

5-5-2022

Technologically Improving Textualism

Jeffrey W. Stempel

University of Nevada, Las Vegas – William S. Boyd School of Law

Erik S. Knutsen

Follow this and additional works at: <https://scholars.law.unlv.edu/nljforum>



Part of the [Contracts Commons](#), [Insurance Law Commons](#), [Judges Commons](#), and the [Jurisprudence Commons](#)

Recommended Citation

Stempel, Jeffrey W. and Knutsen, Erik S. (2022) "Technologically Improving Textualism," *Nevada Law Journal Forum*: Vol. 6, Article 2.

Available at: <https://scholars.law.unlv.edu/nljforum/vol6/iss1/2>

This Article is brought to you by the Scholarly Commons @ UNLV Boyd Law, an institutional repository administered by the Wiener-Rogers Law Library at the William S. Boyd School of Law. For more information, please contact youngwoo.ban@unlv.edu.

TECHNOLOGICALLY IMPROVING TEXTUALISM

By Jeffrey W. Stempel* and Erik S. Knutsen**

The textualist approach to construing statutes, regulations, contracts, and other documents remains dominant but has drawbacks, most significantly its tendency to disregard probative evidence of textual meaning in favor of isolated judicial impressions and dictionary definitions. Although a broader, contextual, “integrative” approach to interpretation is preferable, the hegemony of textualism, even extreme textualism, is unlikely to recede soon.

Textualism can be substantially improved, however, through effective use of a form of big data—the corpus linguistics approach to discerning word meaning. By enlarging the universe of sources about how words are actually used, corpus linguistics represents a significant improvement over imperial judicial pronouncements about word meaning along with episodic and inconsistent use of dictionary definitions for deciding cases. If deployed as tool of textualism rather than formulaic use of a bigger dictionary, corpus linguistics analysis can, at a minimum, serve as a useful supplement to traditional textualist tools.

INTRODUCTION

Textualism is the dominant mode of construing legal text (statutes, regulations, insurance policies, other contracts, and other documents),¹

* Doris S. & Theodore B. Lee Professor of Law, William S. Boyd School of Law, University of Nevada Las Vegas.

** Professor of Law, Queen’s University Faculty of Law-Canada.

Thanks to Ken Abraham, Tom Baker, Bill Boyd, Christoph Brömmelmeyer, Jay Feinman, Chris French, Sara Gordon, Dan Hamilton, Yong Han, Helmut Heiss, Randy Maniloff, David McClure, Ann McGinley, Miguel Martinez Munoz, Dan Schwarcz, Jeff Thomas, Abel B. Veiga Copo, and Oliver William, as well as our colleagues in the American College of Coverage Counsel, the Project Group for the Principles of Reinsurance Law (PRICL), and the ALI Restatement of the Law of Liability Insurance processes and participants in the Third Congreso Internacional de Derecho del Seguro (El Seguro de Personas y la Inteligencia Artificial). The opinions expressed in this article are of course our own and should not be attributed to any of those we cite or thank. In Memory of David Herr (1950–2021), a brilliant lawyer with no fear of math, computers, or rigor. © 2022 Erik S. Knutsen and Jeffrey W. Stempel.

¹ See JOSEPH M. PERILLO, CONTRACTS § 3.12 (7th ed. 2014); E. ALLAN FARNSWORTH, CONTRACTS § 1.8 (4th ed. 2004); ANTONIN SCALIA & BRYAN GARNER, READING LAW: THE INTERPRETATION OF LEGAL TEXTS 6–7 (2012); John F. Manning, *Textualism and Legislative Intent*, 91 VA. L. REV. 419, 419–21 (2005). Compare AMERICAN LAW INSTITUTE, RESTATEMENT THE LAW, LIABILITY INSURANCE § 3 (AM. L. INST. 2018) (“RLLI”) (adopting a “plain meaning” approach to construing insurance policy text that, although not hyper-textualist, privileges policy text relative to policy purpose, background information, and

notwithstanding the methodology's significant drawbacks.² Although a broader, more contextual and nuanced approach to the legal interpretation of disputed text³ would improve adjudication, textualism remains the dominant approach⁴ and enjoys continuing widespread support.⁵

Textualism can be particularly problematic when judges eschew a sophisticated and nuanced reading of text by “outsourcing” interpretation to dictionary definitions, which can become a mechanical exercise that encourages robotic interpretation rather than full consideration of factors affecting textual meaning.⁶ Although a move away from textualism—particularly hyper-literal textualism or textualism that refuses to acknowledge ambiguity and consider non-textual information—would be a welcome jurisprudential development, current judicial sentiment remains highly text-centric.

drafting history), with RESTATEMENT (SECOND) OF CONTRACTS §§ 200–229 (AM. L. INST. 1981) (taking a more “contextualist” approach to interpretation).

² See *infra* Part I; Jeffrey W. Stempel & Erik S. Knutsen, *Textual Literalism, Formalism, and Canon Centric Construction as De facto—and Defective—Artificial Intelligence*, in III CONGRESO INTERNACIONAL DE DERECHO DEL SEGURO (EL SEGURO DE PERSONAS Y LA INTELIGENCIA ARTIFICIAL) (Abel B. Veiga Copo & Miguel Martinez Munoz eds., forthcoming 2022) (Spain) [hereinafter Stempel & Knutsen, *Textual Literalism*]. See generally Jeffrey W. Stempel & Erik S. Knutsen, *Rejecting Word Worship: An Integrative Approach to Judicial Construction of Insurance Policies*, 90 U. CINN. L. REV. 561 (2021) [hereinafter Stempel & Knutsen, *Rejecting Word Worship*].

³ See Stempel & Knutsen, *Rejecting Word Worship*, *supra* note 2, at 622.

⁴ See SCALIA & GARNER, *supra* note 1, at 53–55. A former law professor (writing primarily in the area of administrative law) and judge of the U.S. Court of Appeals for the District of Columbia prior to his appointment to the Supreme Court, Justice Scalia became a particularly pronounced advocate of a strict form of textualism that eschewed consideration of legislative history and consideration of extrinsic evidence of word meaning absent facial ambiguity of the law or legal document in question that could not be resolved through textual analysis aided by traditionally recognized canons of construction. See generally ANTONIN SCALIA, *A MATTER OF INTERPRETATION: FEDERAL COURTS AND THE LAW* (1997) (supporting strict textualism); Antonin Scalia, *The Rule of Law as a Law of Rules*, 56 U. CHI. L. REV. 1175 (1989) (arguing for formalist jurisprudence as well as textualist jurisprudence).

⁵ This includes the newest members of the Supreme Court, appointed by former President Donald Trump in large part not only because of their conservative ideology but also because of their more formalist, textualist approach to legal interpretation. Trump-appointed Justices Neil Gorsuch, Brett Kavanaugh, and Amy Coney Barrett all have professed support for textualism. Although textualism is generally associated with conservative ideology and case outcomes, this is of course not universally the case. For example, in *Bostock v. Clayton County*, 140 S. Ct. 1731 (2020), Justice Gorsuch applied rather strict textualist analysis in the Court's opinion finding that Title VII of the 1964 Civil Rights Act extended protection to gay and transgender persons. Further, of course, textualists are not always consistent in their resistance to consideration of non-textual evidence of meaning such as background context. See, e.g., *id.* at 1822 (Kavanaugh, J., dissenting) (basing opposition to *Bostock* majority opinion in part on non-textual factors, such as legislative intent and social atmosphere at the time of enactment, extrinsic evidence that is arguably a departure from his generally professed textualism); Brett Kavanaugh, *Fixing Statutory Interpretation*, 129 HARV. L. REV. 2118 (2016) (reviewing the late Second Circuit Judge Robert Katzmann's book, *Judging Statutes*). See generally ROBERT A. KATZMANN, *JUDGING STATUTES* (2014).

⁶ See Stempel & Knutsen, *Textual Literalism*, *supra* note 2, manuscript at 17 (on file with authors).

But even if textualism cannot be displaced or expanded to include additional information bearing on the meaning of statutes, contracts, and documents, it can be made more rigorous, useful, and accurate. Although dictionaries provide some insight into word meaning, they are no panacea, particularly if used to provide a (pardon the pun) definitive answer rather than simply information for textualist processing by the court. More promising, if one is to be restricted to reading text alone, is the corpus linguistics approach of going beyond dictionaries and examining word meaning as reflected in large databases containing written material that actually uses the word at issue in multiple contexts.

I. TEXTUALISM AND ITS DRAWBACKS

Textualism addresses both *interpretation*, which focuses on finding linguistic meaning of text, and *construction*, which is the process of giving legal effect to text.⁷ Writings of legal significance require courts to determine the “plain meaning” of text.⁸ But texts consuming legal attention are often not clear and require construction that may not spring self-evidently from the face of the document.

Although dominant now, textualism appears not to have become dominant until the twentieth century, displacing focus on legislative history and intent (or party intent for contracts), which had been the focus of nineteenth century analysis. Prior to that, purposivism—construction according to the objective of the legal text—appears to have been dominant.⁹ For example, in the well-known *Heydon’s Case* of sixteenth century England, the court stressed that its role was to “suppress the mischief” a statute sought to alleviate and “advance the remedy” provided by the law.¹⁰

And prior to emphasis on textualism that became dominant in the latter half of the twentieth century, the first half of the century found courts open to balancing public policy factors affecting interpretation and construction.¹¹ But

⁷ See Lawrence Solum, *The Interpretation—Construction Distinction*, 27 CONST. COMMENT. 95, 96 (2010).

⁸ See RLLI, *supra* note 1, § 3 cmt. a (AM. L. INST. 2018) (endorsing text-based plain meaning approach as primary means of construing insurance policies and, in Comment a. and Reporters’ Note a., characterizing plain meaning as traditional approach to contract construction but also noting substantial support for the contextualist approach generally reflected in RESTATEMENT (SECOND) OF CONTRACTS §§ 200–229 (AM. L. INST. 1981)); PERILLO, *supra* note 1, § 3.10 (characterizing plain meaning approach, which holds that if writing is “plain and unambiguous on its face, its meaning must be determined from the four corners of the instrument without resort to extrinsic evidence of any kind,” as the “dominant” rule of contract construction); FARNSWORTH, *supra* note 1, at §§ 7.10, 7.11 (favoring more contextual, purposive approach but acknowledging the importance of construing text and presumption of applying unambiguous text).

⁹ William S. Blatt, *The History of Statutory Interpretation: A Study in Form and Substance*, 6 CARDOZO L. REV. 799, 800 (1985).

¹⁰ *Heydon’s Case*, 76 Eng. Rep. 637 (Exch. 1584).

¹¹ See generally HENRY HART, JR. & ALBERT SACKS, *THE LEGAL PROCESS: BASIC PROBLEMS IN THE MAKING AND APPLICATION OF LAW* (William N. Eskridge, Jr. & Philip P. Frickey eds.,

during the latter third of the twentieth century, textualism, including a particularly literal and formal brand of textualism, ascended.¹²

Textualism in its traditional form involved what scholars have termed “plain meaning” jurisprudence. The court focused on documentary text (be it in the Constitution, a statute, regulation, contract, or other document) and, if finding the text clear, ended the interpretative inquiry and applied the reading given to the text to decide the case. The text, if sufficiently clear, was given nearly exclusive primacy but, in contrast to more extreme textualism, left some room for consideration of other indicators of meaning and some possibility of countermanding text initially read as plain.¹³

During the last portion of the twentieth century and continuing into the twenty-first, United States courts expressed greater confidence in their ability to achieve a clearly correct reading of text alone, making the plain meaning presumption more ironclad and nearly impossible to displace by non-textual indicia of meaning. Canons of construction, rather than extrinsic information, were applied to aid in interpreting language, with canons and dictionary definitions becoming almost talismanic (even though canons and dictionaries are by definition non-textual or extra-textual). In addition, there was more effort to give a word the same meaning not only in a particular statute at issue

1994); GUIDO CALABRESI, *A COMMON LAW FOR THE AGE OF STATUTES* (1982) (advocating judicial freedom to treat statutes, particularly older or idiosyncratic statutes, as common law).

¹² At the U.S. Supreme Court level, the process began during the Burger Court (which tended to follow the “plain meaning” textualism described below) and increased during the period of the Rehnquist Court (which Justice Scalia joined almost upon inception in 1986), accelerating during the Roberts Court era, which witnessed a surge in citation of dictionaries and resistance to consideration of extra-textual information bearing on text meaning. Increasing numbers of Republican-appointed Justices brought to the court a preference for textualism as well as a conservative ideological and jurisprudential bent. *See* WILLIAM N. ESKRIDGE, JR. ET AL., *CASES AND MATERIALS ON LEGISLATION AND REGULATION: STATUTES AND THE CREATION OF PUBLIC POLICY* Ch. 5 (6th ed. 2020) (tracing history of statutory interpretation theory and noting gradual decrease in focus on statutory purpose and legislative intent and increased focus on statutory text).

Contract interpretation reflected a similar shift from a purposive approach receptive to context and extrinsic evidence to one more narrowly focused on documentary text. *Compare* RESTATEMENT (SECOND) OF CONTRACTS §§ 200–223 (AM. L. INST. 1981) *with* RESTATEMENT OF LIABILITY INSURANCE § 3 (AM. L. INST. 2018) (endorsing a “plain meaning” approach to construing insurance policies where contextual information and extrinsic evidence are not considered if the court deems policy text to have a plain meaning on its face). *See* Lemley, *infra* note 22; *see also* RESTATEMENT OF LIABILITY INSURANCE § 3 cmt. a. (AM. L. INST. 2018) (noting “two main approaches to the interpretation of contracts . . . : the contextual approach and the plain-meaning approach.”). The Restatement of Liability Insurance, promulgated forty years after the Restatement (Second) of Contracts, opts for a more text-centered plain meaning approach.

¹³ *See* ESKRIDGE, ET AL., *supra* note 12, at 499, 521–22 (describing new textualism and distinguishing it from plain meaning and noting general ascendance of textualism) (“In the 1980s, a group of judges and executive officials developed a more constrained version of the plain meaning rule [that] argued that courts interpreting statutes have no business figuring out legislative intent . . .”).

but throughout codified law. Scholars have termed this the “new textualism”—highly resistant to extrinsic evidence (save for canons and dictionaries).¹⁴

To be sure, plain meaning textualism, and particularly new textualism, has critics,¹⁵ but they are largely not on the bench. Federal courts are heavily textualist (resulting from the influence of political appointments), and most state courts are dominated by textualism as an interpretative methodology.¹⁶

As we have elaborated elsewhere, textualism, particularly the more austere new textualism, has several flaws.¹⁷ For example, courts commonly refer to text as “plain,” “clear,” and “unambiguous”—but don’t explain *why* the text is clear¹⁸ when it often is not clear to others or is read differently by others.¹⁹ One

¹⁴ See generally William N. Eskridge, Jr., *The New Textualism*, 37 UCLA L. REV. 621 (1990); SCALIA & GARNER, *supra* note 1 (defending narrow focus on text and use of canons and dictionaries in lieu of non-textual or extra-textual information bearing on meaning); John Manning, *The Absurdity Doctrine*, 116 HARV. L. REV. 2387 (2003) (arguing for literal application of textual meaning without regard to whether it produces an “absurd result”); John C. Nagle, *Textualism’s Exceptions*, 2 ISSUES LEGAL SCHOLARSHIP (2002) (same).

¹⁵ See, e.g., FRANK CROSS, *THE THEORY AND PRACTICE OF STATUTORY INTERPRETATION* (2009); William N. Eskridge, Jr. & Philip P. Frickey, *Statutory Interpretation as Practical Reasoning*, 42 STAN. L. REV. 321, 321–22 (1990).

¹⁶ For example, although it also has decisions taking a more contextualist, eclectic view of interpretation, Nevada caselaw skews toward a textualist approach. See, e.g., *William v. United Parcel Servs.*, 302 P.3d 1144, 1147 (Nev. 2013) (“In the absence of an ambiguity, we do not resort to other sources, such as legislative history.”); *Cromer v. Wilson*, 225 P.3d 788, 790 (Nev. 2010) (“When a statute is clear and unambiguous, we give effect to the plain and ordinary meaning of the words.”); *Kaldi v. Farmers Ins. Exch.*, 21 P.3d 16, 21 (Nev. 2001) (“[W]here a written contract is clear and unambiguous on its face, extraneous evidence cannot be introduced to explain its meaning.” (quoting *Geo. B. Smith Chemical v. Simon*, 555 P.2d 216, 216 (Nev. 1976))); *Siggelkow v. Phx. Ins. Co.*, 846 P.2d 303, 304 (Nev. 1993) (holding that contract terms should be “viewed in their plain, ordinary and popular sense”); *Reno Club v. Young Inv. Co.*, 182 P.2d 1011, 1014 (Nev. 1947) (holding that in the absence of clear evidence of a different intention, words must be presumed to have been used in their ordinary sense, and given the meaning usually and ordinarily attributed to them; finding option agreement to be “in ordinary and plain language” with a “meaning [that] seems clear”); see also *Nev. State Democratic Party v. Nev. Republican Party*, 256 P.3d 1, 5 (Nev. 2011) (“[W]hen a statute is facially clear, a court should not go beyond its language in determining its meaning.” (citation omitted)); *Lowe Enters. Residential Ptnrs., L.P. v. Eighth Jud. Dist. Ct.*, 40 P.3d 405, 411 (Nev. 2002) (finding that where statute’s language is “plain and unambiguous” and “its meaning clear and unmistakable,” there “is no room for construction” or consideration of material beyond the statutory language itself, but where a statute is ambiguous, the plain meaning rule has no application); *Nev. Mining Ass’n v. Erdoes*, 26 P.3d 753, 758 (Nev. 2001) (using “clear statement of [legislative] intent” to resolve “any ambiguity inherent” in statutory language at issue regarding meaning of “120 calendar days” following commencement of legislative session for determining deadline for conclusion of session; adjusting for daylight savings time to conclude that two bills were enacted before expiration of session).

¹⁷ See, e.g., Stempel & Knutsen, *Textual Literalism*, *supra* note 2; Stempel & Knutsen, *Rejecting Word Worship*, *supra* note 2.

¹⁸ See Jeffrey W. Stempel, *What Is the Meaning of “Plain Meaning”?*, 56 TORT, TRIAL & INS. PRAC. L.J. 551, 552 (2021).

¹⁹ Lawrence Solan et al., *False Consensus Bias in Contract Interpretation*, 108 COLUM. L. REV. 1268, 1268–69 (2008) (explaining how judges and other respondents reading texts routinely vastly overrate the degree to which others agree with them as to meaning of texts).

is reminded of Supreme Court Justice Potter Stewart's famous quip about obscenity ("I know it when I see it" and "the motion picture involved in this case is not that") (referring to the French film *Les Amants*).²⁰

But a textualist approach dominates, notwithstanding that observers other than the deciding judge may view the term as ambiguous, a definition that is not particularly hard to satisfy in that it requires only two or more reasonable constructions of text (although the proponent of a particular meaning must give at least some explanation why a proffered meaning is reasonable and not foreclosed).

In their quests to (depending on the posture of the case) argue for word clarity or word uncertainty, advocates often turn to an unrecognized type of de facto artificial intelligence: the dictionary.²¹ While dictionary use is of course not strictly an algorithm or program, it is a means of seeking machine-like objectivity—or the illusion of objectivity—hence our view that it has aspects of AI, albeit a primitive and potentially misleading form of AI. In addition to its potential to leapfrog the type of reflective, multi-faceted analysis one would expect from courts,²² dictionary use is problematic in that, notwithstanding the justifications advanced by its supporters, it is inconsistent and prone to manipulation as judges use preferred or readily available dictionaries and definitions without any particular discipline and perhaps with motivated reasoning deployed in the service of preferred outcomes.²³

We remain puzzled by the popularity of textualism and its emphasis on reading words and applying canons of construction without consideration of other materials permitting a more comprehensive examination of textual meaning. In our view, courts should, at least when presented with information by counsel, consider (if apt in the instant case) background; objective; history; precedent; market conduct; drafting history; purpose; contemporaneous understanding; reasonable expectations; and public policy (including avoidance of absurd or unconscionable results).²⁴

II. CORPUS LINGUISTICS AND ITS POTENTIAL FOR IMPROVING TEXTUALISM

An eclectic, comprehensive and integrative approach is an anathema to textualists. However, to the extent textualists consider canons of interpretation and dictionaries, they should be willing to similarly consider use of corpus

²⁰ See *Jacobellis v. Ohio*, 378 U.S. 184, 197 (1964) (Stewart, J., concurring).

²¹ See Stempel & Knutsen, *Textual Literalism*, *supra* note 2; Stempel & Knutsen, *Rejecting Word Worship*, *supra* note 2.

²² See generally Mark Lemley, *Chief Justice Webster*, 106 IOWA L. REV. 299 (2020).

²³ Scholarship consistently takes courts (particularly the U.S. Supreme Court) to task for dictionary shopping. Ellen P. Aprill, *The Law of the Word: Dictionary Shopping in the Supreme Court*, 30 ARIZ. L.J. 275 (1998); ESKRIDGE ET AL., *supra* note 12, at 595 (explaining that different Justices prefer different dictionaries); 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 (2013).

²⁴ See generally Stempel & Knutsen, *Rejecting Word Worship*, *supra* note 2.

linguistics analysis. The corpus linguistics movement or school of interpretation advocates use of a large database reflecting language use and analysis of that database to aid understanding and interpretation of words.²⁵ Corpus linguistics seeks to examine a massive number of writings to see how terms were used and with what frequency.²⁶

A leading corpus linguist, Douglas Biber, has identified four unifying characteristics of the corpus approach:

- it is empirical, analyzing the actual pattern of use in natural text;
- it utilizes a large and principled collection of natural texts, known as a ‘corpus,’ as the basis for analysis;
- it makes extensive use of computers for analysis, using both automatic and interactive techniques; and
- it depends on both quantitative and qualitative analytical techniques.²⁷

A recent article described corpus linguistics as a

method [that] allows a user to search a large body of text, or corpus, for a particular word to identify patterns in usage that reveal information about a word’s meaning. For example, a user may track a word’s frequency over time, identify the words that most frequently occur in close vicinity to that search term, or review each instance of a word’s usage in context. Advocates argue that this information tells us something important about the meaning of a term. Corpus linguistics, they argue, reveals the common meaning of a word more reliably than dictionaries, ad hoc Google searches, or intuition.²⁸

As noted by two prominent linguistics and law scholars, the “principal goal [of corpus linguistics] is to use big data that is representative of a particular variety of language as a source of information about ordinary meaning in the realm of statutes and original public meaning in constitutional argument.”²⁹

Another commentator describes corpus linguistics as

²⁵ See Stephen C. Mouritsen, Comment, *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)).

²⁶ *Id.*

²⁷ *Id.*

²⁸ Matthew Jennejohn et al., *Hidden Bias in Empirical Textualism*, 109 GEO. L.J. 767, 769–70 (2021) (citing Stefan Th. Gries & Brian G. Slocum, *Ordinary Meaning and Corpus Linguistics*, 2017 BYU L. REV. 1417, 1441 (2017); Thomas R. Lee & Stephen C. Mouritsen, *Judging Ordinary Meaning*, 127 YALE L.J. 788, 829 (2018); Brian G. Slocum, *Ordinary Meaning and Empiricism*, 40 STATUTE L. REV. (2018); Neal Goldfarb, *A Lawyer’s Introduction to Meaning in the Framework of Corpus Linguistics*, 2017 BYU L. REV. 1359, 1367–68 (2017); Thomas R. Lee & James C. Phillips, *Data-Driven Originalism*, 167 U. PA. L. REV. 261, 289–90 (2019); Friedemann Vogel et al., *Computer-Assisted Legal Linguistics: Corpus Analysis as a New Tool for Legal Studies*, 43 L. & SOC. INQUIRY 1340, 1346 (2018); Mouritsen, *supra* note 27); see also Lee & Phillips, *supra*, at 283 (explaining that words draw meaning from other words surrounding them, whereas dictionaries provide the meanings of words in isolation).

²⁹ Tammy Gales & Lawrence M. Solan, *Revisiting A Classic Problem in Statutory Interpretation: Is a Minister a Laborer?*, 36 GA. STATE U. L. REV. 491, 496 (2020).

[t]he study of language in corpora [the plural of “corpus”]. Diverse linguistic research projects employ corpus linguistic methods: from analysis of language acquisition to prediction of what syntactic choices speakers will make, to study of words’ positive or negative prosodies. “Legal corpus linguistics” (LCL) has been used to describe one very small subset of recent work in corpus linguistics and legal interpretation, concerning the *ordinary*, *public*, or *plain* meanings of legal texts. The legal community’s critiques aimed at LCL are not critiques of the broader field of “corpus linguistics.” At the same time, much has been written about corpus linguistics, and some of those discussions inform LCL debates.³⁰

Although relatively recent in legal discourse, corpus linguistics has been part of linguistics for roughly thirty years—a product of increasing data storage and processing capacity.³¹ It established a legal beachhead in the United States in 2010 with the publication of a prominent article,³² along with continuing work by the author and a prominent judicial ally, Utah Supreme Court Justice and former Brigham Young University law professor Thomas Lee.³³ Corpus linguistics analysis has substantial academic support and is increasingly being accepted as a judicial tool of analysis.³⁴

³⁰ Kevin Tobia, *The Corpus and the Courts*, 87 U. CHI. L. REV. ONLINE *1, *3–4 (2021) (footnotes omitted).

³¹ Prior to the advent of relatively low-cost servers and personal computers as well as the internet, world wide web, and software that could “scrub” the web to collect instances of word usage by assembling writings from books, magazines, newspapers, and other sources, it was prohibitively expensive, if not impossible, for linguists to engage in the analysis permitted by modern corpus linguistics.

³² See Mouritsen, *supra* note 27.

³³ See Lee & Mouritsen, *supra* note 28 (summarizing corpus linguistics and applying its methodology to cases).

³⁴ For example, the Seventh Annual Conference on Corpus Linguistics was held at BYU Law School in February 2021, reflecting the degree to which legal corpus linguistics has become an institution of sorts. For examples of corpus linguistics scholarship, see Clark D. Cunningham, *Foreword: Lawyers and Linguists Collaborate in Using Corpus Linguistics to Produce New Insights into Original Meaning*, 36 GA. STATE U. L. REV. vi (2020) (describing corpus linguistics and summarizing articles in symposium issue), and Lawrence Solan & Tammy Gales, *Corpus Linguistics as a Tool in Legal Interpretation*, 2017 BYU L. REV. 1311 (2017) (summarizing and critiquing corpus linguistics). Gales and Solan also took on a famous 1892 case text (*United States v. Holy Trinity Church*) based on the “spirit” of the law seemingly contra its literal text and found it correctly decided according to a richer textual analysis facilitated by corpus linguistics. Gales & Solan, *supra* note 29. Although focused on text, the findings of corpus linguistics analysis may easily align with longstanding purposive analysis. For example, on the basis of corpus linguistics analysis, Gales and Solan reached an interpretation of an immigration law decision consistent with purposivist support. *Id.* See also, e.g., Carol Chomsky, *Unlocking the Mysteries of Holy Trinity: Spirit, 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).

Despite not being widely used in court decisions to date, there are a number of instances in which courts have expressly applied corpus linguistics or related empirical study of word usage as part of their decision-making. See, e.g., *United States v. Costello*, 666 F.3d 1040 (7th Cir. 2012) (using Google News to determine that a woman living with a criminal boyfriend is not “harboring” him); *Am. Bankers Ass’n v. Nat’l Credit Union Admin.*, 306 F.

Of course, in law, where the “mainstream” is wide enough to include both Ronald Dworkin³⁵ and Robert Bork,³⁶ there is criticism of corpus linguistics in which commentators note its methodological limits or arguably myopic focus on text at the expense of the broader context and objective of the contract or statute under scrutiny.³⁷ Those preferring purposivist theories of interpretation (a group with which we identify) or a more holistic approach may approach corpus linguistics with caution but caution need not be rejection. We posit that increased use of corpus linguistics analysis will probably result in improved textualism, which in turn deflates efforts to supplant textualism with a more comprehensive approach that gives sufficient weight to non-textual indicia of meaning.

In our view, better textualism remains an inferior mode of analysis when compared to more comprehensive examinations of statutory meaning. Even if improved by corpus linguistics, textualism should be supplanted by a richer, less reductionist, and less fundamentalist approach to construing laws, contracts, and other documents.

But if one is to be a textualist, better to be a better textualist. Although a role for dictionaries (they can be helpful if not deemed inflexibly

Supp. 3d 44 (D.D.C. 2019) (using Corpus of Historical American English and data from opinions on Westlaw to determine the meaning of “rural district”); *State v. Rasabout*, 356 P.3d 1258, 1271 (Utah 2015) (Lee, J., concurring) (using corpus linguistics analysis to conclude that firing a gun twelve times constituted twelve “discharges” of the gun within the meaning of a criminal statute); *see also* Tobia, *supra* note 30, at *21–31 (collecting cases).

For additional explanation and analysis of corpus linguistics, see Jeffrey W. Stempel, *Adding Context and Constraint to Corpus Linguistics*, 86 BROOK. L. REV. 389 (2021); 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 the use of terms contained in a wide range of writings and accepting that corpus linguistics analysis is not the sole determinant of word meaning); Solan & Gales, *supra* (summarizing and critiquing corpus linguistics). Several databases have been assembled containing millions or even billions of entries from newspapers, books, periodicals, broadcast transcripts, and the like. *See, 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>]. *See generally* Symposium, *Data-Driven Approaches to Legal Interpretation*, 86 BROOK. L. REV. 291 (2021).

³⁵ *See, e.g.,* RONALD DWORKIN, *TAKING RIGHTS SERIOUSLY* (1977) (arguing for comprehensively left/liberal approach to law; criticizing courts and legislatures for excessive conservatism during mid-twentieth century).

³⁶ *See, e.g.,* ROBERT H. BORK, *SLOUCHING TOWARD GOMORRAH* (1996) (arguing for comprehensively right/conservative approach to law; criticizing perceived excessive turn left in law during mid-twentieth century).

³⁷ *See, e.g.,* Tobia, *supra* note 30; Jennejohn et al., *supra* note 28; Anya Bernstein, *Legal Corpus Linguistics and the Half-Empirical Attitude*, 106 CORNELL L. REV. 1397 (2021); Anya Bernstein, *What Counts as Data?*, 86 BROOK. L. REV. 435 (2021); Stempel, *supra* note 34; Evan C. Zoldan, *Corpus Linguistics and the Dream of Objectivity*, 50 SETON HALL L. REV. 401 (2019); Carissa Byrne Hessick, *Corpus Linguistics and the Criminal Law*, 2017 BYU L. REV. 1503 (2017). For a response to criticism of corpus linguistics, see Lee & Mouritsen, *supra* note 34.

determinative) and canons of construction (also helpful if not regarded as talismans or shibboleths) undoubtedly remains, the greater universe of a corpus of word usage or corpora (the term for multiple databases of word meaning) expands the utility of a dictionary and can help illuminate the meaning of a word based on its use in a large variety of contexts.

There are of course, some practical considerations, in particular the logistics and finances required for such efforts. Well-done corpus linguistics work requires expertise (possessed by relatively few lawyers and fewer judges); money; and time. But actual cases move quickly, almost always faster than the writing schedule of legal scholars. In addition, the longer timeline of a civil trial has been largely displaced in the United States by the increased emphasis on disposing of cases via motion that often focuses on the facial meaning of the text at issue.³⁸

To be deployed in court, a corpus linguistics analysis must be submitted at least by the time of a summary judgment motion³⁹ and probably in time for a Rule 12(b)(6) motion.⁴⁰ This is ordinarily too fast a track for extensive corpus linguistics input from counsel, unless the court is willing to delay consideration pending expected corpus linguistics analysis. This limits the ability of litigants to present corpus linguistics information helpful to courts faced with interpretative disputes.

But even if limited in utility in litigation, corpus linguistics analysis remains readily available to legal scholars assessing the meaning of statutes, regulations, and standardized forms (like insurance policies). To the extent that such scholarship takes place, it can be presented to courts.⁴¹

³⁸ See BROOKE D. COLEMAN ET AL., *LEARNING CIVIL PROCEDURE* chs. 6, 7 & 15 (3d ed. 2018) (noting strengthened summary judgment doctrine beginning in 1980s that permits judges to more easily decide cases as a matter of law based on text of documents at issue and strengthened Rule 12(b)(6) dismissal doctrine permitting judges to end cases viewed as insufficiently plausible); Stephen N. Subrin & Thomas O. Main, *The Fourth Era of American Civil Procedure*, 162 U. PA. L. REV. 1839 (2014) (noting increased judicial resolution of cases via motion in lieu of trials, a process that often involves judges determining meaning as a matter of law based on text of statutes and contracts); see, e.g., *Ashcroft v. Iqbal*, 556 U.S. 662 (2009) (Rule 12); *Bell Atl. Corp. v. Twombly*, 550 U.S. 544 (2007) (Rule 12); *Celotex Corp. v. Catrett*, 477 U.S. 317 (1986) (summary judgment); *Anderson v. Liberty Lobby, Inc.* 477 U.S. 242 (1986) (summary judgment); *Matsushita Elec. Indus. Co., Ltd v. Zenith Radio Corp.*, 475 U.S. 574 (1986) (summary judgment). In addition to contracts, statutes, and regulations, civil pleadings, and documents at issue in civil litigation are texts subject to interpretation that increasingly has focused on judicial interpretation rather than factual adjudication.

³⁹ FED. R. CIV. P. 56.

⁴⁰ *Id.* 12(b)(6).

⁴¹ To date, the bulk of corpus linguistics scholarship has concentrated on past decisions rather than pending cases. For example, the superb Gales & Solan, *supra* note 29, article assessed a famous U.S. case that is roughly 130 years old—*United States v. Church of the Holy Trinity*, 143 U.S. 457 (1892). *But see* Clark Cunningham & Jesse Egbert, *Using Empirical Data to Investigate the Original Meaning of “Emolument” in the Constitution*, 36 GA. STATE U. L. REV. 465 (2020) (addressing issue relevant to modern controversy over whether former President Donald Trump’s direction of government patronage to his hotel properties violated the Emoluments Clause of the U.S. Constitution).

III. A PRACTICAL TEMPLATE FOR EMPLOYING CORPUS-BASED INFORMATION FOR INTERPRETATION

While we would prefer that textualism give way to broader, more nuanced interpretative methods, we recognize that courts addressing interpretative issues will of course begin with the text of the document in question. If the court views its task as determining if the language at issue has a meaning sufficiently clear to foreclose consideration of non-textual information (e.g., background, purpose, drafting history, expressions of party intent or understanding, reliance, custom, practice, course of performance, course of dealing, or usage in trade), the court logically should enlist all reasonable textual aids in the task.

This should involve harnessing corpus linguistics analysis as an aid to textual interpretation. So used, corpus linguistics can be a helpful form of AI that collects and presents information about word usage and understanding that can assist a court in determining the clarity (or lack of clarity) of the text under review.

As a practical matter, extensive corpus-based research and analysis may not make economic sense in small stakes cases. But a brief examination of word usage in documents in addition to dictionaries can potentially provide insight at low cost. More extensive or sophisticated corpus linguistics analysis probably demands retention of a linguist as an expert witness, which may not be justified on cost–benefit grounds.

For the most part, we regard this as a decision to be made by counsel in consultation with clients. If an advocate thinks the stakes of the case justify substantial expenditure of resources on corpus linguistics analysis, it will be conducted and presented to the judge. If not, judges should perhaps be wary of *sua sponte* research of this type, just as judges should be wary of conducting any research or investigation out of view of the parties.

The dividing line should be whether the judge’s research is something that logically would have been done by a competent advocate with sufficient time and opportunity for producing the information. For example, no one should think it impermissible for a judge to conduct her own research into caselaw out of fear she will find and be influenced by a case not cited by either party. The additional case(s) considered by the judge were available to be found and discussed by the parties in their briefs. That said, if the court concludes that a particular case not cited by either party is determinative of a dispositive issue, the court should probably inform the parties and permit them to address the case and its implications.

In addition to the different timetables of litigation and legal scholarship, the recent and current zeitgeist in the United States has been unduly dismissive of legal scholarship as relevant to real world cases, exemplified by statements of Chief Justice John Roberts that he has not found such research and analysis helpful to judicial decision-making. See Brent E. Newton, *Law Review Scholarship in the Eyes of Twenty-First-Century Supreme Court Justices: An Empirical Analysis*, 4 DREXEL L. REV. 399, 399 n. 1 (2012) (noting Roberts’ statement that the typical law review article “isn’t of much help to the bar” and similar statement by Justice Stephen Breyer).

A similar approach seems apt for any judicial corpus linguistics examinations performed after briefing and oral argument. If the court's review of corpus-produced information is in the nature of gathering basic information outside the briefs (e.g., the time of sunrise on the day of the workplace accident) or noting additional cases concurring with those cited by one of the litigants, the court's use of this information as factors in assessing text probably does not require notice to the parties and corresponding opportunity for comment. However, if the court is conducting extensive examination of corpora or concluding that a particular finding of collated terms is determinative of the dispute, this should be brought to the attention of the parties if the information was not addressed by counsel in briefs or oral argument.

Although dictionaries, treatises, and judicial precedents are technically extrinsic to the documentary text under review, they are typically not considered extrinsic evidence and can be consulted by courts at the outset of the inquiry.⁴² Corpus linguistics analysis by advocates or even by the court *sua sponte* should enjoy similar status so long as the court is merely considering the corpus-provided information and not using it as the sole determinant of meaning.

As previously noted, corpus linguistics methodology looks at the way words are used across a range of verbal contexts to provide an assessment of the more common use and connotations of words.⁴³ It can thus serve as an empirical "Big Data" way of not only adding perspective to word usage but also for operationalizing canons of construction such as *noscitur a sociis* and *ejusdem generis*.

For example, the *noscitur* canon, which translates as "it is known [by] its associates," posits that the meaning of a word is illuminated by reference to the other words with which it is associated.⁴⁴ The *ejusdem* canon, which translates as "of the same kind, class, or nature," operates in similar fashion by positing that general words should be construed in accord with more specific words used in the text.⁴⁵

By using corpora, courts can obtain more information about the verbal "associates" of a word at issue in litigation and the "kind" or "class" of words to which the disputed term belongs. Although there is of course potential for error, unduly narrow focus, and motivated reasoning when a judge looks at

⁴² See RLLI, *supra* note 1, § 3 cmt. b. (AM. L. INST. 2018) ("Generally accepted external sources of meaning that courts consult when determining the plain meaning of an insurance policy term include: dictionaries, court decisions, statutes and regulations, and secondary legal authority such as treatises and law-review articles. Such external sources of meaning are not 'extrinsic evidence,' except in limited circumstances discussed below. Rather, they are legal authorities that courts consult when determining the plain meaning of an insurance policy term, which is a legal question.").

⁴³ See *supra* notes 27–33 and accompanying text.

⁴⁴ See ESKRIDGE ET AL., *supra* note 12, at 595, 1151.

⁴⁵ *Id.* at 596, 1152.

word collates generated by a corpus search,⁴⁶ these same dangers apply when the judge is merely reading the text in isolation (or, frankly, even when using a traditional printed dictionary). At least with a corpus linguistics approach, the judge is armed with empirical data that can operate to inform the judge that the court's initial impression of word association was off base.

The additional data and its empirical processing provided by corpus linguistics can similarly be used to test the validity of canons such as the surplusage canon, the belt-and-suspenders canon, and various grammar canons. For example, one might examine corpora to see if ordinary speakers and writers really follow high church grammar principles such as the last antecedent rule or are steadfastly consistent regarding the absence or deployment of the Oxford Comma.

To be sure, it is unrealistic to expect extensive, well-done corpus linguistics research for all but the highest stakes cases with the most sophisticated counsel or presiding judges on the high end of the intellectual scale.⁴⁷ And there are in our view real limits on the extent to which one should hitch notions of word meaning to corpus linguistics alone.⁴⁸ And, as with all forms of AI and AI-like research and technology, we should be concerned that use of a machine, algorithm, or big data does not devolve to judicial outsourcing of serious reflection and judgment by the courts in a careful, contextual fashion.

Notwithstanding these concerns, proper use of corpus linguistics analysis has to be a valuable addition to consultation of dictionaries, treatises, or even caselaw. Warts and all, a corpus linguistics approach provides hope for improvements in textualism. It may not displace judicial application of “gut feeling” or dictionary “shopping” or result-oriented twisting of language, but it can at least serve as a check upon the errors these methods often produce.

Although corpus linguistics involves computer searches, so does legal research. And, unlike some forms of AI, corpus linguistics does not purport to provide an answer but instead provides data for review and analysis by a human being.

⁴⁶ As one federal appeals court judge (Harold Leventhal of the District of Columbia Circuit) put it regarding this problem of motivated reasoning, construing legislative history is like “looking over a crowd and picking out your friends.” Patricia M. Wald, *Some Observations on the Use of Legislative History in the 1981 Supreme Court Term*, 68 IOWA L. REV. 195, 214 (1983). Overlooked in this aphorism, however, is the degree to which this tendency applies to reading statutory language of any length or selecting the most apt dictionary definition or legal precedent. See Tobia, *supra* note 30 (noting this criticism of textualists by purposivists); Adam M. Samaha, *Looking over a Crowd—Do More Interpretative Sources Mean More Discretion?*, 92 N.Y.U. L. REV. 554 (2017) (concluding that the answer is “no”); see also sources cited *supra* note 37.

⁴⁷ Where there are interested litigants or counsel (e.g., the legal aid attorney on a mission) wishing to establish a point (e.g., a test case supported by an interest group), one may reasonably expect to see corpus linguistics in lower stakes cases. Similarly, a particularly motivated judge with strong law clerks or access to academic research may introduce corpus linguistics analysis in lower stakes cases. See, e.g., Stempel, *supra* note 34, at 394.

⁴⁸ For a critical analysis of the field, see sources cited *supra* note 37.

Although corpus linguistics lacks the organization of a dictionary, which is the product of authors, the uncurated nature of corpora can be a useful adjunct or counterweight to dictionary consultation—a means of checking to see if the dictionary authors have erred in their assumptions about common use of language.

Corpus linguistics can also operate as a counterweight to erroneous or problematic precedent. Prior case law is the seeming gold standard for informed interpretation. Under the common law system, it makes sense for subsequent courts to generally follow the lead of prior decisions that have delved into an issue rather than expending judicial resources on well-trodden ground, at least for established rules or conclusions. Once a given jurisdiction has conclusively and consistently determined what constitutes “use” of an automobile or the “trigger” of “bodily injury” or “property damage,” it generally makes little sense to re-examine the correctness of these decisions or the rigor of their reasoning through corpus linguistics or other tools.

But where a decision stands alone, is recent, is enmeshed with arguably inconsistent precedent, or displays patently problematic reasoning, it should be fair game for reassessment, including examination according to corpus linguistics.

For example, a significant number of courts, perhaps even a majority, have stated that faulty work by a vendor cannot be an “accident” eligible for insurance coverage (per the general rule that insurance applies only to fortuitous losses and not to intentional misconduct or property destruction).⁴⁹ We regard most of these decisions as illogical and poorly reasoned,⁵⁰ even

⁴⁹ See generally RANDY MANILOFF ET AL., 2 GENERAL LIABILITY INSURANCE COVERAGE: KEY ISSUES IN EVERY STATE ch. 12 (5th ed. 2021) (*Is Faulty Workmanship an “Occurrence”?*).

⁵⁰ Our view, admittedly not verified by corpus linguistics research (a great potential project for a future article or student law journal note), is that vendors frequently make unintentional errors—accidents—when doing work. Think of the carpenter who is not intending to provide standard work to the customer but is simply a bad carpenter. His faulty work absolutely is accidental. He’s a bad carpenter. Similarly, failing to properly align sheetrock or applying wallpaper with air bubbles reflects unintended miscues by the worker. That is different than a vendor consciously choosing not to use rebar when pouring concrete in order to cut costs.

Episodes like the latter example of the evil contractor are not covered by insurance because the resulting harm is expected or intended by the insured contractor or may even be deemed intentional fraud. Episodes like the former are not covered absent injury to other property because they are the “work” of the builder that is excluded from general liability coverage by the “Your Work” exclusion contained in the standard form commercial general liability (“CGL”) policy.

Consequently, even if a court is concerned about preventing a liability insurance policy from becoming a surety performance bond, there is nonetheless no need for a court to deny the accidentalness of vendor error to avoid coverage for the mere repair or replacement of vendor work. The “Your Work” exclusion prevents this and vindicates the purpose of the general liability policy and the intent of insurers.

However, as noted in the subsequent footnote, the implications of our analysis may not differ from that of most courts (which refuse to treat faulty work as an accident unless it damages property other than that of the vendor, in which case it then magically becomes a non-accident and is subject to coverage). See generally MANILOFF ET AL., *supra* note 49.

though most of these courts redeem themselves by finding sufficient fortuity when the purportedly non-accidental work of the vendor causes injury to other property.⁵¹

Precedent like this should be subject to re-examination if it is not so venerable that it has created undue reliance. Even if the ultimate result in these cases (coverage where the leaky roof from faulty work damages the wood flooring of the building) is correct, the notion that negligent⁵² work cannot be an accident is illogical on its face and contrary to real world experience. Correcting this analytic error, although hardly the top priority of legal reform, would improve legal analysis and, more important, remove the risk that this reasoning could result in problematic decisions in other contexts. Corpus linguistics analysis of the use of the word “accident” and its synonyms could well shed light on this and illustrate whether we are right or wrong in our view (based on a collective 100 years on the planet) that in common speech, vendor error is often described as accidental or by chance rather than intended.⁵³

Corpus linguistics, if done with adequate controls, can also be helpful in providing algorithm-like discipline as well as expanding contextual information. A persistent criticism of judicial use of dictionaries is that courts are inconsistent, using different dictionaries from case to case, which permits more opportunity for opportunism in that a judge can keep dictionary shopping until she finds one with a definition most amenable to the desired result.⁵⁴ This

⁵¹ *Id.*

⁵² Negligence is by definition accidental even resulting from foolishness, laziness, or stupidity. For example, a painter working in a vacant rental home in Hilo, Hawaii, (where the average annual rainfall is 142 inches, *Average Weather for Hilo (By Month)*, LOVE BIG ISLAND, <https://www.lovebigisland.com/hilo/average-yearly-weather/> [https://perma.cc/ZM5U-N8ZE] (Mar. 16, 2021)) may forget at the end of the job to close the windows previously opened to allow ventilation. The serious water intrusion damage that greets the landlord inspecting the property a week later is not an intentional loss but results from the painter’s negligence. In situations like this, we think the conduct and result sufficiently “accidental” to fall within coverage. By contrast, the painter’s conscious decision to use a cheaper paint known to be prone to peeling is both intentional conduct making the problem substantially certain to result and, unless the flaking paint is ingested or otherwise causes damage, falls within typical policy provisions barring coverage where the injury was only to the vendor’s “own work.” We thus accept a finding of “no accident” from faulty work in cases where the evidence shows that a policyholder intentionally engaged in conduct that was intended to or practically certain to cause loss but not in cases where the vendor simply was not very good at its trade or was insufficiently vigilant in catching errors made by employees or agents. This is consistent with the typical clauses in liability policies allowing coverage to a general contractor where the deficient work was performed by a subcontractor.

⁵³ For example, we have heard customers say (and perhaps said ourselves) that a vendor “screwed up” a repair much more often than we have heard a customer describe a poor repair or installation as an intended deficiency.

⁵⁴ Another, perhaps more realistic possibility, is that result-oriented judges with motivated reasoning are not consciously dictionary shopping but instead are making decisions based on policy preference or (more innocently) based on their initial reading of text. When viewing the briefs of counsel, they are then more naturally attracted to dictionary definitions contained in the brief of the favored party and cite those in the resulting opinion as ballast for a decision already made on first glance at the words in dispute.

gives the judge the opportunity to plausibly argue that the particular result is compelled by the “plain meaning” of the term without resort to any non-textual information that might contradict the word meaning preferred by the result-oriented judge.⁵⁵

In contrast, a corpus linguistics analysis involves a large data collection of word collates with no single datum representing the word’s meaning. Rather, word meaning is gleaned from its use across the data. Although of course subject to pre-conceived notions (perhaps based on initial consultation of a dictionary as well as pre-existing views of word meaning) or manipulation, the broader nature of a corpus (relative to a dictionary) most likely reduces the confirmation bias courts can display when using dictionaries alone.

There are, of course, different corpora that may be used to assess word meaning.⁵⁶ The number and size of corpora (including specialized or technical corpora) is impressive and growing.⁵⁷ There is consequently the possibility of “corpus shopping” similar to dictionary shopping. But as noted, even a specialized corpus contains many more examples of word usage than a dictionary and most likely many more than any one court’s collective experience with usage of a particular term.⁵⁸

Thus, notwithstanding limitations and risk, corpus linguistics analysis, unless really poorly done, should at least be a useful adjunct in determining

⁵⁵ To achieve this type of result-oriented decision, of course, a judge often needs only a single dictionary in that most dictionaries contain several definitions for each word, providing some discretion in selection without the need to consult additional dictionaries. But that said, the opportunity for finding a sought definition is of course increased by review of additional dictionaries.

⁵⁶ Particularly popular are the *Corpus of Historical American English*, ENGLISH-CORPORA, <https://www.english-corpora.org/coha/> [<https://perma.cc/2CAU-U9N5>] and the *Corpus of Contemporary American English*, ENGLISH-CORPORA, <https://www.english-corpora.org/coca/> [<https://perma.cc/99B2-4X4Z>]. There are also specialized corpora dealing with statutory language, scientific language, engineering terminology, and so on. *E.g.*, Khurshid Ahmad, Prof., Univ. of Surrey, Language Engineering and the Processing of Specialized Terminology, Address at the Language Engineering Convention/ACTES du Journees du Genie Linguistique (Jul. 6–7, 1994) (available at <http://www.computing.surrey.ac.uk/ai/pointer/paris.html>).

⁵⁷ In addition to the widely known Corpus of Contemporary English and the Corpus of Historical English, *supra* note 56, there are now corpora composed of the United States Code and other statutes as well as corpora regarding building trades terms, medical terms, and word usage in other specialized fields. Many of these are specifically constructed by linguistics scholars for their own research and may not be available to the public but the norm appears to be that researchers make their corpora available to colleagues in the field so that their work may be examined and expanded upon. *See Corpus Linguistics*, UNIV. OF ESSEX: W-3 CORPORA PROJECT, https://www1.essex.ac.uk/linguistics/external/clmt/w3c/corpus_ling/content/choosing.html [<https://perma.cc/9R6G-D4VV>].

⁵⁸ A dictionary typically contains a range of word illustrations that seldom exceeds a half-dozen examples. By contrast, a corpus contains thousands of examples of word usage. Although one might question whether all examples are helpful in the legal context (e.g., the use of a word in a teenage pop culture magazine as compared to the more curated examples of the dictionary), even a small corpus will contain many more examples than even a comprehensive dictionary. *See id.*

word meaning even if not deemed determinative. Hence, it looks to us like an improvement over the common textualist methods of simply staring at the words or deferring to a dictionary. In properly deploying corpus linguistics analysis, one commentator's checklist provides useful guidance:

- Analyze text from the relevant time;
- Use representative and balanced corpora;
- Don't commit the Nonappearance or Uncommon Use Fallacies;
- Don't commit the Comparative Use Fallacy;
- Take account of the "context" of the language;
- Acknowledge that corpus data might [in some or even many cases] ultimately be unhelpful;
- Acknowledge the possibilities of linguistic indeterminacy; [and]
- Do not rely on "intuition about linguistic facts," which may be biased.⁵⁹

This is a challenging checklist that will be honored in the breach more than the observance in many cases. We emphasize that we are not suggesting that corpus linguistics is a panacea for improving judicial interpretation or that it should displace other means of assessing disputed text. In addition to considering greater receptiveness to non-textual information (our preferred solution), courts could improve the operation of textualist technique by agreeing on a presumptively limited universe of "major" dictionaries and resorting to other dictionaries only when justified where the term at issue is technical or specialized and more aptly defined by a medical dictionary, an engineering dictionary, or (of course) an industry dictionary,⁶⁰ provided there is no evidence suggesting the parties were not using the term in its technical sense. To implement this, courts would agree on a presumptively limited universe of "major" dictionaries such as the Oxford English Dictionary or the latest Merriam-Webster collegiate dictionary.

Whatever ground rules a court adopts for construing laws such as a constitution, treaty, statute, or regulation, those same ground rules (e.g., dictionaries to consult, adjustment for the time the text came into being, resort to specialized lexicons) should ordinarily apply to contracts, particularly to insurance policies and standardized contracts that operate in the nature of private legislation.⁶¹ There is no compelling reason to approach words differently based on the type of document containing the disputed text,

⁵⁹ Tobia, *supra* note 30, at *9–20 (providing explanation of these comments and citing them in relation to corpus linguistics scholarship, particularly examining Lee & Mouritsen, *supra* note 34).

⁶⁰ See, e.g., HARVEY W. RUBIN, *DICTIONARY OF INSURANCE TERMS* (6th ed. 2013). In addition, most insurance textbooks used in colleges and business schools contain a glossary of terms as an appendix. See, e.g., GEORGE E. REJDA & MICHAEL J. MCNAMARA, *PRINCIPLES OF RISK MANAGEMENT AND INSURANCE* 659–77 (13th ed. 2017); EMMETT J. VAUGHN & THERESE M. VAUGHN, *FUNDAMENTALS OF RISK AND INSURANCE* app. A (11th ed. 2014) (pp. G-1 through G-21).

⁶¹ Jeffrey W. Stempel, *The Insurance Policy as Statute*, 41 *MCGEORGE L. REV.* 203 (2010).

particularly when so-called “private” documents are largely standardized and widely distributed.⁶²

Looking forward, we consider corpus linguistics analysis as a potentially illuminative method of textual inquiry but not a substitute for more comprehensive analysis by human interpreters who, even if confined to a textualist methodology, can use a number of resources and tools for determining meaning, rather than relying on initial, gut-feeling impressions. Adding corpus linguistics to the textualist mix is not as easy a “lift” for courts as more disciplined and even-handed use of dictionaries but neither need it be an unduly expensive, time-consuming, or difficult undertaking.

In addition, like other legal analysis, corpus linguistics analysis will often, perhaps nearly always, be presented to the court by advocates, whose submissions in turn are examined by opposing advocates. Judges receptive to corpus linguistics need not start from scratch or undertake a lengthy *sua sponte* journey, but can have the information brought to them through the adversary system.

Technological advances (e.g., bigger databases, faster computers, better search engines) and methodological advances (more comprehensive corpora, more informative specialized corpora, corpora containing more apt writings such as academic articles rather than magazine articles) (or vice versa depending on the case) should improve the usefulness of corpus linguistics analysis. Analytical and theoretical advances are likely to provide greater guidance to courts as to the limits of analysis and conclusions that can be drawn from word associations. The corpus linguistics tool should improve over time, as will court (and advocate) facility with corpus linguistics methods and technology.

CONCLUSION

Rigid (or even semi-rigid) textualism is not ideal. But it has a long history of use and enjoys current judicial and political support. If we must have textualism, particularly textualism resistant to considerations beyond the face of disputed text, any methodology that improves the process of textualist interpretation should be welcomed by the courts. Judicious use of corpus linguistics offers courts the opportunity to apply the textualist methodology with more rigor so that decisions reduce reliance upon and selective use of dictionaries as well as idiosyncrasy and bias in textual construction.

⁶² *See id.* Even so-called “manuscript” policies drawn more specifically by the parties are seldom written de novo but usually are “drafted” by combining standard language from policies with often only modest revision.