of Objectivity*, 50 SETON HALL L. REV. 401 (2019). Professor Bernstein is careful to distinguish the "corpus linguistics in linguistics" from "corpus linguistics in the law," which is the target of her criticism. *See* Bernstein, *Half-Empirical Attitude*, at 6–16.

[4] In this small addition to a Symposium focusing largely on corpus linguistics, it seems unnecessary to identify and explain the concept. But to clarify terminology, when speaking of corpus linguistics, I refer to the school of thought that advocates analysis of a large database reflecting language use and analysis of that database to aid understanding and interpretation of words. Corpus linguistics seeks to examine a massive number of writings to see how terms were used and with what frequency.

> A leading corpus linguist, Douglas Biber, has identified four unifying characteristics of the corpus approach:
>
> 1. it is empirical, analyzing the actual pattern of use in natural text;
>
> 2. it utilizes a large and principled collection of natural texts, known as a "corpus," as the basis for analysis;
>
> 3. it makes extensive use of computers for analysis, using both automatic and interactive techniques; and
>
> 4. it depends on both quantitative and qualitative analytical techniques.

Stephen C. Mouritsen, *The Dictionary is Not a Fortress: Definitional Fallacies and a Corpus-Based Approach to Plain Meaning*, 2010 BYU L. REV. 1915, 1954 (2010) (citing DOUGLAS BIBER ET AL., CORPUS LINGUISTICS: INVESTIGATING LANGUAGE STRUCTURE AND USE 4 (1998)); *see also* Clark D. Cunningham, *Foreword: Lawyers and Linguists Collaborate in Using Corpus Linguistics to Produce New Insights Into Original Meaning*, 36 GA. ST. U. L. REV. vi (2020) (describing corpus linguistics and summarizing articles in symposium issue); Thomas R. Lee & Stephen C. Mouritsen, *Judging Ordinary Meaning*, 127 YALE L.J. 788 (2018) (summarizing corpus linguistics and applying its methodology to cases); Lawrence Solan & Tammy Gales, *Corpus Linguistics as a Tool in Legal Interpretation*, 2017 BYU L. REV. 1311 (2017) (summarizing and critiquing corpus linguistics). Several databases have been assembled containing millions or even billions of entries from newspaper, books, periodicals, broadcast transcripts and the like. *See generally, e.g.*, *Corpus of Historical American English*, ENGLISH-CORPORA, https://www.english-corpora.org/coha/ [https://perma.cc/2CAU-U9N5]; *Corpus of Contemporary American English*, ENGLISH-CORPORA, https://www.english-corpora.org/coca/ [https://perma.cc/99B2-4X4Z].

could then present an unduly one-dimensional approach[5] to understanding and applying the language of legal documents[6] such as constitutions, statutes, rules, regulations and contracts.[7] To a large extent, my criticisms of and suggested use for corpus linguistics reflect the larger, longer-running battle between advocates of relatively restrictive textualism and proponents of broader interpretative methodologies that consider a range of data bearing on the meaning of a document.

Sound interpretation and construction requires that context supplement text. The two factors are complimentary rather than mutually exclusive. Although corpus linguistics may improve textual analysis, it cannot eliminate the need for contextual consideration. The additional data about word meaning provided by corpus linguistics analysis should be

---

[5] *See* Bernstein, *Half-Empirical Attitude*, *supra* note 3, at 15 ("[O]ne prominent approach that vociferously rejects using pragmatics to evaluate meaning: textualism."); *see also id.* at 15–17 (extending criticism of textualism fused with legal corpus linguistics, noting that textualists vacillate between concern for lay understanding of term and specialist understanding of terms); Lee & Mouritsen, *supra* note 4, at 865–79 (recognizing and responding to criticisms of corpus linguistics approach); Solan & Gales, *supra* note 4, at 1320–31 (sympathetic critique but criticism nonetheless).

[6] Because this is a law review article, I limit my analysis to legal documents, which includes not only statutes—which along with the U.S. Constitution have received the bulk of attention from corpus linguists and interpretative scholars but also regulations (e.g., of a government agency or private organization), rules (e.g., of procedure or governing use of a facility), and contract writings. The meaning of the text of any document—and the document itself— can become disputed which in turn requires interpretation of the words of the document and construction of the meaning and application of the document, activities that in my view are improved by consideration of context as well as text. On the distinction between "interpretation" and "construction," see Lawrence Solum, *Legal Theory Lexicon: Interpretation and Construction*, LEGAL THEORY BLOG (Feb. 8, 2009), https://lsolum.typepad.com/legaltheory/2009/02/legal-theory-lexicon-interpretation-and-construction.html [https://perma.cc/HU92-2LV7] Interpretation is the "activity of determining the linguistic meaning (or semantic content) of a legal text" while construction is the "activity of translating the semantic content of the text into legal rules"); *see also* Lawrence B. Solum, *The Interpretation-Construction Distinction*, 27 CONST. COMM. 95 (2010). I would add "or decisions" to Professor Solum's definition of construction.

[7] In this relatively short commentary, I will not reiterate at length the criticisms of textualism, particularly the "new" textualism that eschews consideration of extrinsic and contextual information even more than does the traditional "plain meaning" approach to interpretation but instead merely note that I am largely in agreement with the criticisms. *See* WILLIAM N. ESKRIDGE, JR., ET AL. CASES AND MATERIALS ON LEGISLATION AND REGULATION: STATUTES AND THE CREATION OF PUBLIC POLICY 496–525 (6th ed. 2020) (describing and distinguishing Plain Meaning approach and New Textualism, with excerpted cases as illustrations).

The words used in legal documents are of course important in construing the documents, but the interpretative process should not be restricted solely to a two-dimensional focus on words divorced from other important considerations. As I hope this commentary also reflects, I am not opposed to corpus linguistics and related textual assessments through use of big data and empirical analysis. I am not even opposed to layperson surveys regarding word meaning. I am, however, quite strongly opposed to making any of these exclusive or controlling factors in determining documentary meaning and dispute resolution.

embraced and given careful consideration by interpreters—as should contextual data. I leave it to the imperialistic agenda entrepreneurs of the respective camps to argue about which tools should be used in first resort or given primacy if they suggest different meanings. My core point is that in a typical case, *all* of these factors should be considered unless there are logistical or veracity factors that point to the contrary.

Documents, be they statutes, regulations, contracts, all emerge in the context of addressing a problem or objective (legal, economic, social, political, commercial). However skilled the drafters of these documents in choosing language, consideration of contextual factors is likely to increase an interpreter's understanding of the document and provide a scale for measuring the accuracy of textual interpretation, even if that textual interpretation is enriched by corpus linguistics or other sophisticated analysis that extends beyond consulting dictionaries or the armchair reaction of a single judge.[8] In cases where document text is not carefully drafted or is being applied to situations not envisioned by the drafters, the case for considering context becomes clearer.

In addition, even if one is a strict textualist, corpus linguistic analysis may be an unproductive use of the interpreter's time. In some cases, the meaning—at least facially or superficially—of a word or provision is so indisputably clear that examination of a database of current or historical use is unnecessary.[9] In these types

---

[8] *See* STEPHEN BREYER, ACTIVE LIBERTY: INTERPRETING OUR DEMOCRATIC CONSTITUTION 85 (2006) ("overemphasis on text can lead courts astray" while consideration of drafter intent and purpose along with text more likely to lead to correct interpretation); HENRY M. HART JR. & ALBERT M. SACKS, THE LEGAL PROCESS: BASIC PROBLEMS IN THE MAKING AND APPLICATION OF LAW (William N. Eskridge Jr. & Philip P. Frickey eds., 1994) (publication of famous Hart & Sacks Legal Process teaching materials advocating multi-factor approach to interpretation as more effective approach); William N. Eskridge Jr. & Philip P. Frickey, *Statutory Interpretation as Practical Reasoning*, 42 STAN. L. REV. 321 (1990) (determination of statutory meaning improved by consideration of multiple factors and iteration of interpretation though dialogue with other viewers of legislation).

[9] For example, where a legal document (constitution, statute, regulation, rule, or contract) contains a provision involving age or date, there is (absent very unusual circumstances such as confusion about applicable calendars) no doubt as to the content of the texts. Even where the word used has some vagueness or ambiguity, what might be termed "intrinsic context" associated with the term and even the slightest background information may make textual meaning so clear that additional investigation is not required or worth the expenditure of resources.

Consider, for example, the well-known thought experiment about textual ambiguity: a "no vehicles in the park" sign. *See* H.L.A. Hart, *Positivism and the Separation of Law and Morals*, 71 HARV. L. REV. 593, 607 (1958). The meaning of the sign is then tested with hypotheticals illustrating the limits of language literalism such as a garbage truck collecting from the park's waste receptacles, a bike, a trike, rollerblades and of course a war memorial statue containing a replica of a Jeep. But if a park patron rolls through the grounds in his new Tesla to minimize the distance traversed with his picnic basket, anyone

of cases, discussed below, the interpretative task is not discernment of word meaning but a judicial decision as to the manner in which the words in question should be applied.

By "superficially" clear text, I refer to situations where everyone is quite certain as to the ordinary meaning of the word(s) in question but where there is substantial doubt that the word was meant to be literally applied to the situation before the court. So-called scrivener's errors belong in this category as well.[10] For purposes of this article, I am categorizing codifications of positive law (e.g., statutes), private communications (e.g., letters and emails), and agreements (e.g., contracts) as documents that courts may be required to interpret and apply in the course resolving disputes.

In Part I, I discuss the reasons why corpus linguistics should not be considered in isolation from contextual factors, as the latter often illuminate meanings that cannot be found from simply chronicling the usage of a given word. In Part II, I demonstrate, through the lens of three Supreme Court cases, that corpus linguistics does not aid interpretation when the words of a statute or document are clear, but their application to the facts at hand is not. My critique of corpus linguistics mirrors the larger, long-running, and ongoing debate of the merits of a more textual approach to interpretation that focuses almost exclusively on the language of a legal document versus a more contextual approach that construes legal documents not only in light of their language but also in view of the background, motivation, context, and purpose of the document as well as credible evidence of author intent.[11] I proudly admit to a preference for an eclectic approach

---

possessed of common intelligence and community values readily accepts that the sign's prohibition "clearly" applies to the Tesla driver.

[10] By a scrivener's error, I mean a mistake in translation between what was meant by legislators (or regulators or contracting parties) or even what was enacted, and the text recorded in the document. For example, if a statute was intended to have an effective date of 2012 that was recorded as 2021, this transposition of numbers would be a classic scrivener's error. Relatedly, but more subtly, a document may contain language that is not erroneous per se but may misrepresent the intent of the drafters or clash with prevailing legal norms if given literal enforcement.

For example, in the well-known case of United States v. Locke, 471 U.S. 84, 89 (1985), the statute provided that holders of mining leases must apply for renewal "prior to December 31," an odd phrasing that may have been a misunderstanding tantamount to scrivener's error and was at least inartful drafting casting doubt on whether Congress intended to use this unusual deadline that could prove a trap for the unwary. Nonetheless, the Court gave strict enforcement to the deadline, meaning that a lease holder applying for renewal on December 31 had missed the deadline and lost its leasing rights. *See id* at 93.

[11] Regarding intent, the position of many textualists—and most courts—is that the language of a legal document is the best evidence of drafter intent (regarding public documents such as statutes or regulations) or party intent (in the case of private documents such as leases, invoices, purchase agreements, retention letters, or the like).

to interpretation that embraces not only textualism and data-driven analysis such as corpus linguistics, but also considers other interpretative indicia.[12]

Even if one rejects narrow textualism and favors wide-ranging consideration of contextual information, corpus linguistics should be appreciated as a significant advance in textual interpretative methodology. Instead of giving a gut reaction, soul searching, pulling the most convenient dictionary from a nearby shelf, or batting around terms with law clerks, court staff, or judicial colleagues (as well as listening to arguments of counsel), a judge facing an interpretative question can instead (or in addition) examine the way in which a word is used by the public at large.

I. THE LIMITS OF CORPUS LINGUISTICS

A. Quality Control and Corpus Linguistics

Although lay word usage in non-legal settings cannot be a talisman, neither should it be ignored if properly presented by counsel. Judicial use of corpus linguistics data will ordinarily be presented through counsel's own work (or that of its retained experts) with relevant corpora. This will entail the procedural cost of ensuring that this information is appropriately absorbed and assessed by courts. Introduction of corpus linguistics information raises evidentiary issues that currently appear under-assessed. At a minimum, corpus linguistics work should proceed from a foundation of admissibility that requires requisite expertise of the user, sufficient comprehension by courts (and by opponents so that due process is satisfied), transparency, replicability, and justifiability. If corpus linguistics evidence is aptly introduced

See, e.g., United States v. Am. Trucking Ass'n, 310 U.S. 534, 543 (1940)(text of statute generally the best evidence of congressional intent) ("There is, of course, no more persuasive evidence of the purpose of a statute than the words by which the legislature undertook to give expression to its wishes."); United States v. Missouri Pacific R. Co., 278 U.S. 269, 278 (1929) (where language of statute is clear and its application does not produce an absurd result or impracticable consequences, the text of the statute controls). Even if this very optimistic view of the writing abilities of human beings (or, more challenging yet, human beings working as a group) is largely correct, it cannot be correct one-hundred percent of the time. Consequently, consideration of contextual factors is likely to increase interpretative accuracy by supplementing the understanding of the reader so long as readers are not affirmatively misled by context.

[12] Examples of other indicia of meaning beyond documentary text include: drafter and party intent; the document's purpose; public policy; the resulting impact of alternative interpretations; legislative-judicial colloquy; precedent and reaction to precedent; enforcement issues; changed circumstances, or; other factors reasonably bearing on word meaning. *See generally* WILLIAM N. ESKRIDGE, JR., INTERPRETING LAW: A PRIMER ON HOW TO READ STATUTES AND THE CONSTITUTION (2016) (setting forth wide array of these and factors bearing on statutory construction).

through a sufficiently skilled expert witness, the evidentiary foundation would appear properly established.

But if this database research is done by judges or staff sua sponte without review by the parties, there is considerable danger that it could be done poorly with no ready mechanism for catching or correcting errors.[13] Although one might excuse this on the ground that corpus linguistics work is in the nature of legislative facts rather than adjudicative facts, I find this view misplaced as well as insufficiently concerned about the dangers posed when courts make conclusions about the workings of the world (legislative facts) without benefit of having a record or scrutiny by advocates. One might regard hidden use of corpus linguistics as the judicial equivalent of a juror improperly visiting a crime or accident scene and conducting her own measurements and physical assessments. This situation may be an inevitable part of judging and may be no worse than judges and staff randomly browsing dictionaries—but its propriety should at least be considered according to the standards we use for assessing evidence, including the receipt of other data in court.

To be properly applied, corpus linguistics data should ordinarily require more vetting than what often accompanies judicial reference to scientific literature or current events, which are often treated as "legislative facts" that courts may decree without much constraint. If corpus linguistics is used as part of a textualist method that rejects consideration of evidence of background, history, context, drafter intent, or purpose of the document in question, it could become a catalyst for error rather than understanding. At a minimum, a strictly textualist court utilizing corpus linguistics analysis should insist on correct and rigorous methodology and ensure that it is consistently applied.

B. A Big Dictionary Is Still Just a Dictionary

Dictionary use by courts has received significant scrutiny, most of it critical.[14] As is the case with corpus linguistics, I self-identify as one who is not opposed to consideration of dictionaries as

---

[13] *See* Daniel C. Tankersley, Comment, *Beyond the Dictionary: Why Sua Sponte Judicial Use of Corpus Linguistics is Not Appropriate for Statutory Interpretation*, 87 MISS. L.J. 641, 668 (2018).

[14] *See, e.g.*, James J. 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, 483 (2013); Samuel A. Thumma & Jeffrey Kirchmeier, *The Lexicon Has Become a Fortress: The United States Supreme Court's Use of Dictionaries*, 47 BUFF. L. REV. 227, 232–33 (1999); Lawrence Solan, *When Judges Use the Dictionary*, 68 AM. SPEECH 50, 50 (1993); Ellen P. Aprill, *The Law of the Word: Dictionary Shopping in the Supreme Court*, 30 ARIZ. ST. L.J. 275, 277 (1998).

part of the interpretative process, but rather as someone who also wants courts to consider additional means of analysis as standard judicial operating procedure—regardless of whether text read in isolation appears clear and unambiguous upon initial reading.[15] Let the dictionaries be pulled from the shelves, at least until the definitions get repetitive or yield only diminishing research returns. But let the dictionaries also be supplemented by other means of analysis, both textual and non-textual and contextual.[16]

To the extent dictionaries are useful interpretative tools, a really big, searchable dictionary would seem to be an even better interpretative tool and an improvement in the state of the art of word construction. Corpus linguistics, if done even modestly well, meets this standard. It certainly helps reduce the problem of selecting the apt dictionary to use when construing a legal document. Instead of favoring Dictionary A (and citing it in an opinion) over Dictionary B, the court can consider a wider analysis of word usage. Although corpora are not literally dictionaries, they have an important functional similarity in that they contain not merely a few illustrations of word usage but potentially thousands of examples.[17]

Calling this a state-of-the-art improvement may be a mild overstatement. While the advantages of a large database of language use (and contextual use at that)[18] are clear, a

---

[15] Even if wrong, I am at least consistent across the decades. *See, e.g.*, Jeffrey W. Stempel, *Unmet Expectations: Undue Restriction of the Reasonable Expectations Approach and the Misleading Mythology of Judicial Role*, 5 CONN. INS. L.J. 181, 279, 291–93 (1998) (arguing that "objectively reasonable expectations" of both policyholder and insurer regarding operation of insurance policies be consistently considered at the onset of policy interpretation without a prerequisite finding of facial ambiguity of policy language); Jeffrey W. Stempel, *The Rehnquist Court, Statutory Interpretation, Inertial Burdens, and a Misleading Version of Democracy*, 22 U. TOL. L. REV. 583, 664, 671–72 (1991) (arguing that when addressing difficult interpretative questions, courts should consider the relative difficulty respective stakeholders would have in obtaining corrective legislation).

[16] *See* James M. Fischer, *Why Are Insurance Contracts Subject to Special Rules of Interpretation? Text Versus Context*, 24 ARIZ. ST. L.J. 995, 1005–06 (1992) (supporting contextual analysis of disputed insurance policy text). One minor quibble I have with Professor Fischer's excellent article is its seeming premise that contextual factors are less important—or at least less used by the judiciary—in non-insurance contract disputes. *Id.* at 1030–32. At the risk of sounding tautological, there may be differences in judicial receptiveness to contextual interpretative factors based on the context of the dispute itself (e.g., consumer contract, business-to-business contract, lengthy, complex contract documents vs. short, simple, plain language memorialization of an agreement, sophisticated investor vs. poorly educated pensioner, presence of duress or undue influence, etc.)—and these considerations may impact judicial approaches to document interpretation far more than the subject matter of the contract (e.g., insurance vs. real estate vs. clothing vs. software vs. investment funds); *see also* YONG QIANG HAN, POLICYHOLDER'S REASONABLE EXPECTATIONS 1–6 (2016) (finding considerable use of reasonable expectations concept in commercial contracts as well as regarding insurance policies).

[17] *See* Mouritsen, *supra* note 4, at 1954–55.

[18] Although I am concerned that corpus linguistics not become a monster "HAL 9000 meets Textual Literalism" conception, it should not be overlooked that a premise of

computerized collection of word use may in some cases be less informed then the context provided by dictionary authors through the use of exemplary sentences and illustrations commonly found with dictionary entries.[19] It certainly can be less informed than consideration by courts (through precedent) and policymakers (through legislative history, background, and purpose, as well as executive reaction and implementation). Some degree of judicial discretion is still required in the form of search guidance and dictionary selection and reliance.

In spite of the promise of corpus linguistics, there remain other concerns. What if the database is not sufficiently broad? Or if the search strategy yielding the empirical data is flawed? More important, does the use of a large corpus simply substitute a computer database for a shelf of dictionaries without addressing the problems of excessive focus on documentary text?[20]

Even if corpus linguistics is nothing more than use of a bigger dictionary with more examples of word usage, it still provides a valuable contribution, particularly if combined with other means of interpretative analysis. But if users of corpus

---

corpus linguistics analysis is looking at language in the context of use and not in the strict isolation one finds with some dictionary definitions. *See* 2001: A SPACE ODYSSEY (Metro-Goldwyn-Mayer 1968) (HAL 9000 supercomputer of high but literal artificial intelligence refuses to follow instructions of human crew, killing one crewmember prior to disabling by surviving crewmember). Having ventured this far down the contextual road, corpus linguistics analysis should be willing to embrace contextual factors other than the surrounding text of language.

[19] The authors of a dictionary do more than merely assemble examples of word use at random or based on frequency of use. They exercise editorial judgment in selecting particularly helpful and clarifying examples. Even if reasonable editors might have some disagreement over those selections, there nonetheless is substantial editorial and analytic value added by the editors' expertise that goes beyond collection or tabulation.

[20] *See* Bernstein, *Half-Empirical Attitude*, *supra* note 3, at 14–15 ("Most legal corpus inquiries have used a few publicly available corpora. Particularly popular have been several large corpora compiled by the legal corpus linguistics project at Brigham Young University, whose law school has been a leading force in promoting the method. These include the Corpus of Contemporary American English (COCA) and the News on the Web Corpus (NOW), as well as historical corpora like the Corpus of Historical American English (COHA), and the Corpus of Founding Era American English (COFEA). There are many other corpora out there—linguists have collected all sorts of texts and recordings to study—but the Brigham Young ones are the legal corpus analysis favorites . . . . These corpora share an emphasis on size: each boasts of the sheer number of words, often numbering in the billions, that they collect. Yet they are sometimes a bit cavalier in their claims about what those billions of words can reasonably be seen to offer. The COFEA, for instance, tells us that it contains 'documents from ordinary people of the day,' but does not give the kinds of demographic information that would be crucial to evaluating its range of representation of language in an era of low literacy, expensive writing materials, and extreme opportunity disparity. Propertied White men and enslaved Black women were both ordinary people of the day subject to founding era laws. But given their different access to text production and preservation, the former are likely to be over-represented, the latter under-represented in a contemporaneous corpus. This does not make the corpus useless; but it does mean that 'ordinary people of the day' fails to explain just what it is the corpus offers.").

linguistics analysis are myopically focused (if one can be myopic in looking at a database with millions of words) on documentary text alone, or to the exclusion of other valuable indicia of word meaning, the benefits of corpus linguistics are reduced.[21] To the extent corpus linguistics analysis becomes a closed universe ousting consideration of other interpretative evidence, this promising school of analysis could become an academy of harm.

C. *Context Is Also "Data": Perhaps Better Data Than Computerized Examination of Word Usage*

Much as corpus linguistics represents a refined improvement over traditional strict textual analysis, it does this at some cost. By appearing to be—and in all likelihood actually being—more reliable than an individual judge's internal understanding of language or a dictionary definition, corpus linguistics discourages judges from considering other indicia of contract meaning. While this is of course true of textualism itself, corpus linguistics arguably exacerbates the problem by making a more convincing argument that one can have enough data about word usage to decide cases without reference to other indicia of meaning.

The notion is dangerous when dealing with corpus linguistics just as it is dangerous when expressed by traditional textualists who have long argued that if a document has "plain meaning" and is not facially ambiguous, there is no need to assess meaning by anything deeper than the parchment of the document.[22] To be sure, one's initial reading of a document is a

---

[21] Others have at least implied this criticism or limitation of corpus linguistics. *See, e.g.*, Evan C. Zoldan, *Corpus Linguistics and the Dream of Objectivity*, 50 SETON HALL L. REV. 401 (2019) (arguing that corpus linguistics analysis alone is insufficient for accurate interpretative results); Lawrence B. Solum, *Triangulating Public Meaning: Corpus Linguistics, Immersion, and the Constitutional Record*, 2017 B.Y.U. L. REV. 1621 (2017) (arguing that examination of constitutional history and background should be deployed along with corpus linguistics to reach sounder interpretative results). Prominent proponents of corpus linguistics analysis have defended it against assertions that it is myopic or one dimensional. *See, e.g.*, Thomas R. Lee & Stephen C. Mouritsen, *The Corpus and the Critics*, 88 U. CHI. L. REV. 275 (2021) (arguing that although corpus linguistics focuses on textual meaning, the process inherently involves context as reflected in use of terms contained in wide range of writings and accepting that corpus linguistics analysis is not sole determinant of word meaning).

[22] *See, e.g.*, ANTONIN SCALIA & BRYAN GARNER, READING LAW 53–55 (2012) (viewing interpretation as discerning meaning of document from its text according to prevailing language conventions without reference to extrinsic evidence); ANTONIN SCALIA, A MATTER OF INTERPRETATION 23–29 (1997) (textualist philosophy applies plain meaning of words in statute without regard to legislative history or other extrinsic evidence, which late Justice regarded as illegitimate inquiry in that only the text of the statute has legal force); Frank Easterbrook, *Statutes' Domains*, 50 U. CHI. L. REV. 533, 539–44 (1983) (if the text of a statute is not sufficiently broad to encompass case at issue, courts should not use legislative history or congressional purpose to expand reach of statutory text).

data point: what one thinks upon first reading is of course relevant to meaning. So are second (or third or . . .) readings and close scrutiny of punctuation and absorption of lists that may have been glossed over during the initial reading. So are dictionary definitions of the words used in a text. So are examples of the word's usage in a variety of settings.

All of this is data—just as examinations of word meaning in corpora produces data that frequently sheds light on word meaning. But so is the contextual information surrounding a document and its terms: the situation prompting the document; the purpose of the document in light of the situation; interaction of participants in the drafting; drafting records; contemporaneous perception of the affect of the document, and so on. Not only is this data, it is often powerful data that often says a great deal about the meaning of a document.

But for reasons that have never been persuasive to me, textualists tend to underweight, or in extreme cases, even ignore all of this data unless there is facial ambiguity of the text.[23] The textualist preference is to make interpretive conclusions solely on the reaction of a single reader (the judge) or small group of readers (an appellate panel, judicial staff, perhaps arguments of counsel if the judge remains sufficiently open-minded).[24] Contextualists by contrast take the view that a rational interpreter should consistently be presumptively open to consideration of all information relevant to the meaning of a document.[25]

Thus, while corpus linguistics is an advancement over the traditional use of dictionary definitions alone, it could also

---

[23] Some of this emerges from the textualist belief that where statutory text is clear, it is illegitimate for judges to consider anything in addition to the law's text. *See* sources cited *supra* note 22. These commentators also argue that one cannot reliably discern the specific intent or general purpose of an enacting legislature. *See, e.g.*, SCALIA, *supra* note 22, at 29–37. Others may accept the legitimacy of these sources but give them low saliency if they run counter to the preferred construction of statutory text. *See, e.g.*, SCALIA & GARNER, *supra* note 22, at 56 ("purpose must be derived from the text, not from extrinsic sources such as legislative history or an assumption about the legal drafter's desires").

[24] In phrasing my description/criticism this way, I am arguing that traditional textualists take a very narrow view of the apt audience or interpretative community for determining textual meaning: essentially the judge (or other judges on a panel or en banc court) and their law clerks, or others they might informally consult (the prospect of which in my view would raise ex parte contact and consideration of extra-record evidence issues that are generally ignored). By contrast, corpus linguistics textualist use an expanded universe for determining word meaning is in my view an improvement on insularity and random dictionary consultation despite the shortcomings of corpora as lay usage of words outside legal context.

[25] *See, e.g.*, Abbe R. Gluck & Lisa Schultz Bressman, *Statutory Interpretation form the Inside – An Empirical Study of Congressional Drafting, Delegation, and the Canons: Part I*, 65 STAN. L. REV. 901 (2013) (examining context from which statutes emerge and finding this important to apt understanding of statutory text); Cathy Mijman, *Ascertaining the Meaning of Legislation – A Question of Context*, 43 VICTORIA U. WELL. L. REV. 629 (200) (finding consideration of context essential to sound interpretation); Eskridge & Frickey, *supra* note 8, at 324 (advocating multi-factor approach to statutory interpretation).

produce the unwise side effect of discouraging consideration of context. To the extent that corpus linguists, like textualists, resist (or even bar) consideration of other data about word and document meaning, the method becomes a tainted tool.

Put another way, which is more likely to yield insight into the meaning of a document: collecting examples of word use in corpora or studying the background and development of the document itself? Although corpus linguists will probably disagree and strict textualists will undoubtedly disagree, it seems compellingly logical to think that contextual data will in most cases be more informative about documentary meaning than data about how the terms in the document are used in the popular press or popular culture.[26]

The additional data about word meaning provided by corpus linguistics analysis should be embraced and given careful consideration by interpreters. But so should contextual data. Here, "data" mean the "little" data such as situation-specific information about the impetus and objective of legislation (and regulations, rules, contracts, or other written documents) as well as the "big" data of statistics and corpora. My core point is that in a typical case *all* of these factors should be considered unless there are logistical or veracity factors that point to the contrary. For example, if proffered information is unreliable (e.g., insufficient proof of purported drafting intent) or will consume inordinate judicial resources to obtain or process, this may justify judicial unwillingness to consider or search out the information. But if not, context should be given a place at the metaphorical interpretative table along with text.

1. Logistical and Cost-Benefit Concerns Surrounding Corpus Linguistics and Contextual Evidence

One implicit myth of textualism is that it consumes fewer judicial resources.[27] The judge need only read the text and, if she

---

[26] *See* Eskridge & Frickey, *supra* note 8, at 383 (arguing for multi-factor analysis as a means to better statutory interpretation.).

[27] *See* Robert J. Pushaw, Jr., *Talking Textualism, Practicing Pragmatism: Rethinking The Supreme Court's Approach to Statutory Interpretation*, 51 GA. L. REV. 121,131 (2016) (suggesting that textualism is more efficient than pragmatism or multi-factor approaches to statutory interpretation because jurists are expert at parsing text and "far more efficient" than multi-factor methods because "it focuses on the statute alone"); *See generally* Robert M. Lawless, *Legisprudence Through a Bankruptcy Lens: A Study in the Supreme Court's Bankruptcy Cases*, 47 SYRACUSE L. REV. 1 (1996) (recognizing and refuting myth in the context of bankruptcy statutes); Nicholas S. Zeppos, *Justice Scalia's Textualism: The "New" New Legal Process*, 12 CARDOZO L. REV. 1597 (1991) (textualism is efficient because it avoids the need to process large volumes of information and utilizes clear rules of word meaning).

finds the text clear, render a decision. While this template for almost instant justice may hold true in some cases (e.g., a rental lease that promises payment in U.S. dollars on or before the first of the month), there will be many cases where analysis "limited" to text (and the information textualists consider acceptable for consideration) does not necessarily lead to swifter, less expensive final disposition than would broader consideration at the outset of all information bearing on document meaning.

Once outside the realm of documents with indisputably clear text (e.g., the U.S. Constitution's requirement that the President be at least thirty-five years old[28]), courts are usually at least consulting dictionaries, treatises and other secondary sources, and of course are consulting precedent as well as canons of construction, which involves the burden of taking time to determine the applicable canons.[29] This may not be as time consuming as a searching review of the origins of the Sherman Act,[30] but neither is it something one does during a television commercial break. Except in very clear cases, even a self-consciously textualist court will need to devote substantial resources to determining word meaning, which raises the question of whether time is better spent on increased examination of text via reading more cases, consulting more dictionaries or conducting a corpus linguistics analysis rather than spending time considering extrinsic evidence of meaning such as context, background, history, drafter intent, document purpose, and function of the text within the area of law in question.

Adding corpus linguistics to this textual inquiry—although promising—comes at the cost of investing additional judicial time in deciphering text. Even a quick breeze through the Corpus of Contemporary American English (COCA),[31] the Corpus of Historical American English (COHA),[32] or other corpora

---

[28] U.S. CONST. art. II, § 1, cl. 5.

[29] I find it interesting that a "Bible" of textualists contains more than 50 canons of construction. *See* ANTONIN SCALIA & BRYAN GARNER, READING LAW: THE INTERPRETATION OF LEGAL TEXTS 69 (2012). For the textualist interested in using any of these canons (but not extrinsic evidence), there remains the task of selecting among the many canons in the book, some of which are in tension if not outright conflict. My favorite canonical clash, noted in Karl Llwellyn's famous article more than 70 years ago, is the clash between the notion that statutes in derogation of the common law should be strictly construed while remedial statutes should be liberally construed. *See* Karl N. Llwellyn, *Remarks on the Theory of Appellate Decision and the Rules or Canons About How Statutes Are to be Construed*, 3 VAND. L. REV. 395, 401 (1950)). By definition, remedial statutes change or "derogate" the common law. There is no way in which both canons can be faithfully applied. Even if it is in denial, a court must choose between the two canons or principles and decide whether it will give more support to restraining legislation or applying it broadly.

[30] 15 U.S.C. §§ 1–38.

[31] *See Corpus of Contemporary American English*, *supra* note 4.

[32] *See id.*

requires some time and thought, both in constructing the inquiry and in assessing the results. A more serious "dive" into corpora databases requires even more time. If this is done by the court outside the view of the parties, this not only is what might be termed "dead weight" judicial time,[33] but also raises ethical concerns in that the parties are not provided the opportunity to check the court's methodology and cross-examine the results. The risk that a court will err in determining facts or assessing information in a vacuum have been recognized in cases involving internet research by judges. The judge may obtain suspect information that is never adequately vetted, or the judge may misinterpret the research without correction from a law clerk, counsel or the parties. Although interpretation of documentary text is a matter of law, database research regarding resolution of a disputed word on which a case may hinge is arguably akin to sua sponte judicial investigation of adjudicative facts, both of which may take place with some frequency.[34]

---

[33] By this I mean, as discussed regarding the cost-benefit issues regarding extrinsic evidence, that sua sponte text research by the court from metaphorical square one may often engulf the court in time-consuming search costs, just as sua sponte judicial examination of legislative history or background imposes costs on the court. But if the court is instead merely evaluating the information brought to it by counsel and the parties (be the information textual or extrinsic to text), fewer judicial resources are expended. The court must of course take the time to evaluate the competing submissions, which may include significant time reading material brought to its attention. (One would not want a supposedly neutral court to automatically accept one side's characterization of background facts while reflexively rejecting the submissions of an opponent). But this will usually require substantially less judicial time than if the court was required to first locate and unearth contextual evidence of meaning as a precursor to weighing and assessing it.

[34] *See* George L. Blum, Annotation, *Allowable Judicial Internet Research*, 44 A.L.R. 7th Art. 8, § 19 (2019); M. Cristina Martin, *"Googling" Your Way to Justice: How Judge Posner Was (Almost) Correct in His Use of Internet Research in* Rowe v. Gibson, 11 SEVENTH CIR. REV. 1, 24, 29 (2015); Elizabeth G. Thornburg, *The Curious Appellate Judge: Ethical Limits on Independent Research*, 28 REV. LITIG. 131, 132–42 (2008); David H. Tennant & Laurie M. Seal, *Judicial Ethics and the Internet: May Judges Search the Internet in Evaluating and Deciding a Case?*, 16 PROF. LAW. 2, 2 (2005); *see also* MODEL CODE of JUDICIAL CONDUCT r. 2.9(c) (AM. BAR ASS'N 2020) (judge "shall not investigate facts in a matter independently and shall consider only the evidence presented and any facts that may properly be judicial noticed" pursuant to Fed. R. Evid. 201 and state counterparts); ABA Comm. On Ethics & Prof'l Responsibility, Formal Op. 478, at 7 (2017) (limiting boundaries of proper judicial research to general background information to improve judge's understanding of subject matter of case but barring research regarding particular facts of dispute); Coleen M. Barger, *On the Internet, Nobody Knows You're a Judge: Appellate Courts' Use of Internet Materials*, 4 J. APP. PRAC. & PROCESS 417, 418–49 (2002) (cataloging citations to internet materials in judicial opinions).

In chambers, review of real evidence or judicial visits to physical locations pose similar risks. In addition, even if the raw information held by the judge is essentially accurate, a judge insufficiently versed in statistics, science, or other relevant fields may be a poor consumer of the information. *See, e.g.*, Michael Saks, *Ignorance of Science is No Excuse*, 10 TRIAL 18 (1974) (criticizing court's jury size cases for failing to realize that the same sociological studies arguing against five-person juries (found unconstitutional by the Supreme Court) made an equally compelling case

Also overlooked in debate about the relative costs and benefits of consideration of non-textual indications of meaning is the role of the adversary system. Where American adversarial litigation is functioning as it should, the parties and counsel are incentivized to conduct research and investigation that will support their positions and present this information to the court. The court is not required to engage in historical archival research rivaling that of a Ph.D. candidate, but rather need only be willing to consider the information brought to it by counsel.

The logistical burden—at least on judges—of considering context is consistently reduced by the adversary system. American judges, unlike their counterparts in some systems, are not charged with investigating claims and developing facts. Although it has attracted criticism, the "umpire" metaphor used so successfully by Justice John Roberts at his confirmation hearings[35] does accurately describe a key aspect of adjudication: the court does not have to do the work of finding and presenting evidence—that's the job of the parties.[36] If the parties bring forth

---

against six-person juries). McKune v. Lile, 536 U.S. 24, 32 (2002) (Court adopts as settled fact that recidivism of sexual offenders drops from 80 percent to 15 percent with treatment based on misreading of problematic article). Robert E. Freeman-Longo & Ronald V. Wall, *Changing a Lifetime of Sexual Crime*, PSYCHOL. TODAY, Mar. 1968, at 58; *see also* EXEC. OFFICE OF THE PRESIDENT: PRESIDENT'S COUNCIL OF ADVISORS ON SCI & TECH., REPORT TO THE PRESIDENT FORENSIC SCIENCE IN CRIMINAL COURTS: ENSURING SCIENTIFIC VALIDITY OF FEATURE-COMPARISON METHODS 1–20 (2016) (noting certain types of "evidence" widely admitted in court lack scientific rigor, using bitemarks, firearms, and latent fingerprints as examples).

The issue is not easily resolved in balancing adversarial fairness and judicial discretion. Some judicial initiative in investigating (through either legal or factual research) issues underdeveloped by the parties can improve judicial decisionmaking. Former Seventh Circuit Judge Richard Posner's opinions have been refreshing illustrations and caution against an absolute bar to sua sponte judicial research. *See, e.g.*, Harbor Ins. Co. v. Cont'l Bank Corp., 922 F.2d 357, 362 (7th Cir. 1990) (applying Illinois law) (conducting research regarding origin and meaning of term "mend the hold" after being disappointed in counsel's failure to provide information).

[35] The umpire analogy has been criticized as simplistic and misleading. *See* Michael P. Allen, *A Limited Defense of (at Least Some of) the Umpire Analogy*, 32 SEATTLE U. L. REV. 525, 526 n.4 (2009) (noting the criticism conveyed against the analogy); Richard A. Posner, *The Role of the Judge in the Twenty-First Century*, 86 B.U. L. REV. 1049, 1051 (2006) ("The formalist conception of judging crudely depicted by Roberts is fancied up in versions intended for academic audiences. No serious person thinks that the rules that judges in our system apply, particularly appellate judges and most particularly the Justices of the U.S. Supreme Court, are given to them the way the rules of baseball are given to umpires."). But it remains popular because it appears to reflect the judges' perception of their role. *See, e.g.*, BOB WOODWARD & SCOTT ARMSTRONG, THE BRETHREN 430 (1st Paperback ed. 2005) (in response to question as to whether he preferred being a lawyer to a jurist, Justice Lewis Powell rhetorically asked NFL star running back Larry Brown whether he would "rather be a player or a referee.").

[36] *See generally* BAILEY KUKLIN & JEFFREY W. STEMPEL, FOUNDATIONS OF THE LAW: AN INTERDISCIPLINARY AND JUDICIAL PRIMER Ch. 5 (1994) (chapter on Law, Dispute Resolution, and the Adversary System) (describing role of adversary system and party-controlled, counsel-led presentation of evidence and argument); STEPHAN A. LANDSMAN, THE ADVERSARY SYSTEM: A DESCRIPTION AND DEFENSE (1984) (same, with

no extra-textual evidence that is, in the slang parlance of youth, "on them," and they implicitly agree to live with the possible consequence that a court will impose an adverse reading of text that could have been avoided had the losing party bothered to bring the contextual evidence to the court's attention.

Opponents of contextual evidence regarding documentary meaning implicitly raise the logistical concern that consideration of contextual evidence may require too great an expenditure of judicial resources. Although this is not the primary pillar of textualist arguments against consideration of contextual information, it is a significant, if often unspoken brief for concentrating on documentary text alone.[37]

The textualist implicitly posits that focusing on text saves judicial time and energy by eliminating or reducing the time required to unearth and assess contextual evidence of word meaning.[38] To the extent that corpus linguistics is seen as a branch of textualism or an advancement upon traditional textualism, it implicitly rides the same "wave" of conventional wisdom that deciding document disputes based on text alone is more efficient than decision-making that includes contextual information and that this efficiency can be achieved at little or no cost to accuracy.

This supposition may have a kernel of truth, but the time-saving and net efficiency (which, if properly calculated, should include the cost of erroneous decisions) supposedly derived from restricting consideration of context appears exaggerated due to the incentives of litigants to provide the judge with strong evidence on their position, which necessarily reduces judicial need for outside research. Nonetheless, the prevailing mainstream approach to statutory construction and contract interpretation implicitly accepts this supposition. If the words of the document are clear on their face, the court decides the case on this view of the text with no consideration of additional material. Only if the court finds that the text is facially ambiguous does it expressly consider contextual information.[39] Although there is significant jurisprudence to the

---

additional express praise for adversary system in unearthing evidence and sharpening presentation of argument regarding dispute resolution).

[37] *See* sources cited *supra* note 27.

[38] *See id.*

[39] *See* RESTATEMENT OF THE LAW, LIABILITY INSURANCE § 3 (AM. LAW INST. 2019) (adopting plain meaning rule for construction of insurance policies but defining broad category of information as acceptable for ascertaining plain meaning without becoming consideration of "extrinsic" evidence); E. ALLAN FARNSWORTH, CONTRACTS § 7.10 (4th ed. 2004); JOSEPH M. PERILLO, CONTRACTS § 3.10 (7th ed. 2014) (discussing "plain meaning rule and ambiguity" and noting that majority of course require facial

contrary, particularly regarding contracts, the contrarian view appears to be the in the minority.[40]

Although much of this view is premised on the belief that the text is "the" law or "the" contract and must rule unless unclear, it is also a popular view for the often under-emphasized and even unstated reason that it seems more efficient in that it lightens the judicial task by reducing the amount of material that is considered by the court.[41]

Regarding context and efficiency, a little perspective is in order. Although consideration of context is not costless, neither is it automatically more costly than restricting analysis to text alone. Even strict textualism requires significant investment of judicial time in examining materials in addition to the text in dispute: the case record; the implicit case background; party submissions; precedent; treatises and other secondary literature, and of course; dictionaries. The idealized (from an efficiency standpoint) notion of a court merely reading disputed text and deciding the question is a mirage. Except in very "easy" cases, adequately performed textual analysis will include at least a reasonable amount of time consulting these sources of information. There is no guaranty that this will involve less time than it would take to consult legislative history or background information relevant to the origin and understanding of the statute.

If corpus linguistics is added to the mix of textualist tasks, this adds another bucket of information and the attendant task of analyzing the information. It may be a worthwhile expenditure of time, but it remains an expenditure. The judge may avoid the visible expense of holding hearings or a trial regarding extra-textual material and remains in chambers focusing on documents or, in the case of corpus linguistics, a database.

In that sense, the textualist enterprise—or at least the judge-in-chambers-with-a-green-eyeshade version of it—has kinship with the modern judicial preference for deciding cases pursuant to Rule 12(b)(6) motions (which can hinge on whether the

---

textual ambiguity as prerequisite to consideration of extrinsic evidence). As reflected in the discussion of *Cont'l Can Co.*, *infra* notes 49–66 and accompanying text, textualists may also resist contextual information out of concern that the contextual information may be misleading or is in any event not "the law" or "the" contract.

[40] *See, e.g.*, Pac. Gas & Elec. Co. v. G.W. Thomas Drayage & Rigging Co., 442 P.2d 641, 645 (Cal. 1968); Kenneth S. Abraham, *A Theory of Insurance Policy Interpretation*, 95 MICH. L. REV. 531, 539–40 (1996) (treating the *Thomas Drayage* court's openness to extrinsic evidence as the dominant approach).

[41] It may also be a popular judicial view for the more disturbing legal realist reason that it places fewer constraints on judges.

judge finds allegations sufficiently "plausible")[42] or summary judgment motions (which hinge on whether the judge finds that a purported dispute of fact is sufficiently "genuine" and "material").[43] What may seem like a simple exercise of reading and deciding can become something considerably more time-consuming.

In addition, if a reviewing court reads the contested language differently, text-reading that is not verified by consistent contextual factors may result in reversal and remand, which restarts the adjudication cycle in a time-consuming fashion.[44] By contrast, the time invested at the outset in contextual inquiry may be little more time-consuming than the traditional textualist inquiry but may save more time in the long run by reducing the number of appeals, reversals, and remands.

To be sure, the task of assessing contextual information, like the task of assessing textual information, is not cost-free. But the adversary system at least removes from the court the burden of discovering contextual indicia of meaning. If contextual information is to be heard by the court, it will be because the parties gathered and presented the information, and briefed its significance.[45]

The relevant cost-benefit question then becomes not whether it takes time to assess contextual information, but whether the additional consideration of contextual information (in addition to the obviously required examination of disputed text) consumes significantly more judicial time at the threshold in a party-driven adversary system and whether that additional time provides sufficient offsetting benefits through better interpretation of documentary text that is less often reversed or otherwise relitigated.[46]

---

[42] *See* FED. R. CIV. P. 12(b)(6); Bell Atl. Corp. v. Twombly, 550 U.S. 544, 556 (2007); Ashcroft v. Iqbal, 556 U.S. 662, 678 (2009); BROOKE D. COLEMAN ET AL., LEARNING CIV. PROC. 284–302 (3d ed. 2018).

[43] *See* FED. R. CIV. P. 56(a); *Iqbal*, 556 U.S. at 674.

[44] *See* Jeffrey W. Stempel, *Taking Cognitive Illiberalism Seriously: Judicial Humility, Aggregate Efficiency, and Acceptable Justice*, 43 LOY. U. CHI. L.J. 627, 631–32 (2012) (explaining that summary judgment grants are frequently reversed on appeal where an appellate panel disagrees with trial court about facts viewed as undisputed by the trial court as well as in cases where reviewing court holds different view of applicable law.).

[45] *See* sources cited *supra* note 36 (discussing adversary system role in U.S. litigation).

[46] By "otherwise relitigated" I mean that a court's construction of disputed documentary text may be challenged in means other than direct appeal. For example, there may be no appellate or certiorari review of an interpretative decision but the decision, if considered erroneous or problematic or lacking in sufficient guidance may continue to be the subject of repeated challenge in other cases. By contrast, an interpretative decision that makes persuasive use of contextual factors as well as sound textual analysis may become sufficiently authoritative that it discourages re-litigation or streamlines later adjudication by making pretrial disposition or settlement more likely.

Regardless of one's views on the cost-benefit equation of contextual information bearing on the meaning of text, there is no denying that use of corpus linguistics analysis adds to the burden of the textualist judge. By going beyond an initial "gut feeling" reading or consultation with dictionaries, the judge considering corpus linguistics evidence invests additional time processing what could be a substantial amount of information and assessing the associations found by the corpus linguistics inquiry.

At a minimum, this calls into question the efficiency of textualist methodology relative to contextualist approaches as well as the relative probative value of the differing approaches. The corpus linguistics judge spends time on corpus linguistics that could have been spent actually investigating the surrounding circumstances of the dispute rather than exercising a tunnel vision focus on text alone. The same time devoted toward understanding an industry, party relations, the goal of the arrangement, the negotiations, and the behaviors of the parties, as well as the dispute itself could well lead to more accurate resolution of disputes, perhaps in less time than is required for constructing and assessing a corpus search.

But as is the case with contextual evidence, presentation of corpus linguistics information takes place (or at least should take place) through the adversary system. Although it may be too extreme to bar judges from sua sponte corpus linguistics investigation in chambers, it is preferable that such information be presented to the court by the parties through use of expert analysis and empirical evidence that is subject to cross-examination by motivated parties armed with equivalent expertise.

If done in this manner, corpus linguistics analysis need not be excessively time consuming and can produce greater insight into documentary meaning. But the same is true for contextual information, which suggests that both should be available for courts in deciding cases and that the former should not be privileged over the latter merely because it is more overtly word-related.

2. Veracity Concerns Regarding Contextual Evidence

Another oft-cited reason to avoid the use of context in interpretation are veracity factors. By veracity factors, I mean the concern that extra-textual information about documentary meaning may be suspect. This can occur for any number of reasons: self-interest of the source; faulty recollection; incomplete records; other sorts of spoliation (e.g., deteriorating real evidence). But the fact that some contextual evidence is

unreliable or unpersuasive hardly justifies an absolute bar to the use of contextual evidence.

Courts and the legal system are expressly designed to separate reliable and probative information from suspect information. They do this every day regarding important matters of criminal and civil law. Just as a court (often with the assistance of a jury) is trusted to determine guilt or innocence, the cause of an injury, the amount of fair compensation, or the intent underlying an action, the court is presumably competent to sort out the "wheat" of contextual information probative of meaning from the "chaff" of tainted, suspect, or misleading contextual information.

Consider legislative history, an important item of contextual information surrounding statutory language.[47] Contextualists consider both overall statutory purpose and specific legislative intent important in discerning textual meaning. But not all legislative history is created equal. For example, although floor statements are part of the "official" legislative history of a bill that becomes law, they have long been regarded as less reliable than other types of legislative history, such as committee reports or the joint statements of the bill co-sponsors—descriptions and discussions of the legislation that go beyond a single legislator and are subject to the discipline of response from others on a committee, its staff, or other sponsors of the bill.[48] By contrast, a single legislator can say anything he or she wants regarding a bill. No matter how inaccurate, the statement is in the congressional record, even if the author did not personally appear on the floor but merely submitted a prepared text to the clerk of the relevant chamber.

Discerning courts recognize that such statements can be used as "throwaway" gifts to a constituent or interest group and may not accurately describe legislation. Perhaps ironically, a good example of discerning differentiation of legislative history is provided by a noted textualist: Judge Frank Easterbrook, in a Seventh Circuit decision that rejected a construction of the statute proffered by the party relying on a single senator's floor

---

[47] *See* Stephen Breyer, *On the Uses of Legislative History in Interpreting Statutes*, 65 S. CAL. L. REV. 845 (1992) (defending frequent use of legislative history for ascertainment of statutory meaning); Leigh Ann McDonald, *The Role of Legislative History in Statutory Interpretation: A New Era After the Resignation of Justice William Brennan*, 56 MO. L. REV. 121 (1991) (reviewing jurisprudential debates between Justice Brennan and Justice Scalia in Public Citizen v. U.S. Dep't of Just., 491 U.S. 440 (1989) (finding that the ABA Judicial Evaluation Committee is not subject to Federal Advisory Committee Act based on majority's consideration of legislative history versus dissent's textual approach based on alleged plain meaning of term "utilized" in statute; noting the importance of legislative history as an indicator of meaning for less textually fixated judges such as the Court's moderate and liberal justices).

[48] *See infra* text accompanying note 49 (discussing regarding *Continental Can*).

statements (which post-dated enactment) in a case involving the eligibility of a pension fund for favorable regulatory treatment.[49]

3. Judicial Weighing of Legislative History and Context: The *Continental Can* Example

The Seventh Circuit considered the correct construction of the term "substantially all" where the question was whether that amount of contributions to the pension fund had come from trucking operations. If so, the company was eligible for favorable treatment in withdrawing from an underfunded pension fund and would not make contributions to the fund as the price of withdrawal.[50] The Court accepted some proffered legislative history but rejected other legislative history that it found not to be probative because it appeared to represent only the attempt of an individual senator to aid a supportive interest group.[51]

> Because language is an exercise in shared understanding, one Senator's idiosyncratic meaning does not count. This was the point of the exchange between Alice and Humpty Dumpty. If everyone accepts a new meaning for a word, then the language has changed; if one speaker chooses a private meaning, we have babble rather than communication.
>
> The question then, is whether anyone other than Senator Durenberger used "substantially all" to signify 50.1% rather than 85%. Nothing in the debates suggests that they did. Representative Thompson's statement, the only one delivered on the floor in advance of passage, shows that he at least (and likely the House) used the phrase in the customary way [of meaning 85%]. Senator Durenberger's two comments, inserted in the Congressional Record after the fact, could not have influenced anyone in the House and probably did not come to the attention of anyone in the Senate; anyway it was too late. Efforts of this kind to change the meaning of a text without bothering to change the text itself demonstrate why the

---

[49] *See* Cont'l Can Co. v. Chicago Truck Drivers, Helpers & Warehouse Workers Union Pension Fund, 916 F.2d 1154, 1156 (7th Cir. 1990). The panel consisted of not only Judge Easterbrook, who wrote the opinion, but also Judges Joel Flaum and Michael Kanne.

[50] *See id.* at 1155, 1159.

[51] *See id.* at 1156–60. The Fund argued that "substantially all" meant 85 percent, and an arbitrator initially deciding the dispute agreed, based on not only construction of the words but almost uncontradicted legislative history suggesting that Congress thought 85% was the correct numerical operationalization of the verbal term. *See id. at* 1155–57. Continental Can, which wished to stop paying and leave the pension fund since it had largely left the trucking business, argued that it was eligible to withdraw if a majority of the fund was composed of contributions from trucker employment. *See id.* at 1155. In other words, Continental Can argued, as the Court put it, that "50.1 percent" is the meaning of "substantially all," and cited as support the floor statements of a single U.S. Senator, David Durenberger (R-MN). *See id.* at 1155. 1157; *see also* John F. Manning, *Second-Generation Textualism*, 98 CALIF. L. REV. 1287, 1309 n.99 (2010) (describing *Continental Can* as a case in which the court revealed "how the Senate sponsor of a piece of legislation sought to use legislative history to manipulate its meaning").

> use of legislative history has come under such vigorous attack, even by former Senators.[52]

As the *Continental Can* decision illustrates, courts are (notwithstanding Judge Easterbrook's jab at legislative history generally) perfectly capable of distinguishing probative legislative history from the efforts of particular legislators or interest groups to "salt" or "plant" in legislative materials attempts to expand, contract, or even contradict the overall legislative purpose and specific intent of a bill. Floor statements in particular have been identified as a type of potentially misleading low-value legislative history.[53] But even floor statements may have significant probative value. For example, a statement by a bill's sponsor in response to a query or criticism regarding the bill at the outset of floor debate may (unlike the postenactment solidary views of a single senator in *Continental Can*) be quite informative regarding congressional understanding of a bill and expectations as to its application.[54]

Textualists often seize upon the potential pitfalls of legislative history as grounds for limiting or perhaps even spurning its use. Indeed, in *Continental Can* itself, Judge Easterbrook makes an eloquent textualist case[55]—but in the end considers legislative history to resolve the meaning of the term "substantially all,"[56] despite some criticism from another

---

[52] *See Cont'l Can Co.*, 916 F.2d at 1158.

[53] *See id.* at 1160 (Flaum, J., concurring) (labeling "post-enactment statements" as a "particularly disfavored form of legislative 'history'" (citing Pierce v. Underwood, 487 U.S. 552, 566–68 (1988); United States. v. United Mine Workers of Am., 330 U.S. 258, 281–82 (1947); United States v. Marshall, 908 F.2d 1312, 1318–19 (7th Cir. 1990)).

[54] For example, recent congressional activity has involved Democratic attempts to enact a higher minimum wage, with Sen. Bernie Sanders (I-Vt) a prominent proponent of the effort and maker of an unsuccessful motion to include the topic in the budget reconciliation process that was largely centered on the Biden Administration's proposal for a $1.9 trillion rescue plan for maintaining/spurring the economy despite the COVID-19 pandemic. *See* Alexander Bolton, *Senate Rejects Sanders $15 Minimum Wage Hike*, THE HILL (Mar. 5, 2021), https://thehill.com/homenews/senate/541826-senate-rejects-sanders-15-minimum-wage-hike [https://perma.cc/FN2C-L27D]. In evaluating any passage or rejection of such proposals, a floor statement by Senator Sanders or other prominent advocates logically has probative value, even if it may need to be read by with some caution and concern about posturing or self-serving statements. The same would be true of floor statements by prominent Republic opponents of the bill such as Sen. Mitch McConnell (R-Ky.). If the legislation had passed consideration of such floor statements should at least merit judicial consideration rather than reflexive rejection merely because this information about legislative intent and purpose is not embodied in the statutory language itself.

[55] *See Cont'l Can Co., 916 F.2d* at 1157 ("The text of the statute, and not the private intent of the legislators, is the law. Only the text survived the complex process for proposing, amending, adopting, and obtaining the President's signature (or two-thirds of each house). It is easy to announce intents and hard to enact laws; the Constitution gives force only to what is enacted.").

[56] *See id.* at 1158 ("[T]he text is law and legislative intent a clue to the meaning of the text, rather than text being a clue to legislative intent." (citing *In re* Sinclair, 870 F.2d 1340 (7th Cir. 1989)).

panelist.[57] As one textualist scholar summarized, although courts recognize that "legislative history has its problems and must be approached with caution" they also appear to accept that "no one has made the case for the inherent unreliability of such materials in all contexts."[58]

In short, legislative history is "data" that should be part of any conversation about data-driven interpretation. So is other contextual information that may credibly bear on word meaning. When considering words in private documents such as contracts, the case for context is even stronger because contracts are isolated agreements and not universal positive law. Contracting parties are making an agreement that is commonly memorialized in writing but the writing itself is not "the" contract[59] in the same manner that statutory text is "the" law.

In assessing the utility of corpus linguistics data relative to other data, one might look to cases like *Continental Can* and ask what types of data are most illuminating as to word meaning.[60] A corpus linguistics examination of what people mean when they use words like "substantially" or "substantially all" could be illuminating and almost certainly provides valuable information beyond what one gleans merely from reading a dictionary entry. But is such information any more illuminating than looking at legislative history?

The *Continental Can* court that spurned the implications of a postenactment floor statement also placed significant weight on one sponsor's pre-enactment operational definition (treating eighty-five percent as "substantially all")[61] that enjoyed analogous support in other areas of regulation.[62] This portion of *Continental Can* illustrates another arguable shortcoming of

---

[57] *See id.* at 1160 (Flaum, J., concurring) ("I write separately because of my reluctance to join what, with all due respect, I view as an unnecessary excursion into areas of legislative motive and functioning.").

[58] *See* Manning, *supra* note 51, at 1308.

[59] *See* DAVID EPSTEIN ET AL., MAKING AND DOING DEALS: CONTRACTS IN CONTEXT 12 (5th ed. 2018) ("People think of . . . a piece of paper as being a 'contract.' But the piece of paper is not a 'contract.' . . . At most, the piece of paper is a memorialization of the 'contract' . . . .").

[60] *See generally Cont'l Can Co.*, 916 F.2d 1154.

[61] *See id.* at 1156 (quoting statement by House floor manager Representative Thompson that "substantially all" is intended to mean 85% and noting "our intent that, as used in this special trucking industry withdrawal liability rule, the substantially all requirement would only be satisfied where at least 85 percent of the contributions to the plan are made by employers who are primarily engaged in the specified industries" (citing 126 CONG. REC. 23040 (Aug. 25, 1980) and noting unanimous passage of H.R. 3904 that same day).

[62] *See id.* at 1158 ("'Substantially all' may have a special meaning. Statutes contain words of art, whose meaning may appear strange to a lay reader. 'Substantially all' is one of those phrases with a special legal meaning. Congress uses it all the time in tax statutes, and the Internal Revenue Service decodes it as meaning 85%. Here are a few examples [listing ten]." (internal citations omitted)).

corpus linguistics: the most prominent corpora, such as COCA,[63] collect popular or layperson usage of language. Although this may be apt for statutes of general applicability such as criminal law prohibitions on conduct or contract language between generalist merchants, it may be misleading if the document under review is one using technical or specialized language. In deciding *Continental Can*, for example, it might have been helpful for the court to have at its disposal a corpus of congressional or governmental or regulatory use of terms. Arguably, the court achieved the same effect by looking at contextual data (taxation usage of the term "substantially all") that informed Congress as it enacted the statute.[64]

If, at the time of the decision (1990),[65] Judges Easterbrook, Flaum, and Kanne had access to the COCA and modern computing capacity,[66] should they have ignored the legislative history in favor of a corpus linguistics analysis? I think not. Further, although it would be interesting to conduct such an analysis, I would hope that courts would not be too quick to privilege corpus linguistics analysis over legislative history. Although the broader collective examples of usage offered by corpus linguistics can be helpful, I am skeptical that the casual use of a term by laypersons could be more informative about a statute than data directly pertaining to enactment of the statute by its authors and the enacting body. In sum, corpus linguistics analysis provides a useful and emerging tool in statutory interpretation.

In sum, corpus linguistics analysis provides a useful and emerging tool in statutory interpretation. Nevertheless, as demonstrated above, consideration of contextual factors should not be displaced by this data-driven approach, unless there are strong reasons to do so.

II. THE CONTOURS OF CORPUS LINGUISTICS: A TYPOLOGY

Corpus linguistics can be a valuable tool for discerning the meaning of words in dispute and can be particularly valuable where words are facially unclear and there is little or no reliable contextual information that illuminates meaning. But in other cases, corpus linguistics may not be a useful tool. Sometimes the terminology of disputed documentary language is clear—at least as a general matter—upon a mere reading of the term. In such

---

[63] *See Corpus of Contemporary American English*, *supra* note 4.

[64] *See Cont'l Can Co.*, 916 F.2d at 1158.

[65] *See id.* at 1154.

[66] *See Corpus of Contemporary American English*, *supra* note 4; *see also Corpus of Historical American English*, *supra* note 4.

cases, the task is not one traditionally regarded as "interpretation" (what do the words mean) but of "construction" (what legal effect will be given to the words and the documentary language as a whole). This section provides examples of these differing situations.

A. *Clear Language But An Unclear Path to Decision May Mean Corpus Linguistics Can Contribute Little to Judicial Resolution*

1. Assessing the Permissibility of "Snap" Removal

Judicial reaction to the practice of "snap" removal presents a dispute where the meaning of a word in a statute is literally or superficially clear, without any need for corpus linguistics analysis, but where it is not at all clear—to roughly half the judges that have considered the matter—whether the clear text should be taken literally in light of the history, purpose, and operation of the statute in question.

The term "snap" removal is derived by analogy to defendants that quickly wrest a case away from the state court in which it was filed by seeking removal before a defendant that is a citizen of the forum state is served with process.[67] Removal on the basis of diversity jurisdiction is subject to a limitation known as the "forum defendant" rule. If any defendant "properly joined and *served*" is a citizen of the state in which the action was brought, then the case may not be removed even though the requirements for diversity jurisdiction are met.[68]

In the last decade or so, defendants have increasingly deployed the tactic of removing a case before a forum defendant has been served as a means of avoiding the forum defendant limitation on removal. Entities that are frequent defendants have begun electronically monitoring state-court dockets. When they identify cases that qualify for diversity jurisdiction—but for which removal would be barred because of the presence of a forum defendant (i.e., a defendant that is a citizen of the forum state)—they immediately remove the case before the plaintiff has a chance to serve the forum defendant. These defendants justify this removal even though it is inconsistent with the long-standing forum defendant rule by pointing to the plain language

---

[67] *Examining the Use of "Snap" Removals to Circumvent the Forum Defendant Rule Before the Subcomm. on Courts, Intellectual Property, and the Internet of the H. Comm. on the Judiciary*, 116th Cong. 2 (2019) (statement of James E. Pfander, Professor, Northwestern University Pritzker School of Law).

[68] 28 U.S.C. § 1441(b)(2) (emphasis added).

of § 1441(b)(2), arguing that removal is not blocked unless and until the forum defendant is served.[69]

The district courts have split on whether this tactic is permitted. Slightly more than half of the district courts that have addressed the issue have held that snap removal is not valid and have remanded to state court.[70] They generally view the practice as contrary to Congress's likely intent as well as to the overall structure of removal jurisdiction. They argue that Congress could not possibly have intended for snap removal to be part of the removal scheme, in part because Congress could not possibly have foreseen electronic docket-monitoring, but more fundamentally because one struggles to identify a policy justification for not applying the forum-defendant bar just because a defendant manages to remove before the process server finds its mark.[71]

A nearly equal number of district courts have upheld the practice as permitted by the plain language of the statute.[72] The three Court of Appeals decisions that directly address snap removal upheld it on plain language grounds.[73] The original removal provision did not include the "properly joined and served" limitation, which Congress included in Title 28 in 1948 during a significant revision of the statutory removal provisions.[74] At that time, 28 U.S.C. § 1441(b)(2) was amended to its current language, which states that cases qualifying for diversity jurisdiction (and hence meeting the statutory requirement that original jurisdiction exist) "may *not* be removed if any of the parties in interest properly joined and served as defendants is a citizen of the State in which such action is brought."[75]

There appears to be consensus that the "properly joined and served" language of 28 U.S.C. § 1441(b)(2) was added to prevent

---

[69] *See* PFANDER, *supra* note 67.

[70] *See* Thomas O. Main et al., *The Elastics of Snap Removal: An Empirical Case Study of Textualism* 69 CLEV. ST. L. REV. 289, 306–12, (2021) ((manuscript at 6) (empirical examination of decisions regarding snap removal and remand; finding slight majority of trial courts remanded cases that were snap removed prior to service on forum defendant); Jeffrey W. Stempel et al., *Snap Removal: Concept; Cause; Cacophony; and Cure*, 73 BAYLOR L. REV. 423, 446–49 (2020)) (reviewing development of snap removal tactic); *see also* Little v. Wyndham Worldwide Operations, Inc., 251 F. Supp. 3d 1215, 1221–23 (M.D. Tenn. 2017).

[71] *See* Main et al., *supra* note 54, at 8–9; Stempel et al., *supra* note 70.

[72] *See* Main et al., *supra* note 54, at 15; *see, e.g.*, D.C. *ex rel.* Cheatham v. Abbott Labs., Inc., 323 F. Supp. 3d 991, 996–97 (N.D. Ill. 2018). ("Courts must give effect to the clear meaning of statutes as written. Thus, where possible, we begin and end our inquiry with the text, giving each words its ordinary, contemporary, common meaning." (internal citations and quotations omitted)).

[73] *See* Texas Brine Co. v. Am. Arbitration Ass'n 955 F.3d 482, 486 (5th Cir. 2020); Gibbons v. Bristol-Myers Squibb Co., 919 F.3d 699, 705 (2d Cir. 2019); Encompass Ins. Co. v. Stone Mansion Rest., Inc., 902 F.3d 147, 153 (3d Cir. 2018).

[74] Arthur Hellman et al., *Neutralizing the Stratagem of "Snap Removal": A Proposed Amendment to the Judicial Code*, 9 FED. CTS. L. REV. 103, 108 (2016).

[75] *See* 28 U.S.C. § 1441(b)(2) (emphasis added).

the abuse of the forum defendant rule[76] by improper or "fraudulent joinder" of a forum citizen defendant that was not seriously being pursued by the plaintiff but was only named in the suit to thwart removal.[77] However, the published legislative history regarding the 1948 changes is inconclusive. For example, there is neither a specific statement from Congress nor from the Advisory Committee on Revision of the Judicial Code (the Committee), regarding the addition of the "properly joined and served" language.[78]

---

[76] *See* Hellman et al., *supra* note 74. The basic purpose of the 1948 "and served" language is agreed upon by commentators although they may differ in their views of the propriety of snap removal. *See, e.g.*, Valerie M. Nannery, *Closing the Snap Removal Loophole*, 86 U. CINCINNATI. L. REV. 541, 547–48, 585 (2018) (finding snap removal invalid in light of overall diversity jurisdiction and forum defendant rule regime; criticizing literal reading of "and served" language in statute); Saurabh Vishnubhakat, *Pre-Service Removal in the Forum Defendant's Arsenal*, 47 GONZ. L. REV. 147, 163 (2011) (finding snap removal appropriate because of plain language of "and served" requirement despite tension with purpose of the language); Matthew Curry, Note, *Plaintiff's Motion to Remand Denied: Arguing for Pres-Service Removal under the Plain Language of the Forum-Defendant Rule*, 58 CLEV. ST. L. REV. 907, 931–32 (2010) (same); Jordan Bailey, Comment, *Giving State Courts the Ol' Slip: Should a Defendant Be Allowed to Remove an Otherwise Irremovable Case to Federal Court Solely Because Removal Was Made Before Any Defendant is Served?*, 42 TEX. TECH L. REV. 181, 214 (2009) (criticizing snap removal as inconsistent with overall structure of removal and diversity jurisdiction). This appears to be the judicial consensus as well, even for judges permitting snap removal. *See, e.g.*, Holmstrom v. Harad, No. 05 C 2714, 2005 WL 1950672, at *2 (N.D. Ill. Aug. 11, 2005); Brown v. Organon Int'l, Inc., No. 07-3092 (HAA), 2008 WL 2833294, at *4–5 (D. N.J. July 21, 2008); Brown v. Organon USA, Inc., No. Civ. A. 07-3092(HAA), 2008 WL 2625355, at *7 (D. N.J. June 27, 2008); Allen v. GlaxoSmithKline PLC, No. 07-5045, 2008 WL 2247067, at *6 (E.D. Pa. May 30, 2008); DeAngelo-Shuayto v. Organon USA, Inc., No. 07-2923(SRC), 2007 WL 4365311, at *5 (D. N.J. Dec. 12, 2007); *In re* Aredia & Zometa Products Liab. Litig., No. 3:06-MD-1760, 2007 WL 2905247, at *2 (M.D. Tenn. Oct. 3, 2007); Stan Winston Creatures, Inc. v. Toys "R" US, Inc., 314 F. Supp. 2d 177, 181 (S.D.N.Y. 2003); *see also* John P. Lavelle, Jr. & Erin E. Kepplinger, *Removal Prior to Service: A New Wrinkle or a Dead End?*, 75 DEF. COUNSEL J. 177, 177–80 (2008) (defense lawyer attracted to snap removal as a tactic recognizes split in opinion and tension between "and served" text and its purpose of preventing plaintiff gamesmanship without vitiating forum defendant rule resulting in divided trial court decisions).

[77] *See* Nannery, *supra* note76, at 548 ("[H]istorical context makes [the service language's] purpose evident: The purpose behind the addition of that language seems fairly clear—to bring into the statute the 'fraudulent joinder' doctrine and to restrict other tactics, like failing to serve a properly joined in-state defendant, which might otherwise be used to prevent removals which Congress had authorized." (quoting Champion Chrysler Plymouth v. Dimension Serv. Corp., No. 2:17-cv-130, 2017 WL 726943, at *2 (S.D. Ohio Feb. 24, 2017)).

The fraudulent joinder doctrine holds that defendants against whom the plaintiff has no plausible claim will be ignored in determining whether there exists complete diversity and corresponding federal diversity jurisdiction pursuant to 28 U.S.C. § 1332. *See In re* Briscoe, 448 F.3d 201, 216 (3d Cir. 2006). *But see* Abels v. State Farm Fire & Cas. Co., 770 F.2d 26, 32 (3d Cir. 1985) (plaintiff's motive of joining forum defendant to defeat diversity jurisdiction is not fraudulent joinder unless plaintiff lacks legitimate basis for claim against forum defendant). *See generally* James E. Pfander, *Forum Shopping and the Infrastructure of Federalism*, 17 TEMP. POL. & CIV. RTS. L. REV. 355 (2008) (defending forum shopping against common criticisms and noting that practice may enhance federalism and relative importance of state judicial systems).

[78] *See* 28 U.S.C. § 1441 (1948); H.R. Rep. No. 80-308 (1947), *as reprinted in* 1948 U.S.C.C.S., Special Pamphlet: Title 28 at 1692; S. Rep. No. 80-1559 (1948), *as reprinted in* 1948 U.S.C.C.S. Special Pamphlet: Title 28, 1675; Letter from Hon. Albert

Notwithstanding that the circuit court scorecard currently weighs in favor of snap removal, the practice is highly problematic, inconsistent with traditional notions of federalism and federal jurisdiction, and appears not to have been intended by Congress.[79] But regardless of whether the tactic is permitted by the removal statute, corpus linguistics is not useful in resolving the issue. The problematic words in the statute—"joined and *served*"—are not ambiguous. Any lawyer reading the words knows what "service" means, what is entailed, and the litigation significance of the word. The question is not one of word meaning but of what effect to give the word in light of other facts relevant to statutory construction.

Unsurprisingly, strict textualists merely apply the unambiguous word unless they view the result as "absurd," while contextualists are inclined not to permit a litigation tactic that is consistent with the bare language of the statute but completely inconsistent with the remainder of removal jurisprudence, which has long construed removal requirements narrowly in order to preserve the concept of the limited jurisdiction of federal courts.[80] Textualist and contextualists can debate at length with corpus linguistics providing no support—indeed no relevant evidence—to either side.

2. Documentary Deadlines as Another Example of Non-Utility for Corpus Linguistics

United States v. Locke[81] provides another example of the type of case where corpus linguistics appears to be of limited or no utility. The case concerned the Federal Land Policy and Management Act,[82] which provided that holders of certain mining claims involving federal land are required to file certain documents "prior to December 31" in order to renew their right to mine on federal land.[83] The Supreme Court applied this language literally and disallowed a renewal submission filed on December 31,[84] giving substantial bite to this "quintessential trap for the unwary," enforcing a deadline where "it seems wholly illogical to

---

B. Maris, Circuit Judge, United States Court of Appeals for the Third Circuit, and Chair of the Committee, to Mildrim Thompson, Jr., Esq. (May 10, 1946).

[79] *See* Nannery, *supra* note76, at 547–48; *see also* Stempel et al., *supra* note 70.

[80] *See* Thomas O. Main et al., *The Elastics of Snap Removal: An Empirical Case Study of Textualism* 69 CLEV. ST. L. REV. 289, 296 (2021). *See generally* John F. Manning, *The Absurdity Doctrine*, 116 HARV. L. REV. 2387 (2003) (arguing against "absurd result" exception to literal approach to statutory language).

[81] *See generally* United States v. Locke, 471 U.S. 84 (1985).

[82] *Id.* at 86; 43 U.S.C. §§ 1701–1787.

[83] *Locke*, 471 U.S. at 88–89; 43 U.S.C. § 1744.

[84] *See Locke*, 471 U.S. at 94–99. *But see id.* at 118–20 (Stevens, J., dissenting).

require someone to file something under severe penalty for default the day before the last day of the year."[85]

In cases like *Locke*, the words in question are clear. We all know what "December 31" means and what "before" means. In *Locke* itself, there was no factual ambiguity (e.g., an attempt to file at 11:59 only to be foiled by a defective time stamp, a power failure, or the like). Rather, the question was whether to give literal enforcement of this odd deadline or whether to read it as a scrivener's error that should not result in a permit holder's disproportionate forfeiture of mining rights.[86] The Court majority opted for literalism, with two dissenters arguing for a construction that saved the individual's property interest from total loss due to a one-day delay in filing.[87] Regardless of which view is right (as a contextualist who thinks law should "abhor a forfeiture," my sympathies are with the dissent), the decision is not one for which corpus linguistics has much to contribute. The words are clear. The hard question—indeed, the only question—is what to do with the words.

B. *Illuminating Language in a Helpful But Not Fully Determinative Manner*

1. *Muscarello v. United States*

Muscarello v. United States[88] has long been a topic of scholarly attention and has received renewed attention from corpus linguists.[89] The case involved a drug sale in which the defendant had a gun in the locked glove compartment of his truck. After his apprehension, the government sought to have his sentence enhanced pursuant to 18 U.S.C. § 924(c)(1), which provides that if a person "during and in relation to any crime of violence or drug trafficking crime . . . uses or carries a firearm"

---

[85] *See* ESKRIDGE, JR. ET AL., *supra* note 7, at 446–47. Making the seemingly harsh literalism of the case even harsher, is the history of the land and an arguable estoppel argument that the Court rejected, which may be unjust but not surprising in that estoppel generally is inapplicable against governments, at least for conduct or statements by lower level agents. *See id.* at 447 ("[T]he Locke family had been exercising rights to mine gravel on federal land since the 1950s. Knowing they were required to file papers in order to retain these rights, they sent their daughter to the nearest BLM office, which told there that claims had to be filed by December 31. The Lockes filed on that date, and the BLM [Bureau of Land Management] rejected the papers on the ground that they were too late[,]" a decision affirmed by the U.S. Supreme Court.).

[86] *See Locke*, 471 U.S. at 93.

[87] *See id.* at 96; *id.* at 117 (Powell, J., dissenting); *id.* at 117–19 (Stevens, J., dissenting).

[88] *See* Muscarello v. United States, 524 U.S. 125 (1998).

[89] *See, e.g.*, Lee & Mouritsen, *supra* note 4, at 845–48; Mouritsen, *supra* note 4, at 1926; *see also* Bernstein, *Half-Empirical Attitude*, *supra* note 3, at 13 (referring to *Muscarello* as "much-trodden case").

there shall "in addition to the punishment provided for such crime of violence or drug trafficking crime, be [an additional sentence of imprisonment] of not less than 5 years."[90] In effect, the drug dealer defendant found to have used or carried a firearm has an additional five years tacked on to whatever other sentence would be imposed for the drug crime.

Muscarello contended that the presence of a gun in his vehicle (particularly in a locked glove compartment) was not what is meant by the "use or carry" language in the statute. The Supreme Court rather quickly agreed that the gun (which remained in the glove compartment throughout the drug deal) had not been used. The Court divided sharply, however as to whether Muscarello had "carried" the gun during his commitment of the crime.[91]

Justice Breyer, writing for a 5–4 majority, held that the gun had indeed been carried.[92] Justice Ginsburg and fellow dissenters both disagreed regarding the most apt construction of the term "carry"[93] but, more importantly from a contextualist perspective, chided the majority for failing to apply the "rule of lenity" that is the supposed norm in criminal law statutory construction disputes.[94]

> Section 924(c)(1), as the foregoing discussion details, is not decisively clear one way or another. The sharp division in the Court on the proper reading of the measure confirms, "at the very least, . . . that the issue is subject to some doubt. Under these circumstances, we

---

[90] *See Muscarello*, 524 U.S at 126–27; 18 U.S.C. § 924(c)(1)(A)–(c)(1)(A)(i).

[91] *See Muscarello*, 524 U.S at 127–32. Prior to *Muscarello*:

> In *Smith v. United States*, 508 U.S. 223 (1993), the Court held that, when a person traded a gun for drugs, he "used" the gun in violation of § 924(c)(1). Justices Scalia's dissenting opinion contended that the ordinary meaning of § 924(c)(1), read in context, is that "uses a firearm" means "used the firearm as a weapon," and that in any event the issues was sufficiently doubtful for the rule of lenity to control the outcome in favor of the defendant. In *Bailey v. United States*, 516 U.S. 137 (1995), a unanimous Court held that the "use" element requires active employment of the firearm by the defendant, such that it did not apply to a defendant who had a gun in the trunk of his car in which illegal drugs were found or to a defendant who had a gun locked in a trunk in a closet in a bedroom where illegal drugs were stored.

See ESKRIDGE, JR. ET AL., *supra* note 7, at 655.

[92] *See Muscarello*, 524 U.S. at 139. The majority elaborates on its linguistic points. *Id.* at 127–32.

[93] *See id.* at 142–43 (Ginsburg, J., dissenting) ("Unlike the Court, I do not think dictionaries, surveys of press reports, or the Bible tell us, dispositively, what 'carries' means embedded in § 924(c)(1)."). The dissent elaborates on its linguistic points. *Id.* at 142–44.

[94] *See id.* at 148–50.

> adhere to the familiar rule that, 'where there is ambiguity in a criminal statute, doubts are resolved in favor of the defendant.'" [95]

In contending that Muscarello had indeed carried a gun in the commission of a crime, Justice Breyer and the majority engaged in a randomly de facto corpus linguistics analysis that included a broad survey of usage of the word "carry" and its variants such as "carries" or "carried," that has no apparent or explained basis for the opinion's selection or ranking of sources. As a result, the majority opinion has attracted criticism not only from civil libertarians, but also from corpus linguists for its lack of professional rigor and for error in discerning the apt construction of the term "carry."[96]

The dissent's critique avoids the majority's flailing search to find a broad connotation of "carry" that sustains conviction when substantive criminal law demands the opposite. The dissent also steps back from merely counting examples to address a larger concern posed by interpretive efforts.

> Unlike the Court, I do not think dictionaries, surveys of press reports, or the Bible tell us, dispositively, what "carries" means embedded in § 924(c)(1). On definitions, "carry" in legal formulations could mean, *inter alia*, transport, possess, have in stock, prolong (carry over), be infectious, or wear or bear on one's person. At issue here is not "carries" at large but "carries a firearm." The Court's computer search of newspapers is revealing in this light. Carrying guns in a car showed up as the meaning "perhaps more than one third" of the time. One is left to wonder what meaning showed up some two-thirds of the time. Surely a most familiar meaning is, as the Constitution's Second Amendment ("keep and *bear* Arms") and Black's Law Dictionary, at 214, indicate: "wear, bear, or carry . . . upon the person or in the clothing or in a pocket, for the purpose . . . of being armed and ready for offensive or defensive action in a case of conflict with another person."
>
> On lessons from literature, a scan of Bartlett's and other quotation collections [in the majority opinion] shows how highly selective the Court's choices are. If "[t]he greatest of writers" have used "carry" to mean convey or transport in a vehicle, so have they used the hydra-

---

[95] *See id.* at 148 (alteration in original) (quoting Adamo Wrecking Co. v. United States, 434 U.S. 275, 284–85 (1978)).

[96] *See* Lee & Mourtisen, *supra* note 4, at 846–48 (summarizing that "[t]o the extent that we view the question of ordinary meaning as involving statistical frequency, the analysis above tells us that carry on one's person is overwhelmingly the most common use, while carry in a care is a possible but far less common use"); Mourtisen, *supra* note 4, at 1946–48, 1951–66; *see also id.* at 1948 ("Justice Breyer is searching for only for sentences containing [the words] *carry*, *firearm*, and *vehicle*. In these circumstances, we might expect to see only sentences referring to the carry. That only one-third of the sentences returned contained this usage is actually surprising . . . Justice Breyer has merely confirmed that *carry* is *sometimes* used . . . but he has certainly not confirmed that *carry* is the *ordinary* meaning. In fact, because the search terms were skewed in favor of *carry*, the paucity of instances of *carry*, suggests that *carry*, is somewhat unusual.").

> headed word to mean, *inter alia*, carry in one's hand, arms, head, heart, or soul, sans vehicle. Consider, among countless examples:
>
> > "[H]e shall gather the lambs with his arm, and carry them in his bosom." The King James Bible, Isaiah 40:11.
> >
> > "And still they gaz'd, and still the wonder grew, That one small head could carry all he knew."
> >
> > O. Goldsmith, The Deserted Village, ll. 215–216, in The Poetical Works of Oliver Goldsmith 30 (A. Dobson ed. 1949).
> >
> > "There's a Legion that never was listed, That carries no colours or crest."
>
> R. Kipling, The Lost Legion, st. 1, in Rudyard Kipling's Verse, 1885-1918, p. 222 (1920).
>
> "There is a homely adage which runs, 'Speak softly and carry a big stick; you will go far.'" T. Roosevelt, Speech at Minnesota State Fair, Sept. 2, 1901, in J. Bartlett, Familiar Quotations 575:16 (J. Kaplan ed. 1992).
>
> These and the Court's lexicological sources demonstrate vividly that "carry" is a word commonly used to convey various messages. Such references, given their variety, are not reliable indicators of what Congress meant, in § 924(c)(1), by "carries a firearm."[97]

My point here is not to retread ground previously and better broken by others regarding *Muscarello*, but rather to note that even in this type of case (correctly) seized on by corpus linguistics advocates as an example of a decision that could be improved through a more structured and scientific analysis than that used by the Court, corpus linguistics may be of limited utility. The corpus linguistics critique of *Muscarello* has an affinity with the critique often made of legislative history: That the Court (particularly the majority) looked into a crowded field of verbiage

---

[97] *See Muscarello*, 524 U.S. at 142–44 (Ginsburg, J., dissenting) (alterations and emphasis in original) (internal citations omitted). In scholarly quarters, Justice Breyer's invocation of the King James Bible has achieved a level of criticism-cum-ridicule resembling that of Justice Harry Blackman's embarrassing "[O]de to [B]aseball," WOODWARD & ARMSTRONG, *supra* note 35, at 229, in the infamous Flood v. Kuhn, 407 U.S. 258, 285 (1972) (exempting Major League Baseball from antitrust law and permitting continued servitude of players on strength of ill-reasoned Fed. Baseball Club v. Nat'l League, 259 U.S. 200 (1922) and Toolson v. New York Yankees, Inc., 346 U.S. 356 (1953)). *See, e.g.*, WOODWARD & ARMSTRONG, *supra* note 35, at 224–31 (noting widespread bemusement at Justice Blackmun's star-struck approach to the sport (but apparently not to the players who wanted to be free from an oppressive league regime of player control) that included ribbing from another Justice who baited Blackmun about omission in his list of baseball greats); *see also* Mouritsen, *supra* note 4, at 1930–35 (criticizing Justice Breyer's majority opinion for error in positing that order of dictionary definitions of "carry" was based on prevailing "primary" usage when most dictionaries present definitions historically from oldest to newest use).

and spotted its "friends,"—in the form of uses of the word carry that fit with the personal preferences of the Justices.[98]

The majority in particular appears to have come to a view and then strung together a potpourri of usages that reflect similar perspectives, in particular the view that any sort of transit of a gun amounts to "carrying" the weapon for purpose of criminal culpability meriting an additional five years in jail.[99] To some extent this is the norm in judging and advocacy, even if this may be a painful admission for the legal profession. The brain excels at "confirmation bias" and assembles information to accord with its prevailing orientation, a trait most likely exacerbated by the advocacy of counsel who can be depended upon to collect examples of word usage that favor a client's position.

In perhaps similar fashion, I have a preferred reading of the term "carry a firearm" that is perhaps unduly the result of reading too many detective stories and watching too many police procedurals on television. Like the *Muscarello* dissenters[100] (and implicitly the dissenting Justice Scalia in *Smith v. United States*[101] who argued that "use" of a firearm means use of the firearm as a weapon), it seems clear to me that when an ordinary native English-speaking American in the late twentieth century (the sentence enhancement was enacted in 1968) speaks of carrying a gun, she is talking about carrying the gun on one's person, not in the vehicle used to take her somewhere, even the location of planned crime.

While I am happy that corpus linguistics analysis supports this view and that of the *Muscarello* dissenters, the case illustrates the limits not only of haphazard data collection about words, but also reveals the limits of even rigorous, thorough, and fair analysis by linguistic professionals. While assessing lay usage of a term through the lens of a large database adds potentially valuable knowledge, it is not a magic bullet. The court must also examine whether a specialized meaning was intended, the role of the disputed statutory term in the larger statutory scheme, relation to other substantive legal policies (e.g., the rule of lenity),[102] and the consequences

---

[98] *See* Patricia M. Wald, *Some Observations on the Use of Legislative History in the 1981 Supreme Court Term*, 68 IOWA L. REV. 195, 214 (1983) (attributing the analogy to D.C. Circuit Judge Harold Leventhal).

[99] *See* Lawrence M. Solan, *The New Textualist's New Text*, 38 LOY. L.A. L. REV. 2027, 2053 (2005) (viewing clash of majority and dissenting opinions as a judicial "food fight" more "than a serious argument among distinguished jurists").

[100] *See Muscarello*, 524 U.S. at 139 (Ginsburg, J., dissenting).

[101] *See* Smith v. United States, 508 U.S. 223, 244 (1993) (Scalia, J., dissenting).

[102] *See* Mouritsen, *supra* note 4, at 1917–18 ("In the midst of all of the *Muscarello* Court's lexical wrangling, it is easy to lose sight of what is at stake. Frank Muscarello set out

and ramifications of its decision. Even in this classic case of debate over the "ordinary" or plain meaning of a word, data analysis confined solely to word usage patterns cannot definitively resolve the dispute.

2. *Lachs v. Fidelity & Casualty Insurance*

Both the limits and potential of corpus linguistics analysis to enhance understanding of word usage, in contracts as well as statutes—particularly words that have fallen out of use—is illustrated in *Lachs v. Fidelity & Casualty Co.*[103] Sadie Bernstein, consulting a travel agent, booked a flight from Newark to Miami for the following day. When arriving at the airport, "[s]he purchased 'Airline Trip Insurance' in the sum of $25,000 from an automatic vending machine. She then went to the Consolidated Air Service counter and completed her flight arrangements" and "entered the plane . . . and in less than an hour was dead as the result of a crash."[104]

When the Bernstein estate sued to collect the insurance proceeds, the insurer asserted that flights by "non-scheduled" airlines, such as Miami Airline, Inc., were not within the grant of coverage.[105] Although both the court majority, which found coverage and ordered payment,[106] and the dissent, which regarded the claim as barred by the non-scheduled airline limitation,[107] bandied the term around at length, both did so without the discipline now available via corpus linguistics analysis.

The majority was less concerned about precise determination of the meaning of terms like "Civilian Scheduled Airline" and "non-scheduled" airline in the policy, terms the insurer read as foreclosing coverage because Miami Airline, Inc. was what people now typically describe as a charter service rather than a mainstream carrier such as American Airlines or United Airlines.

---

to sell marijuana" but had his "likely ten-to-sixteen month [enhanced] sentence . . . [extended] to a mandatory five years, plus the additional six-to-twelve months for his underlying drug offense" even though he never removed his gun from the glove compartment of his car.)

[103] *See generally* Lachs v. Fidelity & Cas. Co., 118 N.E.2d 555 (N.Y. 1954). I regard the case as overlooked not only because it is not well-known, but also because it is typically characterized as a case about flight insurance or vending machine/kiosk insurance when it in fact has much broader implications and illustrates operation of insurance policy construction based on the objectively reasonable expectations of the policyholder, and did so well before Professor Robert E. Keeton coined the term in a famous article, *Insurance Law Rights at Variance With Policy Provisions*, 83 HARV. L. REV. 961 (1970). *See* Jeffrey W. Stempel, Lachs v. Fidelity & Casualty Co. of New York: *Timeless and Ahead of Its Time*, 2 NEV. L.J. 319, 320 (2002).

[104] *Lachs*, 118 N.E.2d at 556–57.

[105] *Id.* at 559.

[106] *Id.* at 560.

[107] *Id.* at 563 (Fuld, J., dissenting).

The majority noted that according to plaintiff (the executor of the Bernstein estate), the setting in which the insurance policy was sold emphasized the words "Airline Trip Insurance," (including illustrations of a plane in flight) without qualification along with advertised policy limits of a maximum of $25,000.[108]

In what the court described as "*much smaller print*" that was allegedly obscured by the vending machine's presentation of the sample policy, the policy sold stated that it

> [c]overs first one-way flight shown on application (also return flight if round trip airline ticket purchased) completed in 12 months within or between United States, Alaska, Hawaii or Canada or between any point therein and any point in Mexico, Bermuda or West Indies on any *scheduled* airline. Policy void outside above limits. For 'international' coverage see airline agent . . . . This insurance shall apply only to such injuries sustained following the purchase by or for the Insured of a transportation ticket from . . . a *Scheduled Airline* during any portion of the first one way or round airline trip covered by such transportation ticket . . . in consequence of: (a) boarding, riding as a passenger in, alighting from or coming in contact with any aircraft operated on a regular or special or chartered flight by a *Civilian Scheduled Airline maintaining regular, published schedules and licensed for interstate, intrastate or international transportation of passengers by the Governmental Authority having jurisdiction over Civil Aviation* . . . . Defendant on its part points to the fact that there was hanging on the wall at the right hand end of the counter where decedent picked up her ticket, a fairly large sign—approximately three feet by four feetF—bearing the caption in large size capital letters: "Non Scheduled Air Carriers Authorized To Conduct Business In This Terminal" . . . . There is, of course, no proof that decedent ever saw the sign.[109]

As one might surmise from this descriptive setting of the factual table, the court majority was more concerned about whether the marketing arrangement adequately informed the decedent of any limitations on coverage. The insurer argued that a "Civilian Scheduled Airline" was an "air carrier which obtains a certificate of public convenience and necessity as provided in Section 401 of the Civil Aeronautics Act"[110] and that this term has

> a clear and definite meaning has caused it to bring forward and present an enormous amount of proof extrinsic to the policy including a statute, regulations, newspaper and magazine articles, etc. By this mountain of work it seems to us that defendant has established that "Civilian Scheduled Airline" is not at all free from ambiguity and vagueness if it were not so the contract of insurance itself would

---

[108] *Id.* at 557 (majority opinion).

[109] *Id.*at 557–58 (second and third alteration in original) (emphasis added).

[110] *Id.* at 560.

> disclose within its four corners the intent of the parties in entering into it.[111]

In other words, this insurer, nearly seventy years ago, was engaging in an informal version of corpus linguistics analysis, marshalling as many examples as it could of the usage of the terms "civilian scheduled airline" and its informal cousin a "scheduled airline." While this fell on comparatively deaf ears with respect to the majority, which took a more contextual, reasonable expectations approach that imposed coverage absent clear language to the contrary, particularly in view of the setting of the sale that raised reasonable expectations of coverage by the decedent policyholder (all good decisions in my view),[112] the dissent was very favorably influenced by the insurer's examples of use of the term in writings aimed at laypersons.[113]

The dissent focused relentlessly on the meaning of policy text mentioning "scheduled" and "non-scheduled" airlines, albeit with a rather different view of the setting of the sale than that held by the majority. The dissent's discussion (reproduced at length to give the reader a sense of the dissent's approach) stated:

> While it might have been sufficient to show that there are two separate and distinct classes of airlines, the *"scheduled"* and the *"nonscheduled,"* the insurer went far beyond that in this case. Recognizing, perhaps, the inconspicuousness of the obvious, defendant has marshalled a veritable mountain of material contained in statute and regulations, in opinions and reports, in newspapers and magazines to demonstrate that the term *"scheduled airline"* has gained a wide and general currency, and that it is a term of clear and precise meaning, which has become part and parcel of the ordinary person's everyday vocabulary. *Defined by any standard and from any point of view, and compressed into a sentence, it simply and solely denotes a common carrier permitted to operate, or to hold out to the public that it operates, one or more airplanes between designated points regularly, or with a reasonable degree of regularity, in accordance with a previously announced schedule .*
>
> Its origin is in the Civil Aeronautics Act and the pertinent regulations issued by the Civil Aeronautics Board . . . . The courts, as well as Congress and the Civil Aeronautics Board, have invariably recognized that the "scheduled airline" is a definite, well-understood class of carrier . . . . As a response to all this, plaintiff contends that, whatever meaning the term may have under the statute or under the regulations, that meaning, in the absence of an express incorporation of such material, does not bind one not chargeable with knowledge of administrative action. *The point might be well taken, if to the so-called*

---

111 *Id.*

112 *Id.*

113 *Id.* at. 562-63 (Fuld, J., dissenting).

> *"average" man upon whose knowledge plaintiff relies the term had a meaning which was unclear or different from that announced by the regulatory agency or the courts. Words may, of course, have different significations to different people, at different times or under different circumstances, but that is not the case insofar as the term "scheduled airline" is concerned. The meaning given it by law has, as it were, passed into the public domain, where, in line with the current zest for abbreviation, it is usually referred to as a "sked," to differentiate it from its counterpart, the "nonsked."*
>
> In point of fact, we might well have taken judicial notice of the term and its meaning, had the record not been as replete as it is with illustrations from the daily press and from popular magazines and books with large nationwide circulations. For instance, the president of a firm leasing planes to a non-scheduled carrier wrote in *Fortune* Magazine for August, 1949: "It is now only a year ago that my air-freight line, California Eastern Airways, Inc., leased one of its DC-4 ships to a 'large irregular' or non-scheduled carrier, popularly known as a nonsked . . . *Non-scheduled airlines* use the same well-proved planes used by all airlines. But Douglas DC-3's of the *scheduled airlines* carry twenty-one seats; those of the nonskeds, twenty-eight. Most of the DC-4's of the scheduled lines contain forty-four seats; we now have sixty-seven in ours". A feature article in the March, 1951, issue of *Cosmopolitan* Magazine captioned, "Don't Fly the Unscheduled Air Lines!", takes for granted the classification of, and the distinction between, scheduled and nonscheduled carriers. And in *Harper's* Magazine for May, 1949, we find this discussion of scheduled airlines and nonscheduled flights: "Including taxes, you would pay $113 for a *non-scheduled flight* from New York to Los Angeles, as against $181 on a *scheduled airline.* These carriers are non-scheduled in the sense that they are not permitted by the Civil Aeronautics Board to fly more than a limited number of trips between any two cities each month, and because they cannot, therefore, represent themselves to the public as running on regular time-tables."
>
> It is thus made evident that the general public encountered the term "scheduled airline" in the course of its normal and ordinary reading of newspapers and magazines, and the context in which the term appeared demonstrates that to the average person it had the same clear and definite meaning as it had for the federal agency which originally used it.[114]

The dissent's assembly of both technical and press usage of the terms "scheduled" and "non-scheduled" flights, aided by the briefing of insurer counsel,[115] is impressive. Sixty-five years

---

[114] *Id.* at 561–63 (third alteration in original) (emphasis added) (internal citations omitted).

[115] *Id.* at 560 (majority opinion). This was an impressive amount of work (and billing to the client) to devote to contesting a $25,000 flight/life insurance policy. Although this amount equates to nearly $240,000 in 2020 (lawyer billings would presumably reflect a similar increase), one wonders about the insurer's cost-benefit analysis in that it surely expended as much or more in defense costs in a case that appears (perhaps with the benefit of hindsight) to be anything but a sure winner and in a state that does not impose counsel fees on a losing party. *See* RANDY MANILOFF & JEFFREY STEMPEL, GENERAL LIABILITY

later, a court faced with the same questions could review a corpus linguistics examination submitted by counsel that may or may not support the dissent's view that, at least in the early 1950s "everyone" knew what these terms meant. This would be a considerable refinement of the type of debate that went on in *Lachs*, in which the majority found different usages sufficient to trigger the rule that ambiguities are resolved against the drafter (almost always the insurer and indisputably the insurer in this case) while the dissent saw a clear pattern, at least in the magazines it cited.[116]

But empirical data—even if done more systematically than by the *Lachs* Court and based on millions of documents—remains just data. An advocate or adjudicator still must decide what to do with the information. To illustrate this dilemma using the traditional judicial framework for interpreting contracts, if the understanding of the terms as set forth in the dissent was reflected in a corpus linguistics examination, the insurer would presumably argue that it was the beneficiary of a clearly written, widely understood limitation on coverage that precluded any judicial consideration of contextual factors or the policyholder's own understanding or reliance. Conversely, if the big data was inconclusive, the estate of the policyholder would argue that the policy was ambiguous on its face, opening the door to receipt of the extrinsic evidence favoring the decedent policyholder or triggering *contra proferentem* in her favor (or both).

Even if corpus linguistics big data supported the dissent, the policyholder's estate still would have the arguments the majority found persuasive: (1) The circumstances surrounding the sale of the flight insurance that could be read as creating an expectation of coverage in ordinary passengers; (2) that data-driven analysis may not be sufficiently strong to preclude deeming a term ambiguous; (3) the norm of insurance law that policy provisions narrowing or defeating coverage are treated as exclusions, even if not so denominated, and that exclusions are strictly construed against insurers with the burden to show applicability of the exclusion placed upon the insurer; (4) public policies favor victim compensation in order to avoid disproportionate forfeiture, and; (5) placing liability on the cheapest cost avoider.[117]

INSURANCE COVERAGE: KEY ISSUES IN EVERY STATE 317–18 (4th ed. 2018) (fee shifting available to policyholders in some cases but not to insurers except pursuant to civil rules of general applicability such as a rule against frivolous litigation).

[116] *See supra* notes 108–114 and accompanying text.

[117] *See generally Lachs*, 118 N.E.2d at 556–60.

In addition, flight insurance is a particularly profitable line of insurance for reasons that auger in favor of deciding close questions (and perhaps even not-so-close questions) in favor of coverage. Although premiums are comparatively low, it does not provide much coverage in that it insures the life of the policyholder only for the few hours of the flight.[118] By contrast, an ordinary life insurer is providing round-the-clock, 24/7 coverage for a premium that is much lower per minute of coverage.

Furthermore, the flight insurance business model involves very low overhead and minimal underwriting.[119] Fine. Bully for the flight insurer for restraining operating costs. But having chosen this business model that provides potentially misleading marketing and no agent that can correct consumer misunderstanding, it hardly seems unfair to hold the flight insurer to the consequences. Put another way, it seems unfair to permit an insurer to deny a claim based on the insurer's understanding of a specialized term that could not be read and reflected upon by a purchaser in an environment the insurer knew was not conducive to careful study of policy language.

Conversely, the dissent could counter that its linguistic analysis involved not only a data-driven account of layperson use of the terms "scheduled" and "non-scheduled" but also an assessment of the technical background and meaning of the terms. The majority would reply that technical or specialized understandings should not control against a claim by a consumer layperson with no prior course of dealing with the insurer. And then there is the matter of the arguably defective delivery of the policy, which took place because the insurer relied on vending machine sales rather than providing a sales counter where a prospective policyholder could obtain a legible copy of the policy and ask questions of the company representative.[120]

---

[118] *See id.* at 557 ("AIRLINE TRIP INSURANCE . . . .'[c]overs first one-way flight shown on application (also return flight if round tip airline ticket is purchased . . . .)'").

[119] In addition to providing only brief (the length of the plane flight) and limited (covering death only from a crash; death by illness unrelated to the flight or other injuries are not covered), airline flight insurance involves low overhead in that the insurer invests no resources in underwriting and few in claims (because there are so few claims). Consequently, the cost of the insurance per amount and time of coverage is very low and the product is very profitable. Or at least it was. Sales of the product have diminished as passengers came to view flying as safer. *See LIMRA: Behavior Economics and Life Insurance: People Respond to Emotions – Not Just Facts*, ADVISOR MAG. (Feb. 8, 2013), https://www.lifehealth.com/limra-behavioral-economics-and-life-insurance/ [https://perma.cc/WL5D-RW6Y]; *see also Flyer Beware: Is Travel Insurance Worth It?*, OFF. SENATOR EDWARD J. MARKEY (D. MASS.) (Aug. 21, 2018), https://www.markey.senate.gov/news/press-releases/flyer-beware-is-travel-insurance-worth-it [https://perma.cc/SDS7-PSJF] (concluding that the answer is no, and that broader based travel insurance that provides more comprehensive coverage than flight insurance, and covers injury and illness during an entire trip, is similarly profitable for insurers).

[120] *See generally Lachs*, 118 N.E.2d. at 556.

Thus, even in a type of case that would seem tailor-made for corpus linguistics, thorough examination does not resolve the dispute unless the court adopts a strictly textualist attitude and is so sure of the corpus linguistics results that it refuses to consider not only contextual factors bearing on meaning, but also the relevant body of substantive law applicable to the case. The "rules of the road" for resolving insurance coverage disputes would be applicable regardless of whether corpus linguistics analysis takes place and may limit the application or decisiveness of even well-done textual data analysis coming to a strong conclusion as to word meaning. The same holds for other areas of substantive law in which particular doctrine and practice may provide ground rules of interpretation that depart from a purely textual analysis. An example would be the preference for enforcing arbitration clauses, particularly in labor agreements.[121]

C. *Shedding More Determinative Light on Language But Continuing to Benefit from Context: Holy Trinity Church as Example*

Holy Trinity Church v. United States,[122] one of the most studied cases of statutory construction,[123] provides an example of a case where analysis driven by textual data can be extremely, perhaps even determinatively illuminating. Because the case has been so thoroughly examined at length by scholars with considerable expertise, I will address it only briefly.

The case involved a law prohibiting payment of transit expenses to import persons who would "perform labor or service of any kind."[124] The Holy Trinity Church faced sanctions because

---

[121] *See* Howsam v. Dean Witter Reynolds, Inc., 537 U.S. 79, 83 (2002) (consumer brokerage agreement); Groves v. Ring Screw Works, Ferndale Fastener Div'n, 498 U.S. 168, 173 (1990) (labor agreement); ESRIDGE, ET AL., *supra* note 7, at 1168 (Appendix B, collecting specific substantive law canons of construction).

[122] Holy Trinity Church v. United States, 143 U.S. 457 (1892).

[123] *See generally, e.g.*, Carol Chomsky, *Unlocking the Mysteries of* Holy Trinity*: Sprit, Letter, and History in Statutory Interpretation*, 100 COLUM. L. REV. 901 (2000); Adrian Vermeule, *Legislative History and the Limits of Judicial Competence: The Untold Story of* Holy Trinity Church, 50 STAN. L. REV. 1833 (1998).

[124] *Holy Trinity Church*, 143 U.S. at 458; *see also* Alien Contract Labor Law of 1885, ch. 164, § 1, 23 Stat. 332 ("[I]t shall be unlawful for any person, company, partnership, or corporation, in any manner whatsoever, to prepay the transportation, or in any way assist or encourage the importation or migration of any alien or aliens, any foreigner or foreigners, into the United states, its Territories, or the District of Columbia, under contract or agreement, parol or special, express or implied, made previous to the importation or migration of such alien or aliens, foreigner or foreigners, to perform labor or service of any kind in the United States, its Territories, or the District of Columbia. [However,] nothing in this act shall . . . be so construed as to prevent any person or persons, partnership, or corporation from engaging, under contract or agreement, skilled workman in foreign countries to perform labor in the United States in or upon any new industry not

it paid for the passage from England of a minister.[125] The Church argued that it was exempt from the law because ministers were not laborers performing labor within the meaning of the statute. The Supreme Court agreed in an opinion that, although widely regarded as correct on the law, has attracted criticism from textualists because it relied on the "spirit" of the law and legislative intent[126] and from secularists and non-Christians because Justice Brewer engaged in a bit of discriminatory-cum-jingoist rhetoric about America being a "Christian Nation" that privileged religion (and Christianity in particular), over other activities[127] and expressed classist notions about the superiority of intellectual work over physical exertion.[128]

A definitive "last word" on the Court's assessment of statutory text has probably been delivered by Professors Gales and Solan, who conducted extensive corpus linguistics analysis of the statutory words "labor" and "laborer" by looking to ordinary use of the terms during the relevant time period in a variety of databases, both general and specialized.[129] Although

---

at present established in the United States: *Provided*, That skilled labor for that purpose cannot be otherwise obtained; nor shall the provisions of this act apply to professional actors, artists, lecturers, or singers, nor to persons employed strictly as personal or domestic servants: *Provided*, That nothing in this act shall be construed as prohibiting any individual from assisting any member of his family or any relative or personal friend, to migrate from any foreign country to the United States, for the purpose of settlement here.").

125 *Holy Trinity Church*, 143 U.S. at 472. In 1888, the Church was "prosecuted and fined . . . for violating the law." Tammy Gales & Lawrence M. Solan, *Revisiting a Classic Problem in Statutory Interpretation: Is a Minister a Laborer?*, 36 GA. STATE U. L. REV. 491, 494 (2020).

126 *See Holy Trinity Church*, 143 U.S. at 459 ("It is a familiar rule, that a thing may be within the letter of the statute and yet not within the statute, because not within its spirit, nor within the intention of its makers. This has been often asserted, and the reports are full of cases illustrating its application. This is not the substitution of the will of the judge for that of the legislator, for frequently words of general meaning are used in a statute, words broad enough to include an act in question, and yet a consideration of the whole legislation, or of the circumstances surrounding its enactment, or of the absurd results which follow from giving such broad meaning to the words, makes it unreasonable to believe that the legislator intended to include the particular act.").

127 *See* Mokhtar Ben Barka, *The Christian Nation Debate and the U.S. Supreme Court*, 6 EUR. J. OF AM. STUD. 1, 2, 6-7 (2011) (noting controversy and religious division in U.S.); Steven K. Green, *Justice David Josiah Brewer and the "Christian Nation" Maxim*, 63 ALBANY L. REV. 427, 427-28 (1999) (noting controversy regarding this rhetoric in case); Alan M. Dershowitz, *Justice O'Connor's Second Indiscretion*, N.Y. TIMES (Apr. 2, 1989), https://www.nytimes.com/1989/04/02/opinion/justice-o-connor-s-second-indiscretion.html [https://perma.cc/HJ7C-FQ8S] (prominent Jewish law professor criticizes Justice Sandra Day O'Connor's seeming endorsement of Christian nation portion of *Holy Trinity*); *see also* Lynch v. Donnelly, 465 U.S. 688, 717–18 (1984) (Brennan, J., dissenting) (criticizing Christian Nation rhetoric in *Holy Trinity*).

128 *See* Green, *supra* note 27at 465–66, 471.

129 *See* Gales & Solan, *supra* note 125, at 502–03. As the authors summarized:

> Analysis of date from USSL [United States Statutes at Large], COHA [Corpus of Historical American English], COCA [Corpus of Contemporary American English], and Google Books suggests the following facts:

professing to "take no stand on the Supreme Court's actual holding,"[130] they find that it "appears to us that the Court got the ordinary meaning argument correct based on corpus analysis" and that "the history of usage [of the terms] in both statutory and ordinary language do support the Court's decision."[131]

The Gales-Solan analysis is something of a textual tour de force. It takes a case where corpus linguistics analysis is likely to be helpful: one where construction of a term is crucial to the decision but that resists "seat-of-the-pants" or "gut" understanding of the term by modern readers because of the age of the case and where legislative intent was not clearly set forth by Congress and where contemporary readers are not personally familiar with the background of the legislation.[132] It then

---

> • "Labor or Service" in either order was used through most of the nineteenth century to refer to manual, slave-like labor.
>
> • In legal contexts, the expression was used to specifically characterize slaves themselves. Article IV of the Constitution contains a fugitive slave provision that uses exactly that language and the Fugitive Slave Acts of 1793 and 1850 do so as well, as do other statutes from that era.
>
> It is therefore not clear whether to consider "labor or service" as a single construction [in which the words should be read as one term expressing a particular concept] meaning the type of work that slaves perform or as an ordinary compound disjunctive noun phrase [in which each word in the phrase should be applied independently]. That is, one can ask either about the expression "labor or service" or about "labor" and "service" separately. The Court seemed to choose the latter course, although we argue here that the former course may have been more faithful to the meaning of the term as understood at the time based on the following:
>
> • When the activities of clergy were represented in the corpora, "labor" was rarely used other than in "labor of the Lord" and similar expressions. Yet, while rare, the corpus demonstrated occasional use of "labor" in connection with the tasks that members of the clergy performed.
>
> • When the corpora demonstrated that clergy perform a "service," it was either in the sense of being "in the service of the lord" or providing a positive "service" to the community. However, the decision placed more focus on the former term—"labor"—than on the latter—"service."

Id. at 532–33.

[130] *Id.* at 533. With all due respect to Professors Gales and Solan, they may disclaim taking a position but one cannot read their article without coming away with the definite and firm conviction that the statute was not intended to apply to clergy and that the "labor"-related language in the statute did not include clergy. The data analysis of the article makes it impossible (in my view) for one to persuasively argue giving the word "labor" a broad reading that encompasses clergy. *See id.* at 520–31.

[131] *Id.* at 533.

[132] Although perhaps cause for concern about courts engaging in extra-record analysis, it seems indisputable that judges reviewing modern legislation (e.g., the Age Discrimination Act) or regulations (e.g., those implementing the Sarbanes-Oxley Act and other legislation seeking to shore up the financial system after the Great Recession of 2008) would bring to the analysis their own sense of the background and purpose of the

conducts a broad and rigorous analysis of word usage of the era that considers the identity of both drafter and reader of the statute. And it finds the data to weigh in favor of the Court's construction of the disputed term.

But even in this instance involving a highly text-based analysis of a statute written long ago, the enterprise of determining meaning is well-aided by consideration of contextual information. The *Holy Trinity Church* Court, whatever its other failings, did this, looking first at the legislative history of the law[133] and the purpose and objective of the law[134] as well as precedent. The Court found that examination of the particular legislative history and general background, purpose and objectives of the statue strongly suggests that when Congress referred to "labor" it meant what we now call manual labor and not what we now deem white-collar work.[135]

It appears uncontested that the law was passed to reduce the importation of "cheap" foreign labor undercutting the wages and employability of native workers. There was undoubtedly anti-immigrant prejudice in the mix as well but that also augers in favor of reading the law (so long as not violative of any civil rights mandates) to permit a white, Anglo-Saxon English minister to be paid traveling expenses to come to preach to a

---

statute or rule. Similarly, judges, especially those that conducted commercial litigation prior to taking the bench, would seem inevitably to view contract language in light of their own commercial experiences. Older documentary writings logically make this factor less important to a judge's assessment of word meaning.

133 *Holy Trinity Church*, 143 U.S. at 461–65.

134 *Id.* at 463 ("[A]nother guide to the meaning of a statute is found in the evil which it is designed to remedy; and for this the court properly looks at contemporaneous events, the situation as it existed, and as it was pressed upon the attention of the legislative body.") The Court further noted that

> [t]he motives and history of the act are matters of common knowledge. It had become the practice for large capitalists in this country to contract with their agents abroad for the shipment of great numbers of an ignorant and servile class of foreign laborers, under contracts by which the employer agreed, upon the one hand, to prepay their passage, while, on the other hand, the laborers agreed to work after their arrival for a certain time at a low rate of wages. The effect of this was to break down the labor market, and to reduce other laborers engaged in like occupations to the level of the assisted immigrant. The evil finally became so flagrant that an appeal was made to [C]ongress for relief by the passage of the act in question, the design of which was to raise the standard for foreign immigrants, and to discountenance the migration of those who had not sufficient means in their own hands, or those of their friends, to pay their passage.
>
> It appears, also, from the petitions, and in the testimony presented before the committees of [C]ongress, that it was this cheap unskilled labor which was making the trouble, and the influx of which Congress sought to prevent.

Id. at 463–64 (quoting United States v. Craig, 28 F.795, 798 (E.D. Mich.1886)).

135 *See id.* at 459–60.

congregation of white, Anglo-Saxon privileged Manhattanites. And, notwithstanding the perhaps cringeworthy "Christian Nation" rhetoric[136] of Justice Brewer, he is correct that the United States has almost always given religion favorable treatment under the law.[137]

Although the Court's elitist, class-based rhetoric,[138] like its "Christian Nation" religiosity, was unfortunate, its contextual assessment appears correct, as was its linguistic assessment. Perhaps more important, the textual and contextual are aligned, which strongly suggests the result is correct. Examination of seeming congressional intent and the background and general purpose of the law buttresses a reading of statutory text that now has been demonstrated to accord with usage of the time.

---

[136] *See id.* at 465–71 ("But beyond all these matters no purpose of action against religion can be imputed to any legislation, state or national, because this is a religious people. This is historically true. From the discovery of this continent to the present hour, there is a single voice making this affirmation . . . . There is no dissonance in these declarations. There is a universal language pervading them all, having one meaning; they affirm and reaffirm that this is a religious nation . . . . If we pass beyond these matters to a view of American life, as expressed by its laws, its business, its customs and its society, we find every where a clear recognition of the same truth. Among other matters note the following: The form of oath universally prevailing, concluding with an appeal to the Almighty; the custom of opening sessions of all deliberative bodies and most conventions with prayer; the prefatory words of all wills, 'In the name of God, amen;' the laws respecting the observance of the Sabbath, with the general cessation of all secular business, and the closing of courts, legislatures, and other similar public assemblies on that day; the churches and church organizations which abound in every city, town and hamlet; the multitude of charitable organizations existing everywhere under Christian auspices; the gigantic missionary associations, with general support, and aiming to establish Christian missions in every quarter of the globe. These, and many other matters which might be noticed, add a volume of unofficial declarations to the mass of organic utterances that this is a Christian nation. In the face of all these, shall it be believed that a [C]ongress of the United States intended to make it a misdemeanor for a church of this country to contract for the services of a Christian minister residing in another nation?"). *See id.* (also reviewing at length state constitutions, U.S. Constitution, government oaths of office and other indicia of the importance of Christianity and religion in American life).

[137] For example, the Supreme Court recently exempted religious gatherings from the public health-prompted restrictions otherwise applicable to gatherings of more than a specified number of persons. *See* Roman Catholic Diocese v. Cuomo, 141 S. Ct. 63, 65, 68–69 (2020). There has not been a similarly successful secular challenge even though the First Amendment presumably accords non-religious gatherings freedom of expression on a par with freedom to worship. In addition, many jurisdictions as a matter of legislative decree or regulatory enforcement have permitted religious exceptions to COVID-related restrictions on assembly. *See* Virginia Villa, *Most States Have Religious Exemptions to COVID-19 Social Distancing Rules*, PEW RES. CTR (Apr. 27, 2020), https://www.pewresearch.org/fact-tank/2020/04/27/most-states-have-religious-exemptions-to-covid-19-social-distancing-rules/ [https://perma.cc/YG2A-7A27]; *see also* Wisconsin v. Yoder, 406 U.S. 205, 234 (1972) (permitting Amish parents to avoid state mandatory education requirements).

[138] *See Holy Trinity*, 143 U.S. at 463–64 ("No one reading such a title would suppose that [C]ongress had in its mind any purpose of staying the coming into this country of ministers of the gospel, or, indeed, of any class whose toil is that of the brain . . . . It was never suggested that we had in this country a surplus of brain toilers, and, least of all, that the market for the services of Christian ministers was depressed by foreign competition. Those were matters to which the attention of [C]ongress, or of the people, was not directed.").

In addition, although subsequent legislative history may be viewed as suspect or even out-of-bounds, it bears mentioning that in 1891 Congress amended the law by adding to the exceptions of its reach "ministers of any religious denomination" as well as "persons belonging to any recognized profession," and "professors for colleges and seminaries."[139] Absent a radical change in the composition and orientation of Congress, it seems likely that these amendments were merely clarifying the original intent of the statute rather than making a policy decision to add new exceptions.

Holy Trinity Church is thus well supported by contextual information with no apparent contradictory contextual evidence. Because of the array of corpora used by Gales & Solan, one might even go so far as to characterize the word usage data marshalled to assess *Holy Trinity Church* as "contextual" data as much at it is "textual" data in that it examines word use in a variety of settings.[140] One might even characterize corpus linguistics and related data collection and analysis as contextual.

CONCLUSION: IMPROVING CORPUS LINGUISTICS BY AVOIDING ISOLATION

Corpus linguistics and related data analysis of word usage naturally focuses more on text than does other contextual information such as the background, purpose, or drafting history of documentary text. But that does not negate the contextual nature of corpus linguistics analysis. The assembly of so many examples of word usage in different settings is part of what can make a corpus a significant improvement over a dictionary.

To be sure, corpora are amalgamations of text—but what makes corpora useful is not only the digital searchability of the text but also the information provided in the corpora regarding the circumstances in which the word is used and the connotation of word meaning in various settings. Although not providing background and context regarding the documents at issue in litigation, corpus linguists provides context about word usage. If word use context is germane to interpretation, context more specific to the document at issue in a case must by germane as well.

Corpus linguistics presents a useful advance for legal interpretation. But usefulness does not make the method a foolproof exclusive route to sound decision-making.[141] Examining

---

[139] *See* Gales & Solan, *supra* note 125, at 494.

[140] *See id.* at 505–33.

[141] A limitation recognized by legal corpus linguists. *See* Mouritsen, *supra* note 4, at 1970 ("The corpus method is not a panacea. The use of corpus data will not do away

only a few cases, some of which have been subjected to extensive textual analysis, strongly suggests that better judicial decisions—e.g., *Holy Trinity Church* and *Continental Can*—accord with contextual indicia of meaning[142] while less satisfying decisions—e.g., *Locke*, cases permitting snap removal, *Muscarello*—are those overly preoccupied by text to the exclusion of other consideration.

A more data-driven analysis of text might have rescued the *Muscarello* majority from embarrassment, but its real failing was insufficient concern for the role of the rule of lenity. And in *Locke* and cases permitting snap removal, computerized assessment of word usage would not be able to rescue relentlessly literal textualist judges from themselves.

To the extent corpus linguistics is promoted as a means of making textualism sufficiently scientific to justify resistance to considering context, its potential becomes peril. If corpus linguistics is not adequately constrained by intellectual modesty and supplemented by consideration of context, its potential remains unrealized, or perhaps even destructive.

---

with disagreements as to the meaning of statutory terms. Instead, the corpus method removes the determination of ordinary meaning from the black box of the judge's mental impression and renders the discussion of ordinary meaning one of tangible and quantifiable reality."). My amendment to this concession would include a further concession that disagreements as to meaning should be addressed by consideration of contextual information as well as improved textual analysis.

[142] Or are, like *Lachs v. Fidelity & Cas. Co*, supported by strong non-textual considerations relevant to the subject matter of the dispute or other legal doctrine bearing on the dispute.