CONTRACT INTERPRETATION WITH CORPUS LINGUISTICS

Stephen C. Mouritsen

Abstract: Courts and scholars disagree about the quantum of evidence that is necessary to determine the meaning of contractual provisions. Formalists favor excluding extrinsic evidence unless the contractual text is found to be ambiguous. Contextualists, by contrast, look to extrinsic evidence to support claims about contractual meaning even absent a finding of ambiguity. The formalist approach is faulted for failing to provide a meaningful account of the parties’ intentions and for placing heavy reliance upon the judge’s own linguistic intuitions and general-use English dictionaries—both problematic guides to plain meaning. At the same time, the contextualist approach may impose significant costs on the contracting parties and invite strategic behavior.

Corpus linguistics offers a middle way. Corpus linguistics draws on evidence of language use from large, coded, electronic collections of natural language—language used in natural settings, rather than language elicited through interviews or surveys. These may include collections of texts from newspapers, magazines, academic articles, or transcribed conversations. These collections of texts are referred to as corpora (the plural of corpus). Linguistic corpora can be designed to model the linguistic conventions of a wide variety of speech communities, industries, or linguistic registers. Because large, sophisticated linguistic corpora are freely available, language evidence from linguistic corpora offers a comparatively low-cost alternative to the vast quantity of extrinsic evidence permitted by contextualist interpretive approaches. Moreover, by evaluating corpus evidence, judges and lawyers can create a more accurate, evidence-based picture of contractual meaning than can be found in the formalist judge’s linguistic intuition or in a general-use dictionary.

Moreover, corpora can provide objective evidence of the linguistic conventions of the communities that draft and are governed by the agreements judges and lawyers are called upon to interpret. Corpus evidence can give content to otherwise vague legal concepts and provide linguistic evidence to aid in the evaluation of claims about the meaning (or ambiguity) of a contractual text. Below I outline how corpus linguistic methods may be applied to the interpretation of contracts.

INTRODUCTION........................................................................................................... 1339
I. CHALLENGES TO PLAIN MEANING IN CONTRACT INTERPRETATION................................. 1342
   A. Formalism, Contextualism, and the Middle Way............. 1342
   B. Theoretical Challenges to Plain Meaning..................... 1344

* Stephen C. Mouritsen is an adjunct professor of the Theory and Practice of Interpretation and Law and Corpus Linguistics at the J. Reuben Clark Law School at Brigham Young University. The author would like to thank Jill C. Anderson, Will Baude, Omri Ben-Shahar, Lisa Bernstein, Richard Epstein, David Hoffman, and Thomas R. Lee for comments on an earlier draft. Thanks as well to the Olin-Searle-Smith Fellows in Law program for making possible the author’s association with the University of Chicago Law School.
C. Operational Challenges to Finding Meaning and Ambiguity................................................................. 1346
   1. The Limits of Intuition about Plain Meaning and Ambiguity............................................................ 1346
   2. Dictionaries and the “Baffled Judge” ......................... 1348
II. WHAT WE MEAN WHEN WE TALK ABOUT MEANING AND CONTEXT .................................................. 1353
   A. Semantic, Syntactic, and Pragmatic Context.......... 1354
   B. Plain Meaning and Historical Context .................. 1358
   C. Meaning and Speech Community .......................... 1359
III. A CORPUS-BASED APPROACH TO PLAIN MEANING ................................................................. 1360
IV. THE CORPUS APPROACH TO PLAIN-MEANING IN PRACTICE ....................................................... 1365
   A. Problem No. 1: Plain Meaning and “Snorkeling” ..... 1365
      1. Dictionaries and the Plain Meaning of “Sport” ...... 1366
      2. Prototype, Family Resemblance, and the Plain Meaning of “Sport” .............................................. 1368
      3. Linguistic Corpora and the Meaning of “Sport” ...... 1371
      4. Corpus Linguistics and the Notion of a Core Meaning........................................................................... 1374
      5. Corpus Linguistics and “Sport” in Context .......... 1376
   B. Problem No. 2: The Plain Meaning of “Anticipated” ........ 1379
      1. Prior Case Law and the Plain Meaning of “Anticipated” ................................................................. 1380
      2. Plain Meaning, Inflection, and Part of Speech ....... 1382
      3. “Anticipated” in the Corpus ............................... 1383
   C. Problem No. 3: The Plain Meaning of “Governable” .... 1386
      1. When Definitions Are Not Found in the Dictionary... 1387
      2. The Mandatory Sense of “able” ......................... 1389
      3. “Governable” in the Corpus .............................. 1390
   D. Problem No. 4: The Plain Meaning of “Structural Damage” .......................................................... 1397
      1. “Structural Damage” in the Dictionary................. 1398
      2. “Structural Damage” and the “Pet Fish” Problem .. 1400
      3. Co-selection, Delexicalization, and the Canon against Surplusage ............................................. 1401
      4. “Structural Damage” in the Corpus ..................... 1404
V. CONTRIBUTIONS AND POTENTIAL CHALLENGES OF THE CORPUS APPROACH ........................................... 1408
   A. The Reference Timeframe for Contract Interpretation.... 1408
   B. Register and Speech Community in Contract
INTRODUCTION

When interpreting the undefined terms in a contract, judges often turn to the so-called Plain Meaning Rule, an interpretive heuristic that requires courts to consider only the text of the contract and exclude extrinsic evidence, provided that the text of the contract is unambiguous. In a familiar formulation, the Rule states that “if a ‘clear, unambiguous’ meaning is discernible in the language of the contract, no extrinsic evidence of surrounding circumstances may be admitted to challenge this interpretation.”

The Plain Meaning Rule is often justified on the grounds that it prevents strategic behavior and the fear that allowing contracts to be “challenged in the name of the parties’ actual intent” will “produce disorder or even chaos, waiting to be exploited by unscrupulous litigants who demand a bonus to do what they already promised to do.” In such cases, “the disagreement often manifests itself in the litigation positions of the parties, whose interpretations may more realistically reflect their lawyers’ clever post hoc arguments than any serious dispute about meaning.” The Rule is also justified on the grounds of efficiency.

Parties may prefer a strong Plain Meaning Rule believing that it lowers judicial search costs by limiting the number of sources a judge has to

1. 5 MARGARET N. KNIFFIN, CORBIN ON CONTRACTS § 24.7, at 33 (Joseph M. Perillo ed. 1998).
   This Article addresses the Plain Meaning Rule and focuses on questions of linguistic ambiguity in contractual texts. While the distinction between the Plain Meaning Rule and the Parol Evidence Rule isn’t always clear, here I adopt the framework suggested by Professor Eric Posner. Posner restates that the Parole Evidence Rule as follow: “A court will refuse to use evidence of the parties’ prior negotiations in order to interpret a written contract unless the writing is (1) incomplete, (2) ambiguous, or (3) the product of fraud, mistake, or a similar bargaining defect.”


consider to find plain meaning. A strong Plain Meaning Rule may also allow judges to resolve more cases at earlier stages of the litigation (for example, at the motion to dismiss or summary judgment stage) and thus “reduce the number of intractable credibility issues in trials for breach of contract.”

While intuitively appealing, the Plain Meaning Rule faces a number of challenges. For example, consider the case of an ecological tourism company that seeks to recover from its insurer for failing to defend a negligence action arising from a customer’s drowning death that occurred while snorkeling. The insurance contract at issue makes clear that it “does not apply to ‘bodily injury’ [including death] to any person while practicing for or participating in any sports or athletic contest or exhibition that you sponsor.” This raises the question: Is snorkeling a sport? To resolve this question the court invokes the Plain Meaning Rule and cites dictionary definitions that show that sport is defined as a rule-based athletic competition, and that snorkeling must therefore fall outside of the plain meaning of sport.

The court, like many courts before it, ignored a perfectly well-attested, alternative sense of the word it was purporting to define. Indeed, the same dictionaries cited by the court also define sport as a “recreational activity,” without reference to competition. The court not only failed to take into account an alternative sense, but had no way of knowing the comparative prevalence of the competing senses in the relevant contractual context and no basis for concluding that one should be preferred over the other. These problems faced in the snorkeling case are hardly idiosyncratic. They are similar to problems faced by other courts when called upon to interpret contractual language.

5. Id.
6. Id.
8. Id. at 1193.
9. Id. at 1197.
10. Throughout this Article, I use the words “attest” and “attestation” in their linguistic sense to mean “to establish or verify the usage of [a word],” see Attest, MERRIAM-WEBSTER, https://www.merriam-webster.com/dictionary/attest [https://perma.cc/QET8-7ZTB], and not in their legal sense to mean “[t]o bear witness; testify,” See Attest, BLACK’S LAW DICTIONARY (11th ed. 2019).
11. Throughout this Article, I use the word “sense” to refer to “[o]ne of several meanings that can be established for a word or phrase and covered by a definition in a reference work.” R.R.K. HARTMANN & GREGORY JAMES, DICTIONARY OF LEXICOGRAPHY 125 (2002) (emphasis omitted).
12. See infra notes 141–148.
Linguistic corpora can help judges and lawyers evaluate and work to resolve problems of finding meaning in contractual language. Corpus linguistics may chart a middle way between the formalist and contextualist approaches to contract interpretation, which permits the consideration of extrinsic evidence even absent a finding of ambiguity. As discussed below, corpus linguistics provides judges and lawyers with objective information about language use without some of the prohibitive costs and risks of strategic behavior associated with contextualist approaches to interpretation. Corpus linguistics may help give content to otherwise vague legal concepts like “plain meaning,” “ambiguity,” and “context.” And corpus linguistics can provide evidence of language usage that cannot be obtained through introspection or from dictionaries—evidence of the syntactic, semantic, and pragmatic contexts in which contractual terms are used, and evidence of the language usage from particular speech communities at a particular point in history. For example, while dictionaries typically only take account of word-level meaning, a corpus user may examine usage at a phrase or even sentence level. Linguistic corpora can be constructed to represent the linguistic conventions of a particular industry or region. They can be designed to reflect the linguistic conventions of different time periods. They can provide their users with evidence of language use that is not available in a dictionary and not available via introspection.

But corpora are not a panacea. They are not a black box that will provide definitive, objective answers to every interpretive question. While linguistic corpora can present objective evidence of language use, corpus design and the interpretation of corpus evidence requires human judgment.

Corpora can provide evidence of the way that language is used by the contracting parties—evidence that was not previously available via dictionaries or introspection. It is not possible in this Article to anticipate all of the ways in which corpus evidence may be brought to bear on questions of contract interpretation. In this respect, the problems discussed below are both illustrative, but also exploratory—they demonstrate the kinds of language evidence that corpora make available, but also highlight questions that require additional research.

Corpus linguistics has already begun to play a role in the interpretation of legal texts, both of statutes and of the Constitution.
And linguistic corpora have also played a role in discussions of trademark law.15 To date, no one has addressed the contribution linguistic corpora can make in providing an evidence-based means of evaluating questions of contractual plain meaning and ambiguity.

Part I outlines how corpus linguistic methods may be applied to questions of contract interpretation. Part II discusses both the theoretical and operational challenges of current approaches to plain meaning. Part III introduces corpus linguistics as an alternative to traditional approaches to plain meaning. Part IV addresses four problems of contract interpretation to illustrate the corpus-based approach. Part V discusses both contributions of—and challenges to—the application of corpus methods to questions of contract interpretation. Part VI concludes this Article.

I. CHALLENGES TO PLAIN MEANING IN CONTRACT INTERPRETATION

Courts face a number of challenges when they seek to determine the meaning of contractual terms. These challenges are addressed below.

A. Formalism, Contextualism, and the Middle Way

Courts and scholars disagree about the quantum of evidence that is necessary to determine the meaning of contractual provisions (even in the absence of a finding of ambiguity).16 Contextualists, following the Restatement of Contracts and the Uniform Commercial Code (UCC),


look to extrinsic evidence even before there is a determination that the language of the contract is ambiguous.¹⁷ This evidence may include “the situation and relations of the parties, the subject matter of the transaction, preliminary negotiations and statements made therein, usages of trade, and the course of dealing between the parties.”¹⁸ Critics fault the contextualist approach to finding meaning in a contract for imposing significant costs on the contracting parties and inviting strategic behavior that may ultimately deny contracting parties the benefit of their bargain.¹⁹ Moreover, it is not always clear that evidence of a consistent usage of trade, for example, actually exists, or that judges’ evaluation of that evidence is likely to be accurate or predictable.²⁰

Formalists, by contrast, justify the Plain Meaning Rule on the grounds of efficiency and predictability, and they exclude extrinsic evidence absent a finding of ambiguity. Critics fault the formalist approach for failing to provide a meaningful account of the parties’ intentions.²¹

¹⁷. See RESTATEMENT (SECOND) OF CONTRACTS § 212 cmt. b (AM. LAW INST. 1981) (noting that consideration of extrinsic evidence should “is not limited to cases where it is determined that the language used is ambiguous”).

¹⁸. Id.

¹⁹. See Trident Ctr. v. Conn. Gen. Life Ins. Co., 847 F.2d 564, 569 (9th Cir. 1988) (critiquing the contextualist approach because “even when the transaction is very sizeable, even if it involves only sophisticated parties, even if it was negotiated with the aid of counsel, even if it results in contract language that is devoid of ambiguity, costly and protracted litigation cannot be avoided if one party has a strong enough motive for challenging the contract.”); Gregory Klass, Contract Exposition and Formalism (Feb. 2017) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2913620 [https://perma.cc/TCE2-7R3G] (“The more evidence one allows into interpretation, the less certain the outcome. The costs of such uncertainty in the contractual setting can be especially high. A party that wants to organize its behavior in light of the legal effects of a contractual agreement needs to be able to predict how an adjudicator will later interpret that agreement. To the extent thicker interpretive rules reduce predictability, they impose an additional cost on the parties.”).

²⁰. Lisa Bernstein, The Questionable Empirical Basis of Article 2’s Incorporation Strategy: A Preliminary Study, 66 U. CHI. L. REV. 710, 715 (1999) (finding that “‘usages of trade’ and ‘commercial standards,’ as those terms are used by the Code, may not consistently exist, even in relatively close-knit merchant communities.”); see also Lisa Bernstein, Merchant Law in a Modern Economy, in PHILOSOPHICAL FOUNDATIONS OF CONTRACT LAW 238, 250 (Gregory Klass et al. eds., 2014) (“[C]ourt determinations relating to the existence, content, and scope of usages are likely to be both inaccurate and highly unpredictable, as they are typically made on the basis of very limited information.”).

²¹. See Klass, supra note 19, at 64 (“Although plain meaning rules reduce the costs of adjudication, increase predictability and can provide interpretive accuracy with respect to literal meaning, they create a higher risk of correspondence errors. Plain meaning rules often produce literalist interpretations, and literal meaning is sometimes a poor proxy for the parties’ actual agreement, and identification of the parties’ actual intentions and understandings is the ultimate goal of contract interpretation. Where there is likely to be a gap, a thicker use meaning rule, despite its extra costs, might be the better design choice.”). But see Charles J. Goetz & Robert E. Scott, The
Moreover, formalist decisionmaking about Plain Meaning often places heavy reliance upon the judge’s own linguistic intuitions and upon general-use English dictionaries, both of which are problematic guides to Plain Meaning.  

The use of language evidence from linguistic corpora may chart a middle way between the formalist and contextualist approaches to contract interpretation. The corpus can provide the court with objective evidence of language usage and give the court a more accurate picture of the linguistic conventions of the contracting parties than is currently available through traditional modes of formalist interpretation. And the use of language evidence from linguistic corpora, while certainly requiring the expenditure of some time and effort, may be less costly than assembling and presenting evidence of usage of trade, course of performance, and drafting and negotiating history that imposes cost burdens on parties in contextualist jurisdictions. Moreover, to the extent that the court or the parties looks to existing corpora for language evidence, it may be possible to address some of the concerns about strategic behavior for current contextualist approaches.

B. Theoretical Challenges to Plain Meaning

When courts speak of plain meaning, they do so without a coherent, shared, well-defined, objective notion of what plain meaning actually means. Similarly, in spite of its conceptual importance in the interpretation of contracts, courts lack a coherent, shared, well-defined, objective definition of what ambiguity actually means. Nor do they have a shared and coherent method for discovering the ostensibly plain language of a contract. While the Plain Meaning Rule is subject to a host

Limits of Expanded Choice: An Analysis of the Interactions Between Express and Implied Contract Terms, 73 CALIF. L. REV. 261, 311–12 (1985) (arguing that “[g]rigorous application of the plain-meaning rule reduces interpretation error by encouraging more careful choices of clear, predefined signals”).

22. See infra section II.B.

23. Indeed, even the justices of the United States Supreme Court appear unable to agree on the definition of plain meaning. See Clark D. Cunningham et al., Plain Meaning and Hard Cases, 103 YALE L.J. 1561, 1563–65 (1994); see also Richard J. Lazarus & Claudia M. Newman, City of Chicago v. Environmental Defense Fund: Searching for Plain Meaning in Unambiguous Ambiguity, 4 N.Y.U. ENVTL. L.J. 1, 15 (1995) (“The definition of ‘plain meaning’ is itself anything but plain. How much ambiguity is required before the meaning of a provision becomes ambiguous? Words are hardly ever entirely free of ambiguity and there is almost always room for disagreement based on at least plausible readings.”); Sandra F. Sperino, Flying Without a Statutory Basis: Why McDonnell-Douglas is Not Justified by Any Statutory Construction Methodology, 43 HOUS. L. REV. 743, 764–65 (2006) (“While the term ‘plain meaning’ exudes a sense of simplicity, such an assumption would be misplaced because the exact contours of plain meaning interpretation are debated.”).
of criticisms and has been rejected by the UCC\textsuperscript{24} and the Restatement (Second) of Contracts,\textsuperscript{25} it is nevertheless the approach taken by the majority of U.S. jurisdictions,\textsuperscript{26} and, thus, continues to play an important role in the interpretation of contracts.

Things don’t improve when courts turn to the identification of ambiguity. “The problem,” observed Professor Lawrence Solan, “perhaps ironically, is that the concept of ambiguity is itself perniciously ambiguous. People do not always use the term in the same way, and the differences often appear to go unnoticed.”\textsuperscript{27} As Professor Brian G. Slocum has observed, “the determination of ambiguity by the judiciary is entirely standardless and discretionary. The definitions of ambiguity used by courts are themselves vague, ambiguous, and unhelpful.”\textsuperscript{28} Most states define a contractual provision as ambiguous if “it is reasonably or fairly susceptible of different interpretations,” or simply “capable or susceptible of different interpretations.”\textsuperscript{29} Such definitions leave open a wide range of potential interpretations for contractual terms and provide little guidance to courts struggling to take their role as the ultimate arbiters of contractual clarity and ambiguity seriously.\textsuperscript{30}

\textsuperscript{24} Peter Linzer, The Comfort of Certainty: Plain Meaning and the Parol Evidence Rule, 71 FORDHAM L. REV. 799, 824–25 (2002) (“Karl Llewellyn had built much of the Uniform Commercial Code upon a rejection of the plain meaning rule and upon a concept of the ‘agreement in fact,’ and by 1968, the Code had been adopted throughout the United States.”).

\textsuperscript{25} See RESTATEMENT (SECOND) OF CONTRACTS, supra note 17, § 212 cmt. b (“Plain meaning and extrinsic evidence. It is sometimes said that extrinsic evidence cannot change the plain meaning of a writing, but meaning can almost never be plain except in a context. Accordingly, the rule stated in Subsection (1) is not limited to cases where it is determined that the language used is ambiguous. Any determination of meaning or ambiguity should only be made in the light of the relevant evidence of the situation and relations of the parties, the subject matter of the transaction, preliminary negotiations and statements made therein, usages of trade, and the course of dealing between the parties.”).

\textsuperscript{26} See Linzer, supra note 24, at 800 (citing KNIFFIN, supra note 1, § 24.7, at 34).


\textsuperscript{28} Brian G. Slocum, The Importance of Being Ambiguous: Substantive Canons, Stare Decisis, and the Central Role of Ambiguity Determinations in the Administrative State, 69 MD. L. REV. 791, 794 (2010) (emphasis added); see also Alan Schwartz & Robert E. Scott, Contract Theory and the Limits of Contract Law, 113 YALE L.J. 541, 570 n.55 (2003) (“Courts seldom distinguish between ‘vague’ and ‘ambiguous’ terms . . . . More narrowly, however, a word is vague to the extent that it can apply to a wide spectrum of referents, or to referents that cluster around a modal ‘best instance,’ or to somewhat different referents in different people.”).


\textsuperscript{30} Even the notion of the parties’ intent is sometimes characterized as ambiguous. See Gregory Klass, Interpretation and Construction in Contract Law 2 (Jan. 2018) (unpublished manuscript) (on file with Georgetown University Law Center), https://scholarship.law.georgetown.edu/cgi/viewcontent.cgi?
Language evidence from linguistic corpora can help provide content to otherwise vague terms in the law or help to identify conceptual weaknesses with terms like “plain meaning,” “ambiguity,” and “context.”

C. Operational Challenges to Finding Meaning and Ambiguity

Even if courts had a shared, coherent theory of what plain meaning actually meant, it is not clear that courts would be able to implement such a theory with the tools at their disposal, in particular the courts’ own linguistic intuition and dictionaries.

1. The Limits of Intuition about Plain Meaning and Ambiguity

Courts sometimes make determinations about the meaning of a contract without referencing dictionaries, extrinsic evidence, or any other evidence of meaning beyond the text of the contract. In such circumstances, we can infer that the court is relying on the judge’s linguistic intuition to determine what the contract means. The human language faculty is an extraordinary system for communicating information. But language users are not particularly adept at objectively and predictably identifying and resolving lexical ambiguities when faced with high-frequency, highly polysemous words—for example, words that occur very often and that have a lot of different senses. This is a problem because word frequency is correlated with polysemy. This is a problem because word frequency is correlated with polysemy.

31. Slamow v. Delcol, 174 A.D.2d 725, 726 (N.Y. App. Div. 1991) (relying only on the text to interpret a contract and holding that “when the terms of a written contract are clear and unambiguous, the intent of the parties must be found within the four corners of the contract”).

32. See generally GISELA HAKANSSON & JENNIE WESTANDER, COMMUNICATIONS IN HUMANS AND OTHER ANIMALS 1–22 (2014) (distinguishing human language from other methods of communication).

33. Scott Crossley et al., Polysemy and Frequency Use in English L2 Speakers, 60 LANGUAGE LEARNING 573, 575 (2010). Historically, the reason for the ubiquity of polysemous words is because people have preferred to take words and extend their meaning rather than create new words. This preference, known as the law of least effort, states that speakers will economize their vocabulary by extending word senses in order to conserve lexical storage space. Thus, over time, word meanings are extended so that individual words possess multiple meanings. This is especially true for more frequent words, which tend to be the most polysemous. Because frequent words have the most senses, learners encounter highly polysemous words most often. However, highly polysemous words, although unavoidable, also exhibit higher degrees of ambiguity and could thus be more difficult to process. Id.

34. Id.
it is to have many different senses. Consider that the verb to carry has forty-three separately defined senses in the Oxford English Dictionary (the “OED”), while the word bathymetry has only one. The more common a word is, the more senses it has, and the more senses it has, the more likely two people are to disagree as to its meaning in a given context. This leads to the counterintuitive result that judges and lawyers are more likely to disagree about the meaning of common words than the meaning of uncommon words.

This cognitive limitation may not be the only challenge that courts face when attempting to address questions of plain meaning and ambiguity. At least one study has found that parties frequently and systematically overestimate the degree to which they are likely to agree with other parties to a contract. This phenomenon is referred to as false consensus bias. Courts and parties may fail to recognize false consensus bias and will, therefore, fail to recognize the legitimacy of different readings of the contract. As a consequence, “a judge may consider language to be plain when in fact different people do not understand it the same way, and this may happen even when the judge’s understanding is shared only by a minority of people in general.”

In addition, at least one study has found that simple judgments about ambiguity may be influenced by a party’s policy preferences. As a result, “[t]he person who has a strong view about who ought to win a case (for reasons apart from the text) has trouble seeing the plausibility

35. Id.
37. Id.; see also ENeko AGRre & PHILLIP EDMONDS, WORD SENSE DISAMBIGUATION: ALGORITHMS AND APPLICATIONS 91 (2007) (noting “[h]igh polysemy has a detrimental effect” on the performance of disambiguation tasks); George Tsatsornis et al., An Experimental Study on Unsupervised Graph-based Word Sense Disambiguation in ALEXANDER GULBUKH, COMPUTATIONAL LINGUISTICS AND INTEllIGENT TEXT PROCESSING 184, 193 (2010) (noting that human annotators have higher rates of disagreement when tasked with disambiguating highly polysemous words); KEITH BROWN & KEITH ALLAN, CONCISE ENCYCLOPEDIA OF SEMANTICS 224 (2009) (noting that accuracy on word sense disambiguation tasks declines where finer-grained sense distinctions are required).
39. Id.
40. Id. at 1294.
of other ways of reading the statute involved.” 42 In such circumstances, “[i]f one person says that both proposed readings of a statute seem plausible, and a colleague disagrees, finding one reading too strained, what is there to do about it but for each to stamp his foot?” 43

Results like those discussed in the preceding paragraph have been called into question by a subsequent study of judges’ responses, designed to “assess legal problems designed to trigger unconscious political bias in members of the general public.” 44 The study found that “[j]udges of diverse cultural outlooks . . . converged on results in cases that strongly divided comparably diverse members of the public,” 45 and that “professional judgment [of judges] can be expected to counteract ‘identity-protective cognition,’ the species of motivated reasoning known to generate political polarization . . . .” 46 However, if judges are equipped to combat motivated reasoning, it does not necessarily follow that judges have special advantages when called upon to identify linguistic ambiguity. Some linguistic information simply may not be available to the judge via introspection.

This is not to say that the use of language evidence from linguistic corpora will eliminate the influence of cognitive biases in judicial reasoning about interpretive questions. Corpus evidence requires interpretation, analysis, and judgment. However, the evidence of language use from a linguistic corpus can allow jurists to check their intuitions about language usage and can allow the corpus user to verify claims about language and meaning and make claims about language use that are themselves verifiable or falsifiable.

2. Dictionaries and the “Baffled Judge”

Faced with the daunting task of finding plain meaning in contractual language, judges will often go “looking for comfortable reassurance” in a dictionary 47—what Justice Robert Jackson referred to as “the last resort of the baffled judge.” 48 The citation to dictionaries often goes hand-in-glove with invocations of the Plain Meaning Rule. A prominent

---

42. Id.
43. Id. at 276.
45. Id.
46. Id.
47. RANDOLPH QUIRK, STYLE AND COMMUNICATION IN THE ENGLISH LANGUAGE 86 (1982).
treatise on New York contract law observes: “When interpreting a contract, words and phrases used by the parties should be given their plain meaning . . . . A dictionary can supply the common sense meaning of words. It may be employed to determine the plain meaning of contract terms." While the treatise couches the use of dictionaries in discretionary terms, dictionary use is not always a matter of judicial discretion. Sometimes it is mandated by precedent and the failure to use a dictionary can result in reversal. In Lorillard Tobacco Co. v. American Legacy Foundation, the Delaware Supreme Court reversed a Chancery Court opinion, holding:

[The] Vice Chancellor’s abandonment of all dictionaries and his innovative review of how legal writers have used ordinary words in their texts to ascertain the plain meaning of the words are not supported by precedent. Under well-settled case law, Delaware courts look to dictionaries for assistance in determining the plain meaning of terms which are not defined in a contract.

The trouble is that dictionaries simply do not say which meanings are plain in the context of a given contract. Dictionaries do not set out to identify plain meaning. As the editors of the Webster’s Third New International Dictionary expressly state, the dictionary “does not evaluate senses or establish an enduring hierarchy of importance among them. The best sense is the one that most aptly fits the context of an actual genuine utterance.” As will be discussed more fully below, linguistic corpora can give access to the contexts in which actual genuine utterances occur in ways that dictionaries cannot.

Even if dictionary editors set out to provide the plain meaning of contractual terms, it is not at all clear that they would be able to do so. Dictionary editors cannot possibly anticipate the contexts in which a given word will be used in a given contract. As Professors Hart and Sacks observed in their influential Legal Process lectures: “A dictionary, it is vital to observe, never says what meaning a word must bear in a

50. 903 A.2d 728, 738 (Del. 2006).
51. Id.
52. WEBSTER’S THIRD NEW INTERNATIONAL DICTIONARY 17a (1971) (emphasis added).
53. Rickie Sonpal, Old Dictionaries and New Textualists, 71 FORDHAM L. REV. 2177, 2206 (2003) ("Dictionaries, by their very nature, do not provide the precise meaning of a word as it is used in a particular context.").
particular context."\(^{54}\) Dictionaries set out to document which meanings are attested, not which meanings are plain or ordinary. Moreover, dictionary making is a decidedly human endeavor. All but the most recent dictionaries were produced with citation files that are liable to be “unrepresentative of the language as a whole,” and reviewed by editors who “all too often ignore common usages and give disproportionate attention to uncommon ones . . .”\(^{55}\)

Dictionaries are not complete repositories of every sense in which a given word has been used and every context in which a given word has appeared. While lexicographers have estimated the total number of English words at two million or more, “[e]ven the largest unabridged American dictionaries contain well under half this total.”\(^{56}\) Moreover, “[a]bridged and collegiate dictionaries include a smaller number of words and—more relevant—they contain fewer and shorter definitions for these words.”\(^{57}\) And “all dictionaries must deal with space limitations that affect the usage listings and amplifications contained in their definitions.”\(^{58}\)

As demonstrated below, courts will sometimes attempt to bolster claims about plain meaning by citing multiple dictionaries. But if dictionaries do not contain the plain meaning of a given word in its contractual context, then citing multiple dictionaries simply compounds the problem. Moreover, it is not always clear that dictionaries have arrived at their definitions independently—“[t]he history of English lexicography usually consists of a recital of successive and often successful acts of piracy.”\(^{59}\) This is because “[d]ictionary editors look at each other’s books, and though editors form their own opinions about what ground should be covered, they dare not depart too far from the area laid out by their competitors.”\(^{60}\)

---


55. SIDNEY I. LANDAU, DICTIONARIES: THE ART AND CRAFT OF LEXICOGRAPHY 104 (2d ed. 2001); see also DOUGLAS BIBER ET AL., CORPUS LINGUISTICS: INVESTIGATING LANGUAGE STRUCTURE AND USE 26 (1996) (“[C]itation slips represent only those contexts that a human reader happens to notice (in some cases representing only the more unusual uses.”).

56. 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, 512 (2013); see also Jean-Baptiste Michael et al., Quantitative Analysis of Culture Using Millions of Digitized Books, 331 SCI. 176, 177 (2011) (“This gap between dictionaries and the lexicon results from a balance that every dictionary must strike: It must be comprehensive enough to be a useful reference but concise enough to be printed, shipped, and used. As such, many infrequent words are omitted.”).


58. Id. at 513.

59. LANDAU, supra note 55, at 43.

60. Id. at 402.
Sometimes courts will appeal to the ranking of senses in a dictionary, claiming that a given sense should be given priority because it is listed “first” in the dictionary. But the dictionaries most commonly relied upon by judges do not rank their senses according to “plainness” or “ordinariness,” and consequently such a reliance on a dictionary’s ranking of senses is fallacious. And courts will sometimes arbitrarily select a single sense from a dictionary as the plain meaning, while ignoring a relevant competing sense.

Courts have failed to adopt a principled basis for choosing between legal dictionaries and general-use dictionaries. And courts have failed to adopt a principled basis for selecting between prescriptive and descriptive dictionaries.

When attempting to account for historical usage, courts will sometimes appeal to a dictionary published around the time the document in question was executed. However, this practice, while intuitively appealing has many drawbacks. The publication date of a historical dictionary can be deceptive as some historical dictionaries were merely reprinted with a new addition and were not updated to reflect contemporary usage. Moreover, historical dictionaries may rely on usage examples “composed centuries before the dictionary was compiled.” And they may simply copy usage examples from prior dictionaries. Courts have failed to adopt a principled standard for when to use a historical dictionary. If interpreters are going to attempt to learn about the linguistic conventions prevailing at the time a document

61. See Mouritsen, The Dictionary is Not a Fortress, supra note 13, at 1924–29 (discussing problems with dictionary usage by courts and identifying the “sense-ranking fallacy”); see also Brudney & Baum, supra note 56, at 514–15.

62. See Mouritsen, The Dictionary is Not a Fortress, supra note 13, at 1924–29 (discussing the Sense Ranking Fallacy and noting that both the Webster’s Third New International Dictionary and Oxford English Dictionary rank their senses historically, according to when a given sense first appeared in the language); see also Brudney & Baum, supra note 56, at 514–15.


64. See Brudney & Baum, supra note 56, at 510.

65. Id. at 507.

66. See Sonpal, supra note 53, at 2209–10 (“[S]ome popular older dictionaries were not only reprinted but even appeared in new editions without any substantive change to the body of the dictionary . . . . Accordingly, judges who carefully choose the printing or edition of an old dictionary . . . risk relying on a dictionary the substance of which far antecedes the statute.”).

67. Id. at 2207.

68. See Joseph W. Reed, Jr., Noah Websters Debt to Samuel Johnson, 37 AM. SPEECH 95 (1962) (noting that fully a third of Webster’s definitions were lifted verbatim or close to verbatim from Johnson without attribution).

69. See Brudney & Baum, supra note 56, at 511.
was executed, they will need a better source for historical usage than a historical dictionary.

Indeed, because the information presented in dictionaries is more complicated than judges sometimes recognize, Arthur L. Corbin urged caution about drawing knee-jerk conclusions about plain meaning:

It is true that when a judge reads the words of a contract he may jump to the instant and confident opinion that they have but one reasonable meaning and that he knows what it is. A greater familiarity with dictionaries and the usages of words, a better understanding of the uncertainties of language, and a comparative study of more cases in the field of interpretation, will make one beware of holding such an opinion so recklessly arrived at.\textsuperscript{70}

This is not to suggest that dictionaries cannot be useful in the interpretation of contracts. To begin with, judges and lawyers can use dictionaries for their most obvious purpose—defining unknown terms. Such a use of dictionaries is uncontroversial.\textsuperscript{71} Dictionaries may also be used for purposes of attestation—to confirm that a contested sense of a word has actually been used and recorded by lexicographers. In this respect, unabridged dictionaries are “historical records . . . of the meanings with which words have in fact been used by writers of good repute. They are often useful in answering hard questions of whether, in an appropriate context, a particular meaning is linguistically permissible.”\textsuperscript{72} Dictionaries can give interpreters a sense of the range of possible uses a given word may have had. And because judges and lawyers are often called upon to describe fine nuances between senses, dictionary definitions can model the way in which to describe these fine differences in meaning. But dictionaries cannot tell their readers the meaning a word must bear in the particular context of a given contract.

Corpus linguistics may offer a more nuanced and accurate picture of language use than a dictionary-based approach to plain meaning. Corpus linguistics is an evidence-based approach to the study of language that

\textsuperscript{70} Arthur L. Corbin, Corbin on Contracts § 535 (1952).

\textsuperscript{71} See Lawrence Solan, When Judges Use Dictionaries, 68 Am. Speech 50, 55 (1993) (“In other cases, the dictionary is used to give the reader a general sense of the word, which seems to me an appropriate use of the dictionary, whether or not it is necessary.”); Craig Hoffman, Parse the Sentence First: Curbing the Urge to Resort to the Dictionary when Interpreting Legal Texts, 6 N.Y.U. J. Legis. & Pub. Pol’y 401, 416 (2003) (“Definition is also used when the Court simply does not know (or believes that the reader may not know) the accurate definition of a word that it is using . . . . Needless to say, [this approach] is completely appropriate when ‘definition’ is the Court’s sole objective.”).

\textsuperscript{72} See HART SACKS, supra note 54, at 1375–76.
involves the examination of language evidence assembled in large, electronic collections of texts known as corpora.\textsuperscript{73} Corpus linguists draw on this language evidence to make observations about the language conventions of a given speech community, from a given timeframe, spoken (or written) in a given context. \textsuperscript{74} Using evidence from linguistic corpora, judges and lawyers can learn to test their hypotheses about the meaning of contractual language against the observable language conventions from a given speech community. Where the speech communities of the contracting parties differ, comparative corpus evidence could be assembled to identify differences in the language use of these communities. Such evidence-based observations allow the corpus user to make informed judgments about the way that a given word or phrase is used in a given context, in a given speech community, and at a given time. Moreover, linguistic corpora are freely available and can be accessed from any computer terminal.\textsuperscript{75} In addition, while the analysis of corpus evidence still depends on human intuition and human judgment, there is an important contrast with intuition-based judgments about plain meaning in that the results of a corpus-based approach are replicable and falsifiable.

II. WHAT WE MEAN WHEN WE TALK ABOUT MEANING AND CONTEXT

Courts agree that the meaning of the words of the contract must be understood “in context.”\textsuperscript{76} But courts rarely identify what is meant when contractual context is invoked. In contextualist jurisdictions (following the Restatement), “context” may include extratextual background information regarding “the subject matter of the transaction, preliminary negotiations and statements made therein, usages of trade, and the course

\textsuperscript{73} SUSAN HUNSTON & DAVID OAKEY, INTRODUCING APPLIED LINGUISTICS 16 (2010) (“A corpus is a large collection of language use, in the form of written texts or transcripts of speech, usually stored on a computer and often designed to be a representation of the way a language is used . . . [T]he object of study becomes the language system itself, rather than individual [language] users’ intuitive knowledge of the language system . . . .”).

\textsuperscript{74} HANS LINDQUIST, CORPUS LINGUISTICS AND THE DESCRIPTION OF ENGLISH 1 (2010) (“[It] cannot be denied that corpus linguistics is also frequently associated with a certain outlook on language. At the centre of this outlook is that the rules of language are usage-based and that changes occur when speakers use language to communicate with each other.”).

\textsuperscript{75} The corpora relied upon for this paper are all freely available without a paid subscription fee. See, e.g., CORPUS OF CONTEMPORARY AMERICAN ENGLISH, https://www.english-corpora.org/coca/ [https://perma.cc/C7E6-Y36P].

\textsuperscript{76} Lorillard Tobacco Co. v. Am. Legacy Found., 903 A.2d 728, 740 (Del. 2006) (“A court must accept and apply the plain meaning of an unambiguous term in the context of the contract language and circumstances, insofar as the parties themselves would have agreed \textit{ex ante}.”).
of dealing between the parties.” But even contextualist courts appear to recognize a separate category of linguistic context in which the text of the agreement is situated and which serves as the most important evidence of the parties’ intent. And formalist courts who closely adhere to the Plain Meaning Rule similarly insist that such linguistic context is essential to determining the meaning of a contractual provision.

The question, then, is what do we mean when we talk about the linguistic context of a contractual provision? This becomes a very important question when attempting to evaluate the meaning of a contractual provision based on evidence of language use. In so doing, jurists may want to examine evidence of language use from contexts that are similar to those that prevail in the relevant contractual provision. This requires a meaningful way to account for the environment in which the words of the contract are used.

A. Semantic, Syntactic, and Pragmatic Context

Context can encompass both verbal and non-verbal aspects of communication. The verbal context in which a given contractual provision occurs may include both its syntactic and semantic environment. Syntax concerns itself with the way words are arranged in a sentence to convey meaning. To understand the meaning of a word or phrase, one must consider how it is used in the sentence and how the sentence is structured. Semantics, on the other hand, deals with the meaning of words and phrases within a sentence. Pragmatics, finally, concerns itself with the context in which a sentence is used and how it is interpreted by listeners or readers.

77. Restatement (Second) of Contracts, supra note 17, § 212 cmt. b.
78. People ex rel. Lockyer v. R.J. Reynolds Tobacco Co., 107 Cal. App. 4th 516, 526 (Cal. Ct. App. 2003) (“The language in a contract must be construed in the context of that instrument as a whole.”); see also Restatement (Second) of Contracts, supra note 17, § 212 cmt. b. (“After the transaction has been shown in all its length and breadth, the words of an integrated agreement remain the most important evidence of intention.”).
79. See, e.g., Olson v. Kehoe Component Sales, Inc., 242 A.D. 2d 902, 903 (N.Y. App. Div. 1997) (“The agreement must be read in context and words must be given their ordinary and plain meaning.”) (citations omitted); see also 7A William Meade Fletcher, Cyclopedic of the Law of Corporations § 3640 (West 2018) (“Where the language of a corporate charter instrument is plain and clear, the court will not resort to extrinsic evidence in order to aid in interpretation, but will enforce the contract in accordance with the plain meaning of its terms. The provision must be read in context.”).
80. In this section, I will layout a framework for linguistic “context” for the interpretation of contracts that is similar to the framework discussed in a prior article on statutory interpretation, though I will add additional insights from the corpus literature here. Cf. Thomas R. Lee & Stephen C. Mouritsen, Judging Ordinary Meaning, Yale L.J. 788, 813–28 (2018).
phrase, interpreters should look for examples of the word or phrase in a similar syntactic environment. Semantics is the study of meaning at the word or phrase level. Words and phrases have embedded within them a number of concepts (sometimes referred to as the semantic features or semantic components of a word) that have bearing on meaning. A usage-based evaluation of contractual meaning must endeavor to evaluate usage evidence that shares semantic features with the word or phrase in the contract.

If the goal is to evaluate meaning on the basis of language usage that is similar to the language of the contract, insights about semantic meaning from corpus linguistics can provide guideposts for the collection of evidence. First, courts often assume that word forms with the same base form (sometimes called a lemma) tend to have the same meaning, regardless of how they are inflected (for example, whether they appear in the singular or plural, or first, second, or third person form). But this is not always the case. Some word forms have markedly different meanings when employed with different inflections. To illustrate, “[t]he forms moving and moved share some meanings with move, but each form has a very distinctive pattern of meaning . . . In the word moving for example there is the meaning of emotional affection, which is quite prominent.” In evaluating usage evidence, interpreters would look for evidence of usage of a given word or phrase with similar inflection to that of the operative word in the contractual provision.

Second, courts may assume that the meaning of words is consistent across parts of speech. But this is not always the case. Indeed, “a word

also concerns the study of these rules and principles. NOAM CHOMSKY, SYNTACTIC STRUCTURES 11 (1957) (“Syntax is the study of the principles and processes by which sentences are constructed in particular languages.”).

83. MICHAEL MORRIS, AN INTRODUCTION TO THE PHILOSOPHY OF LANGUAGE 152 (2006) (“Semantics is the attempt to give a systematic explanation of how the meaning of sentences depends upon the meaning of their parts.”).

84. JOHN I. SAEED, SEMANTICS 260–65 (4th ed. 2016). These features include concepts like number, animacy, gender, humanness, and concreteness (i.e., tangibleness). Id.

85. I use the word “inflection” throughout in its linguistic sense to mean “a : the change of form that words undergo to mark such distinctions as those of case, gender, number, tense, person, mood, or voice[] b: a form, suffix, or element involved in such variation.” See Inflection, MERRIAM-WEBSTER, https://www.merriam-webster.com/dictionary/inflection [https://perma.cc/CT2D-DEUZ].

86. JOHN SINCLAIR, TRUST THE TEXT: LANGUAGE, CORPUS AND DISCOURSE 17 (2004).

87. See, e.g., AT&T, Inc. v. FCC, 582 F.3d 490, 497 (3rd Cir. 2009) (“It would be very odd indeed for an adjectival form of a defined term not to refer back to that defined term.”); Del. River Stevedores v. DiFidelto, 440 F.3d 615, 623 (3d Cir. 2006) (Fisher, J., concurring) (stating that it is a “grammatical imperative[]” that “a statute which defines a noun has thereby defined the adjectival form of that noun” (emphasis added)). The United States Supreme Court expressly rejected this “grammatical imperative” argument, holding: “Adjectives typically reflect the meaning of
which can be used in more than one word class is likely to have meanings associated specifically with each word class.88 Thus, for example, the word combat when used as a noun tends to be concerned with the physical side of combat, while the verb combat tends to be concerned with the social side.89 These separate meanings, which are specific to a particular part of speech, can only be discovered by examining evidence of the use of that part of speech in context.

Third, some information about word meaning and word usage is not available via introspection. One example of linguistic information that is unavailable via introspection is semantic prosody.90 “Words or phrases are said to have a negative or positive semantic prosody if they typically co-occur with [words] that have negative or positive meaning.”91 Consider the following list of words that most frequently co-occur with the ostensibly neutral verb to happen.92 Nothing about the verb itself suggests that it would have a positive or negative semantic prosody, but it happens that happen more frequently collocates with negative outcomes than positive ones.

Table 1: Collocates of “Happen”

<table>
<thead>
<tr>
<th></th>
<th>WHA</th>
<th>7</th>
<th>TERRIBLE</th>
<th>13</th>
<th>UNTHINKABLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>WHAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>SOMETH</td>
<td>8</td>
<td>ACCIDENT</td>
<td>14</td>
<td>COINCIDE</td>
</tr>
</tbody>
</table>


88. See SINCLAIR, TRUST THE TEXT, supra note 86, at 18.

89. Id.; see also id. at 19 (“Each grammatical feature will probably correlate with just one meaning, unless it is a very common word, or a word of very multifarious meaning, in which the case the same grammar may apply to two or three meanings.”).

90. JOHN SINCLAIR, READING CONCORDANCES: AN INTRODUCTION 178 (2003) (”A corpus enables us to see words grouping together to make special meanings that relate not so much to their dictionary meanings as to the reasons why they were chosen together. This kind of meaning is called semantic prosody.”); see also SUSAN HUNSTON & GILL FRANCIS, PATTERN GRAMMAR: A CORPUS-DRIVEN APPROACH TO THE LEXICAL GRAMMAR OF ENGLISH 140 (2000) (listing examples of semantic prosody and noting that “[i]f a collocation is chosen which is at odds with the usual semantic set an ‘extra’ meaning will be implied which may convey irony or even insincerity”); ALAN PARTINGTON, PATTERNS AND MEANINGS: USING CORPORA FOR ENGLISH LANGUAGE RESEARCH AND TEACHING 66–67 (1998) (discussing the negative semantic prosody of commit); Justin Sytsma et al., Causal Attributions and Corpus Analysis, Methodological Advances, EXPERIMENTAL PHIL. (2019) (discussing the negative semantic prosody of the verb cause).


92. For details on how linguistic evidence reveals such co-occurrence, see the discussion of “collocation,” see infra note 123 and accompanying text.
While it is clear from these results that some miracles happen, so too do accidents happen, the unthinkable happens; terrible, horrible, untoward things happen.\(^93\) Thankfully, these collocates do not tell us anything definitively gloomy about the nature of the universe, but they do reveal something about the use of the verb happen that we likely could not have discovered via introspection. Evaluating the meaning of words in a contract may require access to evidence about the common environments in which the words of a contract occur. Some of this evidence cannot be found in a dictionary and cannot be gathered through introspection.

Fourth, there must be a systematic way for gathering evidence of usage because the choice of words by a competent English speaker does not always constitute an independent selection.\(^94\) Word usage is highly influenced by a process of co-selection, in which “the choice of one word conditions the choice of the next, and of the next again.”\(^95\) Consider the sentence, “[t]he restaurant would not seat Mary’s party because it was at [full] capacity.” This is a perfectly well-formed English sentence that carries the same basic meaning whether or not the bracketed word full is included. Full may appear in this or similar sentences through the processes of co-selection and delexicalization, discussed below.\(^96\) This feature of ordinary language use can have important implications for an interpreter’s evaluation of the meaning of an utterance.

Finally, an evidence-based approach to meaning should look for ways to incorporate information about pragmatic context—which may include

<table>
<thead>
<tr>
<th></th>
<th>THINGS</th>
<th>WHATS</th>
<th>UNTOWARD</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>NOTHING</td>
<td>OVERNIGHT</td>
<td>BENGHAZI</td>
</tr>
<tr>
<td>4</td>
<td>WHATEVER</td>
<td>ACCIDENTS</td>
<td>NATALEE</td>
</tr>
<tr>
<td>5</td>
<td>EXACTLY</td>
<td>MIRACLES</td>
<td>INSTANTANEOUSLY</td>
</tr>
</tbody>
</table>

---

93. On the COCA website: (1) Select “Collocates”; (2) Enter “happen_v” in the “Word/phrase” field; (3) Enter an asterisk “*” (a wildcard) in the “Collocates” field; (4) Select “Sort/Limit” and set the “Minimum” to “Mutual Info”; and (5) Click “Find collocates.” Corpus of Contemporary American English, ENGLISH-CORPORA.ORG, http://corpus.byu.edu/coca/ [https://perma.cc/A584-3SGA] [hereinafter COCA Website]. Throughout this Article, wherever language evidence from a corpus is referenced, either a link or detailed instructions on how to exercise a given search will be provided. Because the corpora relied upon for this Article are monitor corpora (corpora that are routinely updated to reflect current usage), some variation in the co-occurrence information may be expected.

94. SINCLAIR, TRUST THE TEXT, supra note 86, at 19.

95. Id.

96. See infra section IV.D.3.
the physical or social setting of an utterance or other information that is not encoded in the words themselves. These contexts will not always be included in the text. However, some pragmatic information may be recorded in the corpus and may be brought to bear on questions of Plain Meaning.

B. Plain Meaning and Historical Context

An evaluation of plain meaning ought to take into account the possibility of language change. If the contract is very recent, then this aspect of plain meaning becomes less important. But where the contract at issue is an older instrument, a method that accounts for the temporal dimension of interpretation is required. Consider the following four definitions of car, listed in chronological order in the Oxford English Dictionary:

1. a. A wheeled, usually horse-drawn conveyance; a carriage, cart, or wagon.
2. a. The passenger compartment of a balloon, airship, cableway, etc.; a gondola.
3. a. A railway carriage or wagon . . .
4. motor car n.2. Now the usual sense.

We can imagine that a contract governing the importation of cars, but executed at a more distant point in history, would have a different meaning. Courts often attempt to take into account the possibility of language change by interpreting contracts according to the meaning of the words in the contract at the time the contract was drafted. Human

97. See Duranti & Goodwin, supra note 81, at 6–9. Pragmatics includes concepts like conversational implicature, where the meaning of an utterance is strongly implied but not expressly stated—as where a spouse who says “there sure are a lot of dishes in the sink” is not just making an observation about the state of the universe but is reminding somebody about whose turn it is to do the dishes.


99. Winn-Dixie Stores, Inc. v. Dolgencorp, LLC, 746 F.3d 1008, 1022 (11th Cir. 2014) (quoting Moore v. Stevens, 106 So. 901, 903 ( Fla. 1925) (“[W]ords used must be given their ordinary, obvious meaning as commonly understood at the time the instrument containing the covenants was executed . . ..” (emphasis added)); Hartig Drug Co. v. Hartig, 602 N.W. 2d 794, 798 (Iowa 1999) (“[T]he words are given the meaning at the time the contract was executed.”); AIU Ins. Co. v. Superior Ct., 51 Cal. 3d 807, 821 (1990) (citing CAL. CIV. CODE § 1636) (“Under statutory rules of contract interpretation, the mutual intention of the parties at the time the contract is formed governs interpretation.” (emphasis added)).
language is constantly, naturally changing. But language change doesn’t occur at a constant, predictable rate, and methods of interpretation need a meaningful way to account for such changes. One way to do this is with language evidence from the relevant timeframe.

C. Meaning and Speech Community

Finally, evaluation of contractual meaning should take into account the possibility of differences in the linguistic conventions of different speech communities involved in contract drafting. It is not difficult to imagine contract cases in which differing linguistic conventions of the different communities involved in drafting the contract might lead to different understandings of the same contractual language. Some contracts are executed between sophisticated commercial parties, with specialized knowledge of both contracting conventions of a particular industry and the characteristics of the underlying industry itself. Other contracts are more public-facing and are executed between firms and their customers. Some such differences are anticipated by interpretive rules in the context of the Uniform Commercial Code (UCC), where variations in trade usage are taken into account. In such cases, where the question is whether competing industries, different geographical

---

100. JOHN LYONS, INTRODUCTION TO THEORETICAL LINGUISTICS 43 (1968) ("All languages are subject to constant change. This is an empirical fact . . . . All living languages . . . . are of their nature efficient and viable systems of communication serving the different and multifarious social needs of the communities that use them.").

101. TERRY CROWLEY & CLAIRE BOWERN, AN INTRODUCTION TO HISTORICAL LINGUISTICS 149–51 (2011) (discussing criticisms of attempts to quantify the rate of language change); LYLE CAMPBELL, HISTORICAL LINGUISTICS 201 (2d ed. 2004) (Glottochronology, a theory attempting to account for the rate of language change "has proven particularly misleading and it is important to understand why it should be avoided. . . . [I]t is important to understand why it does not work for subgrouping, or for any other purpose, for that matter." (emphasis added)).

102. See, e.g., MARCYLENA H. MORGAN, SPEECH COMMUNITIES: KEY TOPICS IN LINGUISTIC ANTHROPOLOGY 1 (2014) ("Speech communities are groups that share values and attitudes about language use, varieties and practices. These communities develop through prolonged interaction among those who operate within these shared and recognized beliefs and value systems regarding forms and styles of communication."); Reed Dickerson, STATUTORY INTERPRETATION: DIPPING INTO LEGISLATIVE HISTORY, 11 HOFSTRA L. REV. 1125, 1154 (1983) (defining speech community as the "group of people who share a common language (or sublanguage) and thus a common culture (or subculture), which in turn defines the context that conditions the utterances that occur within it.") Kamal K. Sridhar, SOCIETAL MULTILINGUALISM, IN SOCIOLINGUISTICS AND LANGUAGE TEACHING 47, 49 (Sandra Lee McKay & Nancy H. Hornberger eds., 1996) ("A conglomeration of individuals who share the same norms about communication is referred to as a speech community. A speech community is defined as a community sharing a knowledge of the rules for the conduct and interpretation of speech.").

103. See U.C.C. § 1-303 (defining "usage of trade").
regions, lawyers or their clients, or sophisticated and unsophisticated parties use a given term differently, we might look to evidence of comparative usage in order to evaluate the plain meaning of contractual language.

III. A CORPUS-BASED APPROACH TO PLAIN MEANING

Corpus linguistics is a tool that can assist judges and lawyers in evaluating claims about meaning and ambiguity in a contract. Corpus linguistics can be understood as “the study of language data on a large scale—the computer-aided analysis of very extensive collections of transcribed utterances or written texts.”104 It is premised “on the idea that the description of the language cannot be made just from the intuition of the linguist, but that it requires the handling of a set of real language samples.”105 Linguistic corpora are typically comprised of samples of natural language—written text and transcribed speech that were produced in a natural setting (and collected after the fact by a corpus architect) rather than speech or text that are gathered through traditional methods of elicitation, like surveys, interviews, or observation.106

Linguistic corpora are collections of digitized texts that are often annotated (or tagged) with additional grammatical information (or metadata) to facilitate the study of language.107 In virtually every circumstance where technology and language cross paths—in machine translation, speech recognition, or language pedagogy—linguistic corpora are used to provide objective and accurate information about the


105. Francisco Moreno-Fernández, Corpora of Spoken Spanish Language—The Representativeness Issue, in Linguistic Informatics State of the Art and the Future 120 (Yuji Kawaguchi et al. eds., 2005); see also David Oakey, English Vocabulary and Collocation, in Susan Hunston & David Oakey, Introducing Applied Linguistics 16 (2009) (“A corpus is a large collection of language use, in the form of written texts or transcripts of speech, usually stored on a computer and often designed to be a representation of the way a language is used. . . . [T]he object of study becomes the language system itself, rather than individual [language] users’ intuitive knowledge of the language system . . . .”); James Simpson, The Routledge Handbook of Applied Linguistics 597 (2011) (“At the heart of empirically based linguistics and data-driven description of language, corpus linguistics is concerned with language use in real contexts.”); Elena Tognini-Bonelli, Corpus Linguistics at Work 1 (2001) (“[I]t cannot be denied that corpus linguistics is also frequently associated with a certain outlook on language. At the centre of this outlook is that the rules of language are usage-based and that changes occur when speakers use language to communicate with each other.”).


107. Id. at 433.
linguistic conventions of the speech community in question. Contemporary lexicographers use corpora to provide objective and accurate information about the way in which words are used. And because sophisticated linguistic corpora are freely available to anyone with access to a computer and an internet connection, the same type of evidence used by contemporary lexicographers to draft dictionaries is now available to anyone interested in gathering objective evidence about language use.

Linguistic corpora are designed to be “maximally representative finite sample[s]” of language use, which “enable[] results to be quantified and compared to other results in the same way as any other scientific investigation which is based on a data sample.”

The distinctive characteristic of corpus linguistics is the claim that it is possible to actually ‘represent’ a domain of language use with a corpus of texts, and possible to empirically describe linguistic patterns of use through analysis of that corpus. Any research question relating to linguistic variation and use can be approached from this methodological perspective.

While corpus linguistics follows in a long tradition of using empirical linguistic evidence to understand human language (in fields such as field linguistics and lexicography), the emergence of corpus linguistics as a separate discipline is fairly recent and has moved in tandem with the rise of the personal computer. The exponential growth in computing

108. See, e.g., id. at 102, 513.

109. See, e.g., OXFORD DICTIONARY OF ENGLISH xi (3d ed. 2010) (“The general principle on which the senses in the Oxford Dictionary of English are organized is that each word or part of speech has at least one core sense or core meaning, to which a number of subsenses may be attached...Core meanings represent typical, central uses of the word in question in modern standard English, as established by analysis of the Oxford English Corpus and our other language databases.”). Note that while some recent, contemporary dictionaries rely on linguistic corpora in their construction, many of the dictionaries most commonly cited by courts—including WEBSTER’S THIRD NEW INTERNATIONAL, WEBSTER’S SECOND NEW INTERNATIONAL DICTIONARY, and the OXFORD ENGLISH DICTIONARY (2d ed. 1989), see infra note 215, were developed prior to the emergence of large-scale sophisticated corpora and have not relied upon linguistic corpora in their construction. And even where the lexicographer uses corpora to prepare a dictionary, the dictionary itself, with limitations of space and coverage cannot anticipate every specific context in which a word is used.


111. TONY McENERY & ANDREW WILSON, CORPUS LINGUISTICS: AN INTRODUCTION 75 (2011).


113. O’KEEFE & McCARTHY, supra note 106, at 15 (discussing the relationship among corpus linguistics, lexicography and field linguistics).

114. See TOGNINI-BONELLI, supra note 105, at 5.
power and the increase in the availability of computers to individual users has resulted in the concomitant increase in size, availability, and sophistication of linguistic corpora.115 Today, corpus linguistics allows for real-time examination of “a quantity of information that could hardly be envisaged by a team of informants working over decades even 50 years ago . . . .”116

“Empiricism lies at the core of corpus linguistics . . . .”117 The promise of corpus linguistics is that the corpus—like many other instruments of scientific observation and discovery—can reveal facts about the natural world that cannot be perceived through ordinary means of human perception.118 Corpus linguists test their hypotheses about language through rigorous observation with evidence that only a corpus can render observable and quantifiable. And, importantly, the results of any such corpus-based experiment can also be rendered replicable and falsifiable.119

Linguistic corpora may vary widely in terms of size, design, and purpose.120 They can be tailored to represent the linguistic behavior of a particular speech community or linguistic register.121 And corpora can be

115. Id.
116. Id.
117. MCENERY & HARDIE, CORPUS LINGUISTICS: METHOD, THEORY AND PRACTICE, supra note 91, at 49; PAUL BAKER, GLOSSARY OF CORPUS LINGUISTICS 65 (2006) (“In linguistics, empiricism is the idea that the best way to find out about how language works is by analyzing real examples of language as it is actually used. Corpus linguistics is therefore a strongly empirical methodology.”).

118. Writing about the Google Books N-Gram Viewer, a corpus constructed from the contents of Google Books, Jean-Baptiste Michel et al. commented: “The corpus cannot be read by a human. If you tried to read only English-language entries from the year 2000 alone, at the reasonable pace of 200 words/min, without interruptions for food or sleep, it would take 80 years. The sequence of letters is 1000 times longer than the human genome: If you wrote it out in a straight line, it would reach to the Moon and back 10 times over.” See Michael et al., supra note 56, at 176.

119. MCENERY & HARDIE, CORPUS LINGUISTICS: METHOD, THEORY AND PRACTICE, supra note 91, at 66 (“As a key goal of corpus linguistics is to aim for replicability of results, data creators have an important duty to discharge in ensuring that the data they produce is made available to analysts in the future.”).

120. For further discussion of the types of linguistic corpora and the tools they offer, see Thomas R. Lee & Stephen C. Mouritsen, Judging Ordinary Meaning, YALE L.J. 788 (2018).

121. A register is “any of the varieties of a language that a speaker uses in a particular social context.” Register, MERRIAM Webser, https://www.merriam-webster.com/dictionary/register [https://perma.cc/6C8U-JCAH] (last visited Sept. 1, 2019); see Lawrence M. Solan, The New Textualists’ New Text, 38 LOY. L.A. L. REV. 2027, 2059 (2005) (“When the legal system decides to rely on the ordinary meaning of a word, it must also determine which interpretive community’s understanding it wishes to adopt. This choice is made tacitly in legal analysis, but becomes overt when the analysis involves linguistic corpora because the software displays the issue on a screen in front of the researcher.”) (emphasis added)).
constructed to represent language use from any point in history for which there are surviving texts.

Corpora can provide objective evidence of how frequently different words occur in particular semantic environments through a process called collocation. “Collocation is the co-occurrence of words with a frequency that is much higher than it would be by chance.” Courts frequently interpret contractual terms according to the context in which they occur. Collocation can give us a sense of the scope of that context that is not available through introspection. One way collocation does this is by displaying semantic prosody, as discussed above.

Another important feature of a linguistic corpus is the concordance or key word in context (KWIC) function. “A concordance is a collection of the occurrences of a word-form, each in its own textual environment. In its simplest form it is an index. Each word-form is indexed and a reference is given to a place of occurrence in a text.” A concordance allows its user to review a particular word or word form in hundreds of contexts, all on the same page of running text. This allows a corpus user to evaluate words in context systematically, a task that, again, cannot be performed using intuition.

Corpus evidence can be brought to bear to evaluate claims of plain meaning or ambiguity of the contractual provisions, as illustrated below. The primary source for language evidence discussed below is the Corpus of Contemporary American English (COCA). The COCA is “the only large, and genre-balanced corpus of American English.” “The corpus contains more than 560 million words of text (20 million words each year 1990-2017) and it is equally divided among spoken, fiction, popular magazines, newspapers, and academic texts.” The COCA is a monitor corpus that is occasionally updated to reflect usage.

122. See MCENERY & WILSON, CORPUS LINGUISTICS: AN INTRODUCTION, supra note 111, at 82.
123. O’KEEFE & MCCARTHY, supra note 106, at 435; see also SUSAN HUNSTON, CORPORA IN APPLIED LINGUISTICS 68 (2002).
124. O’KEEFE & MCCARTHY, supra note 106, at 167 (quoting JOHN M. SINCLAIR, CORPUS, CONCORDANCE, COLLOCATION, 32 (1991)).
125. See infra note 169 for instructions on viewing a concordance display.
126. See infra Part IV.
128. Id.
129. O’KEEFE & MCCARTHY, supra note 106, at 110 (discussing the features of the COCA); see also id. at 430 (defining monitor corpus).
By contrast, one of the contracts at issue below was executed in 1961. If we are to correctly assess the linguistic conventions prevailing during that period, we need usage evidence from that time period. The Corpus of Historical American English (COHA) has evidence from that time period. The COHA is “the largest structured corpus of historical English.” It contains “more than 400 million words of text from the 1810s-2000s (which makes it 50–100 times as large as other comparable historical corpora of English) and the corpus is balanced by genre decade by decade.”

Using evidence from the written portions of the COCA and the COHA, we can gather information about the linguistic conventions of Standard American English for the time periods in question. But it is not necessarily obvious that Standard American English is the appropriate speech register for the interpretation of all contracts. In the problems addressed below, the courts and the parties turned to unabridged, general-use dictionaries in order to define the operative terms. This suggests that in assessing plain meaning and ambiguity, courts often have a notion of Standard American English in mind when performing interpretive tasks. Moreover, two of the problems below address contract interpretation issues in the insurance context, in which courts often resolve ambiguities against the drafting party and in favor of the insured. When a court attempts to determine the meaning of a contract drafted by a national or multi-national insurance company and entered into by insureds from a variety of locations within the United States, the court may decide to attempt to interpret the contract according to the shared linguistic conventions of a nationwide speech community. Such an interpretive approach may suggest that some notion of Standard American English might be relevant to the resolution of

130. See infra Part IV.C.


132. Id.

133. WALT WOLFRAM & NATALIE SCHILLING, AMERICAN ENGLISH: DIALECTS AND VARIATION 410 (2015) (defining “Standard American English” as a variety of English that is “held to be the linguistic norm and that is relatively unmarked with respect to regional characteristics of English.”).

134. See infra Part IV.

135. See infra section IV.A. (Problem No. 1); infra section IV.B. (Problem No. 2). Linguist Goffrey Nunberg observed that the American Heritage Dictionary’s usage panel “merely reflected the predominant practices of Standard English,” which he defined as “the variety that happens to have been adopted by the educated middle-classes . . . .” LANDAU, supra note 55, at 248.
some interpretive problems in insurance contexts. However, as we will see (and as we might expect), some contractual language reflects specialized legal or industrial usage, and there may be instances when a specialized corpus of legal texts or contracts from a given industry and time period may be necessary to properly address the relevant interpretive questions.

IV. THE CORPUS APPROACH TO PLAIN-MEANING IN PRACTICE

This section will examine a collection of four contract cases, in which courts were faced with difficult interpretive questions. This section will demonstrate some of the language evidence that linguistic corpora can provide, as well as some of the questions posed by the availability of the evidence corpora provide.

A. Problem No. 1: Plain Meaning and “Snorkeling”

An insured ecological tourism company sued its Insurer for refusing to defend or indemnify the Insured in a negligence action arising from customer’s drowning death while snorkeling. The policy at issue, drafted in 2004, states, “[T]his insurance does not apply to ‘bodily injury’ [including death] to any person while practicing for or participating in any sports or athletic contest or exhibition that you sponsor.”

Question: Is snorkeling a sport?

The Insurer sought summary judgment, arguing that snorkeling is a sport and therefore the death of the Insured’s customer was subject to the exclusion. The court disagreed and invoked the Plain Meaning Rule,

---

136. In addition, Professors Alan Schwartz and Robert E. Scott have argued that commercial parties may benefit from a default rule that enshrines interpretation according to Standard Written American English, which is roughly analogous to what they term “majority talk.” Alan Schwartz & Robert E. Scott, Contract Theory and the Limits of Contract Law, 113 Yale L.J. 541, 570 (2003). “[M]ajority talk” is defined as “the language that people typically use when communicating with each other.” Id. Schwartz and Scott advocate the default rule in favor of “majority talk” for commercial parties “because the default would (1) reduce contracting costs, (2) minimize the opportunities for strategic behavior, (3) reduce the risk of judicial error, and (4) expand the set of efficient contracts that parties could write.” Id. at 584.


138. Id.

139. Id.
noting that “it is not clear that snorkeling falls within the plain meaning of ‘sports.’”

1. *Dictionaries and the Plain Meaning of “Sport”*

The court found that the plain meaning of sport excluded snorkeling. Turning to dictionary definitions of sport, the court observed that sport is “defined by Webster’s as ‘an activity [that gives enjoyment or recreation], especially when competitive, requiring more or less vigorous bodily exertion and carried on, sometimes as a profession, according to some traditional form or set of rules.’” The court also cites the American Heritage Dictionary, which defines sport as “physical activity that is governed by a set of rules or customs and often engaged in competitively.” The court then concludes the insurer “does not even suggest that snorkeling is governed by any traditional set of rules or customs, as provided by the dictionary definitions.” Based on these dictionary citations, the court concludes that the “rule-based athletic competition” sense of sport is the term’s plain meaning.

The court’s dictionary-based reasoning is highly suspect. To begin with, the court refers to Webster’s dictionary as if it were a single, unitary book. In fact, there are three American publishers that publish a dictionary (or suite of dictionaries) under the name Webster’s—Houghton Mifflin Company, Random House, and Merriam-Webster, Inc. The use of the Webster name for dictionaries has been in the public domain since 1908. While this offhand reference to Webster’s may...

140. *Id.* at 1197.

141. *Id.* (citing *WEBSTER’S NEW WORLD COLLEGE DICTIONARY* 1297 (3d ed. 1997)).

142. *Id.* (citing *THE AMERICAN HERITAGE DICTIONARY* 1680 (4th ed. 2006)).

143. *Id.* at 1198. The court also makes a grammatical argument, noting “sports” in the “any sports or athletic contest or exhibition that you sponsor” may be functioning as a nominal adjective of “contest or exhibition” and observes that “if ‘sports’ is an adjective modifying ‘contest’ and/or ‘exhibition,’ the Athletic Exclusion is clearly inapplicable.” *Id.* at 1197. Because the court is ruling on the insurer’s summary judgment motion (a motion for which the insurer bears the burden of proof), the court does not affirmatively rule that sports is an adjective or a noun in this context. And the court cites precedent suggesting that such a construction is ambiguous regardless. *Id.* (citing Garcia v. St. Bernard Par. Sch. Bd., 576 So.2d 975, 976 (La. 1991)). In Garcia, the Louisiana State Supreme Court repeatedly reads a similar provision as applying to a “sports contest,” and still finds the provision to be ambiguous. *Garcia*, 567 So.2d at 976–77.


145. G. & C. Merriam Co. v. Ogilvie, 159 F. 638, 640 (1st Cir. 1908) (holding that “[t]he statutory monopoly having expired under statutory limitation, the word ‘Webster’ used in connection with a dictionary, became public property, and any relief granted upon the idea of title or proprietorship in the trade-name of ‘Webster’ would necessarily involve an unwarrantable continuance of the statutory monopoly secured by the copyright.”); *see also* Merriam-Webster, Inc.
not have any great impact on the court’s analysis, it is indicative of the abstract way in which courts sometimes think about dictionaries. As Professor Lawrence Solan has observed, “we commonly ignore the fact that someone sat there and wrote the dictionary which is on our desk, and we speak as though there were only one dictionary, whose lexicographer got all the definitions ‘right’ in some sense that defies analysis.”

While there is not a single Webster’s dictionary that got all the answers right in some abstract sense, there is one thing that all three Webster’s dictionaries have in common (aside from their name): they all define sport as both a competitive activity (“a game or contest esp. when involving individual skill or physical prowess on which money is staked”) and a physical activity that is “a pleasing or amusing pastime or activity: recreation.” The same thing is true for the American Heritage Dictionary, which defines sport as both “[a]n activity involving physical exertion and skill that is governed by a set of rules or customs and often undertaken competitively,” and as “[a] usually challenging activity undertaken for amusement.” Both of these competing senses of sport are attested in common usage. The court never justifies the choice between these two senses, nor the basis for selecting one set of dictionary definitions and ignoring the other set.

---

v. Random House, Inc., 35 F.3d 65, 72–73 (2d Cir. 1994) (finding the terms “Webster’s” and “collegiate” to be generic with respect to dictionaries).


147. Sport, WEBSTER'S THIRD NEW INTERNATIONAL DICTIONARY 2206 (1975) (defining sport as “a source of diversion,” “recreation” . . . “physical activity engaged in for pleasure,” “a particular activity (such as an athletic game) so engaged in”); see also THE RANDOM HOUSE DICTIONARY OF THE ENGLISH LANGUAGE: THE UNABRIDGED EDITION (1994); WEBSTER’S NEW WORLD COLLEGE DICTIONARY 1297 (3d ed. 1997).


149. It is easy to generate examples of the rule-based competition notion of sport. The sports page of any newspaper provides a variety of examples. A search in a newspaper archive reveals a variety of instances in which sport is used to describe snorkeling. Lottie Bogan, Challenge Met With a Splash, THE NORTHSIDE SUN (JACKSON MISSISSIPPI), Mar. 18, 2004, at 5C (“On our day in Saint Thomas, my daughter-in-law Gail and I decided to go snorkeling, a sport I have enjoyed in the past.”); Patricia Farrell, Explore by Snorkel, BOY’S LIFE, Aug. 1978, at 20 (“Snorkeling is a sport you can enjoy for itself or as practice for the day you can do deep dives with air tanks and scuba equipment.”); Alison S. Wellner, AMERICANS AT PLAY: DEMOGRAPHICS OF OUTDOOR RECREATION & TRAVEL 217 (1997) (“Snorkeling is a sport that can be enjoyed even at older ages . . . .”).
2.  *Prototype, Family Resemblance, and the Plain Meaning of “Sport”*

Like many questions of legal interpretation, the question of whether snorkeling is a sport is deceptively simple. But, as Corbin cautioned, we should be skeptical of knee-jerk responses (especially our own knee-jerk responses) to such questions—these questions often prove more difficult and more nuanced than they seem on first impression. Indeed, the question, “what is a sport?” (and the related question, “what is a game?”) have played an important role in foundational texts of both the philosophy of language and cognitive linguistics.

In his posthumously published Philosophical Investigations, language philosopher Ludwig Wittgenstein stated,

Consider for example the proceedings that we call ‘games’. I mean board-games, card-games, ball-games, Olympic games, and so on. What is common to them all? . . . In ball games there is winning and losing; but when a child throws his ball at the wall and catches it again, this feature has disappeared. . . . Think now of games like ring-a-ring-a-roses; here is the element of amusement, but how many other characteristic features have disappeared! . . . And the result of this examination is: we see a complicated network of similarities overlapping and criss-crossing: sometimes overall similarities, sometimes similarities of detail . . . . I can think of no better expression to characterize these similarities than ‘family resemblances’ . . . .

Wittgenstein’s argument is that “the extension of a concept [like a game] may be united not by common characteristics but by overlapping similarities between the members.” And he attempts to show that our use of general terms can only proceed because “overlapping similarities or resemblances between the instances exist rather than a unique set of characteristic marks that they all share.”

The court’s notion of sport as a rule-based athletic competition and the insurer’s preferred sense of physical activity for recreation potentially share a number of features. Both involve physical exertion. Both take place (sometimes, if not often) out of doors. Both can be undertaken for recreation. But, of course, they have a number of

150. See supra note 70, at 496–97.
153. Id.
differences. Wittgenstein reminds us that it is not always easy to make fine distinctions about which terms fall into a category and which do not. Nor is it always possible to list a series of necessary and sufficient conditions that will describe everything that ought to be considered a sport, but that will exclude everything that is not a sport.

Another way to conceptualize the question of the plain meaning of sport is through the linguistic concept of prototype. A prototype can be understood as “the best example of a category.” A robin may a prototypical bird; a chair may be a prototypical piece of furniture. In a landmark paper in the field of cognitive linguistics, Eleanor Rosch conducted a series of experiments that revealed, among other things, that the human notion of concepts “appear to be represented in cognition not as a set of criterial features with clear-cut boundaries but rather in terms of prototype (the clearest cases, best examples) of the category.”

Rosch’s methodology involved ranking experiments in which 209 survey subjects were asked to rate, on a seven-point scale, which words were “good examples” of one of ten categorical terms. The results of the study showed “a high agreement between subjects concerning these rankings.” Thus, chair is a more prototypical example of furniture than footstool, automobile is a more prototypical vehicle than yacht, and robin is a more prototypical bird than ostrich. The categories Rosch examined included fruit, vehicle, weapon, vegetable, carpenter’s tool, bird, toy, clothing, and, importantly for our purposes, sport. As with the other categories examined, Rosch’s data showed that some activities are “better examples” of sport than others. Rosch’s respondents ranked the following activities from most sport-like to least sport-like:

Football, baseball, basketball, tennis, softball, canoeing, handball, rugby, hockey, ice hockey, swimming, track, boxing, volleyball, lacrosse, skiing, golf, polo, surfing, wrestling, gymnastics, cricket, squash, badminton, racing, pole vault, fencing, bowling, water skiing, ice skating, jai alai, skating,

156. See id. at 193.
157. Id. at 198, 229–33.
158. Id. at 198.
159. Id. at 229.
160. Id. at 230.
161. Id. at 232.
162. Id.
skydiving, sailing, diving, archery, judo, car racing, ping pong, rowing, fishing, horseback riding, running, horse racing, hiking, weight lifting, croquet, horseshoes, boating, pool, billiards, hunting, jump rope, camping, chess, dancing, checkers, cards, sunbathing.\(^{163}\)

Of the ten most prototypical sports in Rosch’s study, only one, canoeing,\(^ {164}\) appears to fit within the more narrow confines of the recreational physical activity sense of sport. However, several other recreational activities are rated more sport-like than their rule-based counterparts—for example, sky diving is ranked more sport-like than judo, and fishing is ranked over horse racing, croquet, and horseshoes.\(^ {165}\)

At least some scholars have argued that when judges search for plain meaning, they are sometimes searching for linguistic prototypes.\(^ {166}\) While Rosch’s article suggests the physical recreation notion is represented in our concept of sport (as many such sports rank above their rule-based counterparts), the article also makes clear that the rule-based competition notion of sport predominates. To the extent that we believe our search for plain meaning is a search for prototype only, we could view Rosch’s data as evidence that snorkeling does not come within the plain meaning of sport. Alternatively, we might rely on Wittgenstein’s reasoning above to conclude that there is no set of necessary and sufficient conditions which can definitively describe everything that is a

\(^{163}\) Id. at 200–01, tbl.1.

\(^{164}\) Like many of the more or less sport-like activities ranked in Rosch’s study, canoeing can be both a recreational activity and a rule-based completion. See Canoe Sprint, OLYMPIC.ORG, https://www.olympic.org/canoe-sprint [https://perma.cc/NP7F-JSPP]. However, as will be shown below, an examination of usage demonstrates that canoeing is most often used with reference to recreational activities.

\(^{165}\) One obvious response to this data is to assume that prototypicality judgments are a mere reflection of the frequency with which these sports (or references to them) are encountered in common speech. After all, it is not hard to imagine that a collection of American test subjects would name football, baseball, and basketball as the most prototypical sports. But if that were the case, we would have to accept that the respondents in Rosch’s study had more experience with (or exposure to) judo than they did with fishing, and running. This seems unlikely. Judgments about prototypicality do not predictably map on to measures of statistical frequency. See John R. Taylor, Prototype Theory, in 1 SEMANTICS: AN INTERNATIONAL HANDBOOK OF LANGUAGE MEANING 649–50 (Claudia Maienborn et al. eds. 2011) (“[M]any people are inclined to say that prototypes (or prototypical instances) are encountered more frequently than more marginal examples and that that is what makes the prototypical. [T]his cannot be the whole story.”).

sport and exclude everything that isn’t. Because snorkeling bears a “family resemblance” to some, but not all, sport-like concepts (it takes place out of doors, involves physical exertion, is undertaken for recreation), it comes within the plain meaning of sport. We might also conclude that because snorkeling as sport is at the very least an attested use of the word sport, the recreational activity reading of sport is a possible reading of the word and, as such, the contract is ambiguous and should be interpreted against the drafter. Each of these approaches is a plausible way of thinking about the plain meaning of sport, but courts have not yet decided what they mean when they invoke the plain meaning canon or when they identify contractual ambiguity.  

3. Linguistic Corpora and the Meaning of “Sport”

Rather than arbitrarily selecting among competing dictionary definitions of sport, linguistic corpora can demonstrate the range of potential meanings of sport. We begin with collocation. Collocation data can be used to get a general sense of the range of possible uses of the term sport. Because the general liability policy at issue in Problem No. 1 was executed in 2004, we will rely on collocation data from the COCA. The fifty most common collocates of sport are listed in the COCA as follows:

usa, professional, illustrated, teams, utility, fox, fans, pro, entertainment, marketing, complex, youth, olympic, medicine, o, athletes, bar, writer, authority, vehicles, soccer, vehicle, nbc, cbs, betting, arena, fan, section, columnist, bureau, elias, espn, leagues, motor, fame, tennis, utilities, franchises, gambling, governing, swimming, franchise, fitness, camps, memorabilia, extreme, coupe, venues, coat, volleyball.

There are a number of collocates of sport that suggest the rule-based athletic competition sense. These include straightforward instances of such rule-based athletic competitions like soccer, tennis, and volleyball. They also include collocates that strongly suggest a rule-based competition context, including Olympic, athletes, and franchises. In addition, while it is certainly possible that sports coverage on major networks would include a feature on snorkeling, it seems likely that

167. See supra notes 23–30.

168. On the COCA website: (1) Select “Collocates”; (2) Enter “SPORT_n” in the “Word/phrase” field (capitalization makes the search lemmatized—assuring that we find all inflections of the word; the “_n” is to limit the search noun forms); (3) Enter an asterisk “*” (a wildcard) in the “Collocates” field; (4) Select “Sort/Limit” and set the “Minimum” to “Mutual Info”; and (5) Click “Find collocates.” COCA Website, supra note 93.
those tuning into sports coverage on FOX, NBC, CBS, and ESPN are looking for coverage of rule-based competition. These collocation results appear to track the findings of the Rosch study referenced above. The collocate data give us the initial impression that the prototype of sport in the Rosch study matches the way in which the word appears to be most frequently used. A corpus user can confirm these predictions by examining the use of sport in context, using the concordance (or KWIC) function of the corpus.\textsuperscript{169}

\begin{table}
\centering
\caption{“Sport” in the COCA\textsuperscript{170}}
\begin{tabular}{|c|c|}
\hline
\textbf{television has driven} & \textbf{sports} \\
\hline
the IOC says the channel would promote Olympic & in the years between the games \\
\hline
win. Despite the performance, finding a footing in the & of MMA was tough. Wanting to have a consistent training regimen \\
\hline
rooting him on to his eighth gold. The & of swimming has come a long way \\
\hline
who owes his vast fortune to Andy Griffith reruns and & on television. Turner signed pitcher Andy Messersmith, one of \\
\hline
Like the rest of Cuban society, the nation’s powerful & apparatus has been hit by the economic hard times that have \\
\hline
with some family members not to exchange gifts. Lessons and & are another major expense. Fees for Kelcee’s gymnastics \\
\hline
Edge Magazine, a skateboarding publication. Before, the & has been just a bunch of loud-mouthed people trying to get \\
\hline
\end{tabular}
\end{table}

\textsuperscript{169} To view the concordance lines for sport: (1) Select “KWIC” on the COCA homepage; (2) Enter “SPORT_n” in the “Word/phrase” field; (3) Select “Sections” and select “MAGAZINE” and “NEWSPAPER”; (4) Click “Keyword in Context (KWIC).” \textit{Id.}

\textsuperscript{170} These concordance lines are selected for illustration purposes. A complete sample of concordance lines for sport can be viewed using the instructions in supra note 169.
Per my analysis of the 100 concordance lines examined, half featured uses of sport that explicitly referenced a *rule-based athletic competition* sense. Many of the other concordance lines strongly suggested a context in which rule-based athletic competition made the most sense, including references to the sports page, or sports betting or sports gambling, and there were a handful of references to the magazine *Sports Illustrated*. A dozen concordance lines had entirely unrelated uses of sport, including sport car or sport jacket. A small percentage of the uses of sport were ambiguous.

There were only three references to sports that might qualify as *physical recreation* rather than *rule-based competition*. Each of the following was described as a sport and could plausibly fall within either sense: *kayaking, skateboarding, and bungee jumping*. In the case of both kayaking and skateboarding, however, the expanded context shows that these concordance lines were taken from articles discussing the evolution of each sport from a recreational activity to a rule-based competition. Such uses of sport make clear that the word has a range of meanings that we have already seen listed about in a number of dictionaries. Out of one hundred concordance lines examined, only one has reference to an unambiguous use of the recreational activity use of sport. One article in the corpus discusses the rise of the recreational, non-competitive sport of bungee jumping.

Here, the *rule-based competition* sense overwhelmingly predominates, while the *recreational activity* sense is attested but comparatively rare. If the court understands the search for plain meaning to be a search for the most common sense of a word, then the court might conclude that the plain meaning of *sport* is *rule-based competition*. By contrast, the court understands contractual ambiguity to include any contractual term that is susceptible to more than one reasonable interpretation, then the court might conclude that because both senses of sport are attested, the word is ambiguous in the context of the contract. Under such circumstances, courts may choose to fall back on the *contra proferentem* canon, which holds that where the meaning of a contractual provision is found to be ambiguous, courts should “adopt
the meaning that is less favorable in its legal effect to the party who chose the words.**171**

4. **Corpus Linguistics and the Notion of a Core Meaning**

Neither of these approaches to plain meaning or ambiguity seems satisfactory. While courts sometimes frame the plain meaning determination in terms of mere frequency of competing senses—suggesting that the words of a contract should be given their “most common meaning.”**172** And while corpora can certainly provide evidence of the comparative frequency of different senses of a contractual term,**173** merely applying the most frequent sense regardless of context, timeframe, and speech community would be largely arbitrary, but possibly not wholly arbitrary.

Consider that many words have “at least one core sense or core meaning, to which a number of subsenses may be attached,” and that these “[c]ore meanings represent typical, central uses of the word in

---

171. KNIFFIN, supra note 1, § 24.27, at 282–83; see also RESTATEMENT (SECOND) OF CONTRACTS, supra note 17, § 206 (“In choosing among the reasonable meanings of a promise or agreement or a term thereof, that meaning is generally preferred which operates against the party who supplied the words or from whom a writing otherwise proceeds.”).

172. RLS ASSOCs., LLC, v. United Bank of Kuwait PLC, 380 F.3d 704, 710–11 (2d Cir. 2004) (“A court’s primary task in interpreting a contract is to determine intentions of the parties . . . as objectively manifested by the language of the contract . . . . [G]iving the words used in a contract their most common meaning generally serves this goal . . . .” (emphasis added)); Ragins v. Hosps. Ins. Co., Inc., 4 N.E.3d 941 (N.Y. 2013) (“Similarly, the parties evidently intended that ‘damages’ would retain its most common meaning . . . .” (emphasis added)); Prudential Kahler Realtors v. Schmitendorf, 673 N.W.2d 663, 666 (S.D. 2003) (“[W]here the parties to a contract cannot agree on the interpretation of a word in the contract, this Court will apply the ‘plain and ordinary meaning’ of the disputed term. The most common meaning of the word shown is ‘to cause or allow to be seen; display.’” (emphasis added)).

173. This is a common misreading of corpus-based approaches to legal interpretation—that they merely find and apply the most common sense of a word. See Carissa B. Hessick, Corpus Linguistics and the Criminal Law, 2017 B.Y.U. L. REV. 1503, 1505–14 (2017); Ethan J. Herenstein, The Faulty Frequency Hypothesis: Difficulties in Operationalizing Ordinary Meaning Through Corpus Linguistics, 70 STAN. L. REV. ONLINE 112, 117 (2017). Such an approach would be arbitrary and is not the approach advocated by proponents of corpus-based approaches to legal interpretation. See In re Baby E.Z., 266 P.3d 702, 726 (Utah 2011) (Lee, J., concurring) (“I share the view that we should not blindly attribute to every statutory term its most frequent meaning. Such an approach would be arbitrary and would lead to statutory incoherence. This is not the approach I have articulated, and not the one I have followed in my consideration of corpus linguistic data.”) (citation omitted); Stephen C. Mouritsen, The Dictionary is Not a Fortress: Definitional Fallacies and a Corpus-Based Approach to Plain Meaning, 2010 B.Y.U. L. REV. 1915, 1962 (2010) (“My contention is not that because [sense 2] is far more common than [sense 1], [the statute] ought to be interpreted with the [sense 2] meaning. Such a reading would be arbitrary. There are undoubtedly circumstances in which Congress employees the less frequent of two senses of a word.”).
question in modern standard English.”\textsuperscript{174} These core meanings can be revealed by examination of corpus data.\textsuperscript{175} Computational linguists have worked for decades to develop algorithms that will successfully choose the correct sense of a polysemous word (a word with more than one sense) for a given context. This process is called word-sense disambiguation.\textsuperscript{176} Algorithms that favor the most common sense of a word have a higher rate of accuracy and tend to choose the correct sense in as many as two-thirds or three-fourths of cases.\textsuperscript{177} If we can generalize these results to all questions of contractual ambiguity, then a heuristic that is right about two-thirds to three-fourths of the time cannot be said to be arbitrary. It is probabilistically non-arbitrary.

Professors Alan Schwartz and Robert E. Scott have argued “[g]reater accuracy is lower variance, and business parties commonly are indifferent to variance. Thus, courts that interpret contracts as typical parties prefer would be indifferent to variance as well, and sensitive only to the costs of administering their evidentiary standard.”\textsuperscript{178} If it is true that commercial parties elevate concerns about efficiency over concerns about accuracy and variance, then a heuristic that applies the most frequent sense (and does so in an automated fashion) may make sense.\textsuperscript{179} Such a heuristic would be more likely than not to be accurate and it would certainly reduce costs of administering the evidentiary standard. But it is not clear that commercial parties would desire an interpretive heuristic that would provide one wrong answer for every two or three right ones. Moreover, it is hard to imagine parties being able to draft contracts against the backdrop of a rule that would ignore less common, but perfectly well-attested senses of a word.

\textsuperscript{174} Oxford Dictionary of English xi (3d ed. 2010).
\textsuperscript{175} Id.
\textsuperscript{176} See Timothy Chklovskii & Rada Mihalcea, Exploiting Agreement and Disagreement of Human Annotators for Word Sense Disambiguation, in PROCEEDINGS OF THE CONFERENCE ON RECENT ADVANCES ON NATURAL LANGUAGE PROCESSING 3, 4 (2003).
\textsuperscript{177} See id.
\textsuperscript{178} See Alan Schwartz and Robert E. Scott, Contract Theory and the Limits of Contract Law, 113 Yale L.J. 541, 577 (2003). Schwartz and Scott define “variance” as, “a measure of how far an outcome can deviate from the mean of a distribution,”—if variance is large “the court makes in an actual case may well be very far from the correct answer,” and “when the variance is small, the court is likely to be close to the correct answer.” Id. at 555 n.65.
\textsuperscript{179} Indeed, algorithms that merely apply the most frequent sense of a word to a word-sense disambiguation problem have been shown to be accurate in 55% of the instances tested. See Keith Brown & Keith Allen, Concise Encyclopedia of Semantics 224 (2009).
5. **Corpus Linguistics and “Sport” in Context**

As set forth above, words may take on a particular, dominant sense when used as a particular part of speech, or with a particular inflection, or when surrounded by particular collocates, or when serving a particular function or serving a particular function or performing a particular role in a sentence. Linguistic corpora allow us to examine not merely how often a word is used in a particular way but how often a word is used in a particular syntactic, semantic, or pragmatic context, in the speech or writing of a given speech community, at a given timeframe. This is what is meant by the examination of context in this Article. While courts agree that context is an important consideration in legal interpretation, courts lack a shared, well-defined notion of what context actually means and lack methods for gathering evidence of a word’s use in a given context.  

The corpus can assist in the examination of the use of sport in more detailed context. The examples above already begin with the correct part of speech—looking at the use of sport the noun, rather than the less common verb to sport meaning “to mock or ridicule” or “to display or wear usually ostentatiously.” We can further limit the contextual evidence by matching inflection (looking only to plural uses of sports).

Different inflections of a given word—including the singular and plural forms of a noun—are often used in very different contexts. One way to demonstrate this is by comparing the most frequent collocates of the singular and plural forms. For example, the most frequent collocates of the singular and plural eye and eyes. Color terms tend to collocate only with the plural eyes while expressions that have to do with visualizing and evaluation appear with only the singular, like keep an eye out or turn a blind eye. A similar phenomenon can be observed in the most common collocates of sport and sports.

---

180. See supra notes 76–80. As such, it is not sufficient to say that “everyone agrees that context is important.” Carissa B. Hessick, *Corpus Linguistics and Criminal Law*, 2018 B.Y.U. L. REV. 1503, 1514 n.43 (2018). Even if it were true that everyone agreed that context is important, judges and lawyers lack any meaningful way to gather evidence of usage in context, and, as noted, lack a shared definition of what context actually means.


183. Id. at 119.

184. Id. at 120.

185. Id.
Table 3:
Collocates of “Sport” and “Sports”

<table>
<thead>
<tr>
<th>collocates of sport</th>
<th>collocates of sports</th>
</tr>
</thead>
<tbody>
<tr>
<td>utility, tires, medal, fish, vehicles, trucks, mode, fisherman, pickup, compact, polo, vehicle, blood, hunting, suspension, species, bikes, coats, wagon, sandals, boots, demands, requires, jumping, Mitsubishi</td>
<td>medicine, fox, editor, news, cbs, activities, authority, radio, network, complex, usa, section, bar, betting, music, memorabilia, Denver, entertainment, stadium, programs, car, arts, fantasy, camps</td>
</tr>
</tbody>
</table>

There is no overlap between the collocates of sport and the collocates of sports. This suggests that in order to determine the meaning of sports, evidence of the use of the plural form ought to be examined.

In addition to examining uses of sports, the corpus permits the examination of instances of sports with a similar syntactic relationship to the other syntactic components of the contractual provision and similar collocation (such as the words nearest to sports). For example, in the passage in question, sports is the head of a noun phrase proceeded by what is sometimes classified as a determiner, any. This may not seem like a great deal of context to take into account. But consider that the corpus contains 46,383 instances of the noun sport (13,750 singular, 32,633 plural), and only 84 instances of the noun phrase any sports.

An examination of these reveals that about two-thirds of these examples explicitly referenced athletic competition—a proportion that increases to about four-fifths when examples that strongly suggest athletic competition are included (for example, references to sports broadcasting, sports networks, etc.). A handful of the remaining instances of any sports are uncertain. But there are no clear instances of the physical recreation sense of sport. By taking into account a small number of contextual factors (inflection, syntax, collocation), we may be able to reduce the uncertainty about the meaning of sport in that context. This does not mean that it is impossible for the phrase any sports to have reference to physical recreation, only that such a reference is unattested.

in the corpus evidence and is, apparently, far less likely to occur than the alternative sense.

One of the challenges for examining usage in context in a corpus is that the greater the specificity of the search, the fewer examples appear in the corpus. As noted above, the switch from sport to any sports reduced the returns from approximately 45,000 to 84. Adding additional contextual information to a search containing any sports may not be possible in the COCA. For example, there are no examples of any sports proceeded by the verbs practice or participate in the COCA. To locate this usage, the corpus user would need a larger corpus. One such corpus in the News On the Web (NOW) Corpus, which is a corpus of approximately 7.4 billion words. And a search in the NOW Corpus reveals only twenty-four instances of sentences in which the verb participate or practice appears with any sport. Of these, only four appear to be ambiguous, while three of the four contain contrastive elements that suggest the athletic competition sense of sport (for example “sports or physical exercise” and “sports or exercise”). Here again, this does not mean that it is impossible that in the phrases practicing for any sports or participating in any sports, the physical recreation sense would occur. Rather, the corpus evidence simply shows that the physical recreation sense of sport does not appear to be attested in the semantic environment of practicing for or participating in. Not only does this evidence suggest that the physical recreation sense of sport is less likely to occur in this semantic environment, but it also suggests that by adding additional contextual parameters, we may be able to reduce our level of uncertainty about a contractual provision’s meaning.

The examination of language evidence from linguistic corpora certainly imposes costs beyond opening a dictionary, but it can provide detailed information about the usage of the operative terms, without


188. There are, essentially, two phrases of primary concern in this passage: practicing for any sports and participating in any sports. These two phrases are collapsed into one by a syntactic process referred to as conjunction reduction—in which “two syntactic heads of one phrase type are followed by syntactically similar complements.” Craig Hoffman, Parse the Sentence First: Curbing the Urge to Resort to the Dictionary when Interpreting legal Texts, 6 N.Y.U. J. LEGIS. & PUB. POL’y 401, 431 (2003). See also Pieter Seuren, Semantic Syntax, in 74 LANGUAGE 664–65 (1998).

189. Supra note 187. To view instances of the phrase any sport near the verbs participate or practice (1) Select “COLLOCATES” on the NOW Corpus homepage; (2) Enter “any sports” in the “Word/phrase” field and “participate_v” or “practice_v” in the “COLLOCATES” field; (3) Click “FIND COLLOCATES,” and (4) click “participate” or “practice.”
incuring the costs associated with discovering information about trade usage, contract negotiation, and course of dealing called for by the contextualist approach. The use of corpus evidence also raises important questions, including what types of contextual evidence ought to be considered in evaluating claims of contractual meaning, and how many instances of the usage of sports ought to be considered. And even the largest corpus will have limited evidence of rare or highly contextualized uses of a word.

Still, corpus evidence can give the judges and lawyers access to information about the context in which the words or phrases of a contract are used that cannot be accessed with a dictionary or through introspection. The corpus allows interpreters to take account of the specific inflection and context with which the operative term is used. It allows interpreters to examine language use from the timeframe in which the contract at issue was executed, and the possibly from the relevant speech community. The availability of this usage information can allow judges and lawyers to check their intuitions about the meaning of a legal text and to verify and falsify claims about usage and meaning.

B. Problem No. 2: The Plain Meaning of “Anticipated”

In UMB Bank, National Ass’n v. Airplanes Ltd.,190 A Lessor financed the acquisition of a fleet of aircraft by issuing $3.7 billion in notes. The notes were governed by an indenture and trust agreement, executed in 1996, that required the Lessor to maintain a reserve for required expenses. The agreement defined the “Required Expense Amount” as “(i) the amount of Expenses . . . due and payable on the Calculation Date relating to such Payment Date or reasonably anticipated to become due and payable before the end of the Interest Accrual Period beginning on such date.”191

A foreign Lessee challenged Lessor’s ability to collect on its lease. An international court found that Lessor’s attempts to collect were unlawful and ordered Lessor to pay Lessee a large cash judgment and punitive damages. Lessor appealed and an appellate court vacated the award against Lessor. Lessee filed a motion for reconsideration. Historically, such motions have an extremely low probability of being granted in the relevant jurisdiction. Nevertheless, during the pendency of the motion, the Lessor refused to make interest payments to the Noteholders because a required expense was “reasonably anticipated to

191. Id. (emphasis added).
become due and payable before the end of the Interest Accrual Period.\footnote{192}

Question: Does anticipated refer only to events that are likely to occur, or does it include events that are merely possible, but unlikely?

There seems little question that anticipated can refer to both expected and unlikely events. Both uses of anticipated are well attested,\footnote{193} and the court was presented with dictionary definitions that appeared to support both interpretations.\footnote{194} The court tacitly acknowledged this dilemma by refusing to arbitrarily select among two competing senses of anticipated and instead turning to precedent.\footnote{195}

1. Prior Case Law and the Plain Meaning of “Anticipated”

The UMB Court sought a “middle ground” by turning to what it characterized as the “assembled case law.”\footnote{196} But the three cases examined by the court in Problem No. 2 perpetuate the same dictionary-based errors in reasoning. For example, the court cites \textit{SN Sands Corp. v. City & County of San Francisco},\footnote{197} in which the California Court of Appeals held that “determination of . . . ‘anticipated expenditures’ must be based on more than the mere possibility of incurring an expenditure.”\footnote{198} But the \textit{SN Sands} Court relies on the same Merriam-
Webster’s definition cited by the Noteholders to reach this conclusion, and fails to acknowledge any other definitions or explain why the probable definition (one of six for anticipate) is the plain meaning.

The court in Problem No. 2 also cites Cyze v. Banta Corp.,\(^199\) in which “the Northern District of Illinois interpreted an employment contract that required an employer to provide severance benefits if an employee was terminated . . . ‘in anticipation of’ a change in control of the company.”\(^200\) The Cyze Court reasoned:

First, Webster’s defines anticipation as “intuitive preconception or a priori knowledge.”\(^201\) The next definition is “a prior action that takes into account, deals with, or prevents the action of another.” Finally, it offers “occurrence before the normal or expected time; the act of looking forward.”\(^202\)

The court fails to note that the Cyze Court is faced with a different interpretive question. The provision at issue in Cyze features the noun anticipation, while Problem No. 2 deals with the past participle form anticipated. As discussed below, words like anticipation and anticipate, which have similar derivational roots, can have markedly different uses when they are used as separate parts of speech. If instead of looking in the Webster’s Third for anticipation, the Cyze Court had been looking for definitions of anticipate, it would have found the same competing definitions cited by the Lessor and Noteholders.\(^203\)

Finally, the court cites the decision in Al-Kasid v. L-3 Communications Corp.,\(^204\) in which the Eastern District of Michigan reviews the same Merriam-Webster.com definitions reviewed above, and concludes that the sixth of these (“to look forward to as certain”) must apply.\(^205\)

\(^{199}\) No. 07 C 2357, 2009 WL 2905595 (N.D. Ill. Sept. 8, 2009).

\(^{200}\) Id. at *3.

\(^{201}\) Id. at *3 (citing Anticipation, WEBSTER’S THIRD INTERNATIONAL DICTIONARY UNABRIDGED 94 (1986) (noting that this is the second definition of “anticipation” in Webster’s).

\(^{202}\) Id. This is not the last definition of “anticipation” in Webster’s Third—it’s the fourth of six.

\(^{203}\) The Webster’s Third is the flagship, unabridged dictionary of the Merriam-Webster Company and all Merriam-Webster dictionaries (including the website definitions cited by the Lessor and Noteholders) are either heavily influenced by its definitions in the Webster’s Third or reproduce its definitions outright. See MCI Telecomm. Corp. v. AT&T, 512 U.S. 218, 226 n.2 (1994) (“The Webster’s New Collegiate Dictionaries, published by G. & C. Merriam Company of Springfield, Massachusetts, are essentially abridgments of that company’s Webster’s New International Dictionaries, and recite that they are based upon those lengthier works.”).


\(^{205}\) Id. at *7.
Each of the cases cited merely repeats the errors that the court was ostensibly attempting to avoid in the first place, and none of them have the force of precedent in the jurisdiction in which the relevant case was adjudicated (the United States District Court for the Southern District of New York). The UMB Court insists that a contract is ambiguous only if “the provisions in controversy are reasonably or fairly susceptible of different interpretations or may have two or more different meanings.”

But like the court in Problem No. 1, the court here ignores a well-attested alternative sense.

2. **Plain Meaning, Inflection, and Part of Speech**

There is another problem with both the court’s and the parties’ dictionary-based jurisprudence. Neither the parties, the court, nor the “assembled case law” cited by the court attempt to define the relevant word form at issue in the contract. In each case, the parties and the cases cited rely on dictionary definitions of the verb anticipate (with the exception of the Cyze case, which defines the noun-form anticipation), but none of them offers a definition of the past participle form “anticipated.” There is at least one good reason for this omission: none of the dictionaries cited by the parties (for example, Webster’s Third, McMillan’s Dictionary) contains a separate definition for anticipated.

It is not uncommon for dictionaries to make such omissions. Often when a word is formed by a process of derivational affixation (the adding of a prefix or suffix), only the base form is included in the dictionary. But the omission of a separate definition for anticipated does not necessarily mean that the word does not take on a specialized meaning in certain contexts. Indeed, different word forms of the same base or root word (often referred to as a lexeme) can take on particularized meanings. We can observe this phenomenon by conducting a simple comparison of the most common words to co-occur

---


207. Id. at 396.

208. The OED is not alone in this omission. The American Heritage Dictionary and Black’s Law Dictionary do not separately define anticipated.

209. P.G.J. van Sterkenburg, A Practical Guide to Lexicography 121 (2003) (“Restrictions on the inclusion and description of derivations that are regular/predictable in their form-content systems are most clearly visible in paper-based dictionaries. In order to keep the volume of a dictionary within reasonable limits, lexicographers have to use frequency criteria when selecting items for [inclusion in the dictionary].”).

with *anticipated* and *anticipated* (i.e., their collocates). The table below shows very little crossover between the most common words that co-occur *anticipate* and the past participle *anticipated*.

### Table 4: Collocates of “Anticipate” and “Anticipated”

<table>
<thead>
<tr>
<th>collocates of <em>anticipate</em></th>
<th>collocates of <em>anticipated</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>problems, failed, changes, ability, failure, consequences, fully, reaction, experts, investors, challenges, impossible, needing, reasonably, react, analysts, professionals, eagerly, arrival</td>
<td>than, highly, eagerly, longer, benefits, originally, faster, widely, hotly, demand, outcomes, consequences, guilt, actual, reasonably, arrival, stronger, films, anticipated, sooner</td>
</tr>
</tbody>
</table>

These collocates of anticipate and the past participle anticipated suggest that these different parts of speech may appear in very different contexts. This fact, at least, should make us skeptical of the practice of looking to definitions of anticipate or anticipation to find the plain meaning of anticipated.

### 3. “Anticipated” in the Corpus

Turning to evidence from the corpus, we can use collocation data to get a general sense of the use of anticipated. Because the Security Trust Agreement at issue in Problem No. 2 was executed in 1996, we will look at usage evidence from the COCA, which contains usage evidence dating from that timeframe. And unlike the dictionary citations referenced by the court, we will be able to limit our search to past participial forms of anticipated. The fifty most common collocates of anticipated are listed in the COCA as follows:

---

211. This data is based on data from the COCA, which will be discussed below. The COCA and other BYU corpora are available without a subscription. See COCA Website, supra note 93. To generate a list of collocates in the COCA you take the following steps: (1) Select “Collocates” on the COCA homepage; (2) Enter “anticipate” in the “Word/phrase” field; (3) Enter an asterisk “*” (a wildcard) in the “Collocates” field; (5) Select “Sort/Limit” and set the “Minimum” to “Mutual Info”; and (6) Click “Find collocates.” Repeat the same process using “anticipated.”

212. On the COCA website: (1) Select “Collocates”; (2) Enter “anticipated” in the “Word/phrase” field; (3) Enter an asterisk “*” (a wildcard) in the “Collocates” field; (4) Select “Sections” and select “MAGAZINE” and “NEWSPAPER”; (5) Select “Sort/Limit” and set the “Minimum” to “Mutual Info”; and (6) Click “Find collocates.” Id.
than, highly, eagerly, longer, benefits, originally, faster, widely, hotly, demand, outcomes, consequences, guilt, actual, reasonably, arrival, stronger, films, anticipated, sooner, revenues, earnings, losses, slower, correctly, shortfall, sequel, planners, impacts, retirements, debut, departure, announcement, tougher, surge, warming, designers, feared, matchup, quicker, organizers, scenarios, crowds, dreaded, shortages, reunion, volumes, emergence, accommodate, endanger.

These collocates of anticipated seem to suggest contexts in which anticipated events are regarded as more probable than not, though not necessarily certain, such as revenues, losses, expenses, restaurant openings, novel releases, etc. However, it is not necessarily clear that such revenues, losses, expenses, restaurant openings, or novel releases are more probable than not.

Turning to the KWIC evidence from the COCA, nearly all of the contexts in which anticipated appears in the corpus evidence refer to circumstances in which a particular outcome is expected and believed to be probable.213

213. To view the concordance lines for anticipated (1) Select “KWIC” on the COCA homepage; (2) Enter “anticipated” in the “Word/phrase” field; (3) Select “Sections” and select “MAGAZINE” and “NEWSPAPER”; (4) Click “Keyword in Context (KWIC).” Id.
Table 5: “Anticipated” in the COCA

<table>
<thead>
<tr>
<th>this is Singer’s much</th>
<th>anticipated</th>
<th>follow-up to his breakthrough [the Usual] Suspects</th>
</tr>
</thead>
<tbody>
<tr>
<td>with a look at the most</td>
<td>anticipated</td>
<td>performances in movies of the season</td>
</tr>
<tr>
<td>Markets have developed . . . Much more slowly than earlier</td>
<td>anticipated</td>
<td>worst-case scenario and explain how evacuations would be</td>
</tr>
<tr>
<td>commission has received 4,105 of these plans, which present</td>
<td>anticipated</td>
<td>benefits. A supportive drug substitution and</td>
</tr>
<tr>
<td>cannabis takers. The control initiative of MDT has not produced</td>
<td>anticipated</td>
<td>big book tour. Too much loss, says a character in</td>
</tr>
<tr>
<td>radio interviews that Wasserstein would be doing as part of her</td>
<td>anticipated</td>
<td>book of the fall season</td>
</tr>
<tr>
<td>want to have read it before you pick up the most</td>
<td>anticipated</td>
<td></td>
</tr>
</tbody>
</table>

My analysis of the evidence of from the COCA suggests that in its most common use, *anticipated* is used to describe events that are expected, believed to be probable, and looked forward to, such as the *anticipated* release of an album, or the anticipated sequel to a film. In nearly every other context found in the concordance evidence in the corpus, the word anticipated can easily be substituted for expected without a change in meaning, whereas *unlikely*, but possible cannot be substituted without changing the meaning. Though *expected* is the most common sense of *anticipated*, the *unlikely, but possible* sense is certainly attested, and anticipated is sometimes used to describe preparations for *possible*, but *unlikely* events. Thus, while it is clear that *anticipated* can be used to mean both expected or *unlikely, but possible*, a review of 100 instances of anticipated reveals only a single instance where the unlikely, but possible sense is attested.
If courts continue to define ambiguity as “susceptible to more than one reasonable interpretation,”\textsuperscript{214} then we can make a case for the contractual language in Problem No. 2 to be ambiguous. As with sport in Problem No. 1, there are, after all, two attested uses of anticipated and it is not clear what would make either reading more reasonable except that one sense appears to be much more commonly used that the other. With such a stark contrast in the use between the \textit{expected} and \textit{unlikely}, \textit{but possible} senses of anticipated, there is a reasonable case for the notion the overwhelming prevalence suggests a likely the plain or plain meaning.

While it is true that anticipated may be used to describe both expected and unlikely events, the usage evidence from the corpus demonstrates that where an unlikely event is intended, the surrounding context appears to make this sense explicitly clear by including words like “unlikely,” “unexpected,” or “worst case.” This suggests that the base or unmarked case for anticipated is the \textit{expected} sense, while the \textit{unlikely, but possible} sense must be specially indicated. Because there is no such indication contractual provision at issue, we can feel more confident that the base or unmarked sense of anticipated was intended.

C. \textit{Problem No. 3: The Plain Meaning of “Governable”}

In 1961, a Lessor separately negotiated mineral leases with two parties—Lessee A and Lessee B—for minerals mined in Forum X.\textsuperscript{215} Both Lessees separately agreed to pay Lessor no less than the highest royalty rate that each paid to any other lessor in Forum X. The original royalty rate was 5%. Lessor sought to ensure that Lessee A and Lessee B would pay the same royalty rates during the 100-year pendency of their respective leases. Lessee B, concerned that Lessor and Lessee A would unilaterally negotiate a higher royalty rate, agreed to pay the same interest rate as Lessee A, subject to the following exception:

\textit{Notwithstanding any of the provisions of this agreement to the contrary, the royalties to be paid to Lessor by Lessee B for minerals mined from the licensed premises during any quarter shall in no event be at rates less than royalty rates pursuant to which royalties are paid to Lessor by Lessee A during the same quarter, if and to the extent that the royalty rates paid to Lessor by Lessee A pursuant to their agreement are governable by the}

\textsuperscript{214} Hegel v. First Liberty Ins. Corp., 778 F.3d 1214, 1220 (11th Cir. 2015).
\textsuperscript{215} These facts are adapted, altered, and greatly simplified from the case of \textit{Ciner Wyoming, LLC v. Rock Springs Royalty Co.}, No. C-16-77-L (D. Ct. Wy 2016).
royalty rates paid by Lessee A pursuant to the provisions of any lease or license from the United States of America for the mining, removing, and disposing of minerals upon lands in Forum X.

In 2006, the National Heritage Areas Act reduced the federal royalty rate on the relevant mineral from 6% to 2%. Lessor continued to charge both Lessee A and Lessee B a 5% royalty rate. Subsequently, Lessor and Lessee A negotiated an increase in the royalty rate to 8%. Lessor began to charge Lessee B at the increased 8% royalty rate. Citing the exception clause above, Lessee B sued claiming that the agreement between Lessor and Lessee A was no longer “governable by” the federal rate (now 2%) and so Lessee B was no longer required to match the rate paid by Lessee A. Lessor argued that the plain meaning of “governable” is “capable of being governed.” While Lessee A was now paying a higher rate than the federal rate, its agreement with Lessor was still “capable of being governed” by the federal rate at the parties’ discretion or in the event the federal rate increased to 8%.

Question: Does governable always mean “capable of being governed” or can it also have a mandatory sense of “must be governed”?

It is hard to imagine that the Lessee would have intentionally entered into a lease agreement that leaves the royalty rate up to the discretion of the lessor and a third party. However, courts routinely state that the best evidence of the parties’ contractual intentions is the “plain language” of the contractual text. It is therefore possible that a court would read “governable” as “capable of being governed” and as granting discretion to the Lessor and Lessee A to set the royalty rate paid by Lessee B.

1. When Definitions Are Not Found in the Dictionary

Lessee B needs to demonstrate that governable can have a mandatory sense. The trouble is that no such definition of governable can be found in the dictionaries most commonly cited by courts. The Webster’s Third and the Oxford English Dictionary (two of the dictionaries most

---

216. Glen Banks, 28 N.Y. Prac., Contract Law § 9:3 (2018) (“The best evidence of what parties to a written agreement intended is what they said in their writing. The primary objective of contract interpretation is to give effect to the parties’ intent as manifested in the words they chose to use.”).

217. Indeed, the author was unable to locate any dictionary that defines governable with a mandatory sense.
often cited by American courts)\textsuperscript{218} records only the “capable of being governed” definition of governable.\textsuperscript{219} And, as discussed above, courts routinely turn to general-use dictionaries to search for the plain meaning of contractual terms.\textsuperscript{220} Indeed, courts have gone so far as to conclude that a sense of a word that is not recorded in a dictionary cannot be the word’s plain meaning.\textsuperscript{221}

But dictionaries are not complete repositories of every possible way in which a given word has ever been used. They often omit infrequent words or infrequent senses. “This gap between dictionaries and the lexicon results from a balance that every dictionary must strike: It must be comprehensive enough to be a useful reference but concise enough to be printed, shipped, and used. As such, many infrequent words [and word senses] are omitted.”\textsuperscript{222} Lexicographers draft dictionaries using impressionistic observations of incomplete usage evidence, so it is not uncommon for dictionaries to omit rare words or rare senses.\textsuperscript{223} In addition, dictionaries often omit separate definitions for words that are formed through a process derivational affixation—as when a new word is formed by adding a new word ending (\textit{to govern + -able}).\textsuperscript{224}

\begin{itemize}
\item \textsuperscript{218} Samuel A. Thumma & and Jeffrey L. Kirchmeier, \textit{The Lexicon Has Become a Fortress: The United States Supreme Court’s Use of Dictionaries}, 47 BUFF. L. REV. 227, 262–63 (1999).
\item \textsuperscript{221} Taniguchi v. Kan Pac. Saipan, Ltd., 566 U.S. 560, 569 (2012) (“Any definition of a word that is absent from many dictionaries . . . is hardly a common or ordinary meaning.”).
\item \textsuperscript{222} Michel et al., supra note 56, at 176; see also BIBER ET AL., supra note 55, at 36–41 (demonstrating that general-use dictionaries omit relevant senses).
\item \textsuperscript{223} BIBER ET AL., supra note 55, at 26 (“[C]itation slips [from which lexicographers draft dictionary entries] represent only those contexts that a human reader happens to notice . . . .”); LANDAU, supra note 55, at 104 (“Even very large citation files, because they are collected by a process of selection, cannot be used reliably for statistical studies of frequency because they are apt to be unrepresentative of the language as a whole.”).
\item \textsuperscript{224} Johan de Caluwe & Johan Taeldeman, \textit{Morphology in Dictionaries, in A PRACTICAL GUIDE TO LEXICOGRAPHY} 114, 121 (Piet van Sterkenburg ed., 2003) (“Restrictions on the inclusion and description of derivations that are regular/predictable in their form-content systems are most clearly visible in paper-based dictionaries. In order to keep the volume of a dictionary within reasonable limits, lexicographers have to use frequency criteria when selecting items for [inclusion in the dictionary].”).
\end{itemize}
The Lessor’s argument may also have some intuitive appeal not only because the “capable of being governed” sense is the only sense of governable recorded in any dictionary, but also because the suffix -able is very commonly added to transitive verbs to give them the meaning of capable of—teachable means “capable of being taught”; washable means “capable of being washed.”

2. The Mandatory Sense of -able

Even though no dictionary presents a mandatory sense of governable, there are numerous words formed with the suffix -able that take on a mandatory sense when used in a legal context. For example, rent that is “payable” on the first of the month is not “capable of being paid”; it must be paid on the first of the month. Income that is “reportable” on a tax filing is not “capable of being reported”; it must be reported. Taxpayers are not merely capable of paying taxes on their “taxable” income; they must pay their taxes. This use of words formed with -able is highlighted in the Cambridge Grammar of English, which makes clear that such words often have a mandatory sense:

The modal meaning [of -able] is generally like that of can. Sometimes, however, it is stronger, like that of must or will. Compare, for example, It is payable at any post office (“can be paid”) with It is payable by 15 June (“must be paid”) or The question isn’t answerable (“can’t be answered”) with The minister is answerable to Parliament (“has to answer”). Note also The deposit is refundable (“will be refunded”).

While the mandatory sense of governable is not reflected in any dictionary, the mandatory sense of -able is. The Oxford Dictionary of English defines -able as follows:

---

225. See Payable, AMERICAN HERITAGE DICTIONARY OF THE ENGLISH LANGUAGE (5th ed. 2012) (defining payable as “[r]equiring payment on a certain date; due,” or “[r]equiring payment to a particular person or entity” (emphasis added)).

226. Reportable, WEBSTER’S THIRD NEW INTERNATIONAL DICTIONARY 1925 (1971) (defining reportable as “2: required by law to be reported . . . .” (emphasis added)).

227. See Taxable, BLACK’S LAW DICTIONARY 1688 (10th ed. 2014) (defining taxable as “1. Subject to taxation”; Taxable, 17 OXFORD ENGLISH DICTIONARY 678 (2d. ed. 1989) (defining taxable as “1. Liable to be assessed (to a tax, impost, or charge); assessable . . . . 2. Liable to be taxed; subject to a tax or duty.”) (emphasis added).

228. RODNEY HUDDLESTON & GEOFFREY K. PULLUM, CAMBRIDGE GRAMMAR OF ENGLISH 1707 (2002).

229. Though published by the Oxford University Press, the Oxford Dictionary of English is unrelated to the OED. The former is a contemporary dictionary compiled with the aid of sophisticated electronic linguistic language database.
-able → suffix forming adjectives meaning: 1 able to be: calculable. 2 due to be: payable. 3 subject to: taxable. 4 relevant to or in accordance with: fashionable. 5 having the quality to: suitable | comfortable. 230

What this means is that a mandatory sense of governable is, at least, possible in a legal context, just like the mandatory senses of payable, taxable, and reportable are all possible senses in a legal context. The next question is whether this use of governable was attested in speech or writing at the time of the drafting of the agreement and whether it makes sense in the context of the agreement. 231

3. “Governable” in the Corpus

Turning to the corpus evidence, we are faced with a series of challenges. The first is frequency. Governable is an uncommon word, and the contested mandatory sense of governable is an even less common sense of an uncommon word. In the COCA, governable has eight collocates that co-occur with governable three or more times:

more, country, space, time, subjects, democracy, area, less

None of these collocates appear to have any reference to a mandatory sense of governable. This result is supported by an examination of the concordance lines of governable in the COCA. 232

---

230. Able, OXFORD DICTIONARY OF ENGLISH 4 (2010). The Oxford Dictionary of English also notes that contrary to popular perception, the morpheme -able and the adjective able, are etymologically unrelated, even though they have similar meanings. Id. ("[O]RIGIN from French –able or Latin -abilis; originally found in words only from these forms but later used to form adjectives directly from English verbs ending in -ate, e.g. educable from educate; subsequently used to form adjectives from verbs of all types (influenced by the unrelated word ABLE), e.g. bearable, saleable.").

231. WEBSTER’S THIRD NEW INTERNATIONAL DICTIONARY 17a (1971) ("The system of separating by numbers and letters reflects something of the semantic relationship between various senses of a word. It is only a lexical convenience. It does not evaluate senses or establish an enduring hierarchy of importance among them. The best sense is the one that most aptly fits the context of an actual genuine utterance.").

232. COCA Website, supra note 93. The concordance line search in COCA is executed as follows: (1) Select “KWIC” on the COCA homepage; (2) Enter “GOVERNABLE_j” in the “Word/phrase” field; (3) Select “Sections” and select “MAGAZINE” and “NEWSPAPER”; (4) Click “Keyword in Context (KWIC).” Id.
Table 6: “Governable” in the COCA

<table>
<thead>
<tr>
<th>Move toward making Washington</th>
<th>Governable</th>
<th>Again. And I suspect he’s going to actually be a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bosnia today is a mess, barely</td>
<td>Governable</td>
<td>and economically dead.</td>
</tr>
<tr>
<td>We have a country that is</td>
<td>Governable</td>
<td>and which people participate with joy and with freedom and</td>
</tr>
<tr>
<td>Look, unless Gaza becomes a relatively peaceful area, a</td>
<td>Governable</td>
<td>area, an area from which attacks do not emanate, then</td>
</tr>
<tr>
<td>Act expanded the common market and rendered many areas of policy</td>
<td>Governable</td>
<td>by a qualified majority, rather than allowing every member a</td>
</tr>
<tr>
<td>It’s critical for the United States to have a</td>
<td>Governable</td>
<td>country on our—on our southern border. And NAFTA,</td>
</tr>
<tr>
<td>Is happening on the ground</td>
<td>Governable</td>
<td>country, in your opinion, if Saddam Hussein was replaced?</td>
</tr>
<tr>
<td>Franken Would it be a more</td>
<td>Governable</td>
<td>country. So, I think we are putting too much on</td>
</tr>
<tr>
<td>Are the next step, so it is not exactly a</td>
<td>Governable</td>
<td>democracy can take root. However, since the prisoner’s</td>
</tr>
<tr>
<td>Prisoner’s dilemma can be broken, and a more</td>
<td>Governable</td>
<td>democracy and that is especially critical during the</td>
</tr>
<tr>
<td>Hamper the willingness to compromise that is needed in a</td>
<td>Governable</td>
<td></td>
</tr>
</tbody>
</table>

In the 560 million words that make up the COCA, governable occurs only forty-two times. Of these, most instances of governable fairly unambiguously refer to the notion of “capable to be governed.” 233

In addition, the contract at issue in Problem No. 3 was executed in 1961. This may militate in favor of looking to linguistic evidence from a

233. Id.
time period closer to the date of drafting. To accomplish this, we can attempt to look for concordance evidence in the Corpus of Historical American English (COHA). The COHA is “the largest structured corpus of historical English.” It contains “more than 400 million words of text from the 1810s-2000s (which makes it 50–100 times as large as other comparable historical corpora of English) and the corpus is balanced by genre decade by decade.” This allows the user to gather language evidence from the decade in which a text is drafted, going back approximately 200 years.

The challenge is that for each decade, the COHA has a comparatively small amount of data—only twenty million words. When dealing with low frequency senses of low frequency words, the COHA may not have sufficient data to reliably attest the ways in which a word was used. Indeed, in looking at evidence from the COHA for 1950 through 1970, governable appears in the corpus only three times—twice in the same paragraph from the same text:

Table 7:
“Governable” in the COHA—1950s through 1970s

| The slogan now is: Is New York | governable | ? And I agree with John Lindsay when he says,
| when he says, Of course it is | governable. | Because we haven’t any choice but to make it
| His admiration of his captain, Robert Gray, was | governable | ? none of Gray’s mates of record thought he was a

Here again, each of these instances of governable supports the “capable of being governed” sense.

As discussed above, some words formed with –able can have a mandatory sense—e.g., payable, reportable, and taxable. This suggests that a similar mandatory sense of governable is, at least, possible. If this is the case, we might expect to see this use of governable attested in a text. The apparent absence of the mandatory sense of governable in the

234. Id.
235. Id.
COCA and COHA could be taken as a strike against Lessee B’s argument in Problem No. 3.

The search could be further specified by including the phrase “governable by” followed by a noun or noun phrase, similar to the way in which the word is used in the contract. However, neither the relevant portions of the COCA nor the evidence from the COHA for the 1960s records a single instance of this usage of governable. The larger NOW Corpus records some twenty-four instances of “governable by” followed by a noun or noun phrase. In each case, the capable of being governed sense occurs.

The COCA and COHA are both corpora that largely reflect different genres of Standard Written American English, while the mandatory senses of words formed with -able tend to appear in legal contexts (a lease making rent payable on the first of the month, a tax code making certain income reportable). If we are looking for a specialized legal sense of governable, then a corpus of Standard Written American English may not be the right dataset. By contrast, if we turn to a legal database like Westlaw, we can find numerous instances of the mandatory, legal sense of governable, including a number of instances of this sense from the timeframe in which the contract at issue in Problem No. 3 was executed:

236. To confirm the lack of instances of the phrase “governable by” followed by a noun or noun phrase in the COCA, (1) Select “KWIC” on the COCA homepage; (2) Enter “governable by” in the “Word/phrase” field; (3) Select “Sections” and select “MAGAZINE” and “NEWSPAPER”; (4) Click “Keyword in Context (KWIC)”; To confirm the lack of instances of “governable by” followed by a noun or noun phrase in the COHA, (1) Select “KWIC” on the COHA homepage; (2) Enter “governable by” in the “Word/phrase” field; (3) Select “Sections” and select 1950, 1960, 1970; (4) Click “Keyword in Context (KWIC).”

237. To view instances of the phrase “governable by” followed by a noun or noun phrase, (1) Select “KWIC” on the NOW Corpus homepage; (2) Enter “governable by” in the “Word/phrase” field; (3) Click “Keyword in Context (KWIC).”
Table 8:  
“Governable” in Legal Context

<table>
<thead>
<tr>
<th>Description</th>
<th>Governable Status</th>
<th>Reason for Governable Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>[family] status that has long been recognized to be governable</td>
<td>governed</td>
<td>by the law of the domicile for purposes of marriage, divorce, [etc.](^{239})</td>
</tr>
<tr>
<td>In this diversity suit</td>
<td>governable</td>
<td>by the law of the State of New York(^{240})</td>
</tr>
<tr>
<td>the conduct . . . herein is solely governable</td>
<td>by the law of the Commonwealth of Pennsylvania(^{241})</td>
<td></td>
</tr>
<tr>
<td>The situation here encountered is governable</td>
<td>by the rule established in the case of <em>Hatch v. Bassett</em>(^{242})</td>
<td></td>
</tr>
<tr>
<td>their subject matter is constitutionally governable</td>
<td>by rules promulgated by this court(^{243})</td>
<td></td>
</tr>
<tr>
<td>the contract . . . is thus governable</td>
<td>by the statute of frauds, not the statutes of wills(^{244})</td>
<td></td>
</tr>
<tr>
<td>the transaction between the parties was governable</td>
<td>by the law applicable to a promissory note(^{245})</td>
<td></td>
</tr>
<tr>
<td>the contract . . . having . . . been made in Louisiana, and therefore governable</td>
<td>by the laws of that state(^{246})</td>
<td></td>
</tr>
</tbody>
</table>

---

\(^{238}\) These examples were located by searching for the phrase “governable by” in the “All State & Federal” search field on Westlaw.


\(^{240}\) Lee v. St. Joe Paper Co., 371 F.2d 797, 797 (2d Cir. 1967).


In the examples above, family status isn’t merely “capable of being governed” by the law of domicile; it must be governed by the law of domicile. A contract is not merely capable of being governed by the statute of frauds; it must be governed by the statute of frauds. The instances of governable in Table No. 8 demonstrate that even though the mandatory sense of governable is a less common (virtually unattested) sense of the word in Standard Written American English, the mandatory sense is well attested in legal contexts (just like the mandatory senses of other words formed with –able—payable, taxable, and reportable). Because it is virtually unattested in Standard Written American English, the mandatory sense of governable may not be the first sense that comes to mind when a judge examines the text of a contract. Indeed, the existence of the mandatory sense may not even occur to the judge. Nor could the judge find the mandatory sense by looking to a dictionary. This is one of the principal utilities of corpus linguistics: to reveal information about language that is not available through introspection or dictionaries.

Similar evidence for a more mandatory sense of governable can be found by searching through a database of contracts that are included with public company filings with the United States Securities and Exchange Commission (SEC). One such database is LawInsider.com. LawInsider.com is a database that collects “over 3,264,137 contracts [that are contained in] SEC document filings that have been catalogued and indexed by contract type, state/jurisdiction, company, signees and filing date.” A search for governable in these documents reveals only a handful of uses, some of which appear to suggest a mandatory, legal sense. For example, one financing agreement contains an express “Affirmative Covenant” which states: “7.8 Compliance with Laws: Continue at all times to comply with all laws, ordinances, regulations, requirements and Governable Regulations of any Governmental Authority relating to Borrower’s business, property or affairs.” Though capitalized, “Governable Regulations” is not a defined term. However, the context makes clear that compliance with these regulations is not a matter of discretion. Similarly, a loan agreement contains a provision for payment of taxes, which states: “(b) Payment of Other

Taxes by the Borrowers. Each applicable Borrower shall timely pay any to the relevant Governable Authority in accordance with applicable law, or at the option of the Administrative Agent, timely reimburse it for, Other Taxes. 250 Here again, “Governable Authority” is capitalized, but is not a defined term in the contract. But there is no suggestion that the payment of taxes or the identity of the taxing authority are matters of discretion under the agreement. These examples suggest that there exists a specialized, legal use of governable that has a mandatory, rather than a discretionary, sense.

Both Westlaw and LawInsider have a number of limitations as linguistic corpora. Neither allows a user to easily gather information about usage over time, or information about collocation, nor does either allow for the concordance of usage evidence. But in the absence of a principled linguistic corpus of American legal discourse, Westlaw and LawInsider can nevertheless be useful for gathering evidence about the legal usage of a word, especially when that evidence can be compared against usage evidence from a corpus of Standard Written American English like the COCA or COHA.

Problem No. 3 illustrates the importance of matching the right corpus to the right question. The evidence suggested that governable may have a specialized, mandatory sense that appears only in legal contexts. If that is the case, then we would expect that a specialized legal usage would be attested in a corpus of legal texts. To be sure, Westlaw and LawInsider leave much to be desired as a corpus. Neither allows its user to see frequency information about the use of a given word, including frequency distribution through time. The texts in each have not been annotated with linguistic metadata, like part-of-speech tagging, which limits its utility as a linguistic corpus. Neither database offers much of the typical functionality of a linguistic corpus, like the ability to search for collocates or review concordance lines. Nor are they designed to be representative of the speech or writing of a given speech community (lawyers) or to present a representative sample of legal language use (including samples of legal language usage in lawyer speech, or text or email correspondence). Still, both Westlaw and LawInsider.com share a few features with a linguistic corpus. They are very large collections of easily searchable texts and can be useful tools to locate and examine the

specialized legal use of a given word or phrase. That is the task that they perform here.

In this case, the contract makes the most economic sense if governable takes on a mandatory meaning. But this sense of governable does not appear in any dictionary and does not appear to occur in common usage. Using an electronic collection of legal texts, we are able to demonstrate that an uncommon sense of a word in fact exists and is well-attested in legal texts. Once we have demonstrated that the contested sense of governable exists, we are better positioned to argue that this previously unknown sense best fits the context of the contract and ought to control in the court’s interpretation of the contract.

D. Problem No. 4: The Plain Meaning of “Structural Damage”

A homeowner’s insurance policy provides coverage for structural damage to a home resulting from a sinkhole.\(^\text{251}\) The home in question experienced “widespread, minor cracking to both the exterior and interior of the home” as the result of a sinkhole, but did not suffer any damage that would “impede[] the structural components from supporting the loads that they are intended to support.”\(^\text{252}\) The policy does not define structural damage, and there is no relevant statutory definition. Prior decisions interpreting similar language and applying relevant state law have reached opposite conclusions about the meaning of structural damage.\(^\text{253}\)

Question: Does structural damage mean any “physical damage to a building,” or only “‘damage that impairs the structural integrity [or load bearing capacity] of the building’”?\(^\text{254}\)

The court faced with Problem No. 4 appealed to the “plain meaning” of structural damage,\(^\text{255}\) reasoning that “[i]f the relevant policy language

\(^{251}\) This problem is based on Hegel v. First Liberty Ins. Corp., 778 F.3d 1214, 1216–22 (11th Cir. 2015).

\(^{252}\) Id. at 1218, 1222 (citing Fla. Stat. § 627.706(2)(c) (2005)).


\(^{254}\) Hegel, 778 F.3d at 1219, 1221.

\(^{255}\) Id. at 1220 (quoting State Farm Fire & Cas. Co. v. Steinberg, 393 F.3d 1226, 1230 (11th Cir. 2004) (“Florida courts start with the plain language of the policy as bargained for by the parties. If that language is unambiguous, it governs. If the relevant policy language is susceptible to more
is susceptible to more than one reasonable interpretation, one providing coverage and the other limiting coverage, the insurance policy is considered ambiguous, and must be interpreted liberally in favor of the insured and strictly against the drafter who prepared the policy.”

Acknowledging that “differing interpretations of the same provision [are] evidence of ambiguity,” the court found that “no genuine ambiguity” existed in the interpretation of the phrase structural damage. The court insisted that the words structural and damage “cannot be viewed in isolation” and that the court must “construe an insurance contract in its entirety, striving to give every provision meaning and effect.”

The court further appealed to a familiar canon of contract interpretation known as the canon against surplusage, which holds that courts “must read the contract to give meaning to each and every word it contains,” and “avoid treating a word as redundant or mere surplusage ‘if any meaning, reasonable and consistent with other parts, can be given to it.’” With this in mind, the court reasoned that because “structural damage” was necessary for the homeowner to recover, equating “physical damage” with “structural damage” would read “structural damages” out of the policy. Indeed, the court noted that that “at oral argument, [the homeowner’s] counsel . . . was unable to identify what type of damage to the property would not be ‘structural’ under their preferred definition.”

1. “Structural Damage” in the Dictionary

Even though the court insisted that structural and damage cannot be viewed in isolation, the court looked to dictionary definitions of structural and structure only in order to determine the plain meaning of

256. Id. (quoting State Farm Fire and Cas. Co. v. Steinberg, 393 F.3d 1226, 1230 (11th Cir.2004)).

257. Id. The court noted that “at oral argument, counsel . . . was unable to identify what type of damage to the property would not be structural under their preferred definition.” Id. at 1221.

258. Id.

259. Id. (emphasis added) (quoting Equity Lifestyle Props., Inc. v. Fla. Mowing & Landscape Serv., Inc., 556 F.3d 1232, 1242 (11th Cir. 2009)).

260. Id.

261. Id.
structural damages. The court noted: "‘Structural’ is an adjective, defined in the Oxford English Dictionary as ‘[f]orming a necessary part of the structure of a building or other construction, as distinct from its decoration or fittings.’ The noun ‘structure,’ on the other hand, is simply a synonym for a building." The court cites only the Oxford English Dictionary when seeking to find the plain meaning of structural damage, and cites precedent suggesting that citation to a dictionary is a best practice when interpreting contracts. Consistent with prior instances of the dictionary/intuition approach, the court does not explain why the court references only a single definition from the Oxford English Dictionary, rather than numerous competing senses of structural listed in other common, unabridged dictionaries, which make clear that both senses of structural relevant here are attested. For example, Webster’s Third New International Dictionary defines structural as both “1 a : of or relating to structure or a structure [and] . . . b : of or relating to the load-bearing members or scheme of a building as opposed to the screening or ornamental elements.” Nor does the court account for the fact that both senses of structural damage are well attested in common usage.

262. Id. Interestingly, the court fails to note that the Oxford English Dictionary does, in fact, separately define structural damage, as “damage to the structure or overall stability of something; esp. damage to a building, bridge, etc., sufficient to impair its structural integrity.” Oxford Univ. Press, Structural, OXFORD ENG. DICTIONARY, http://www.oed.com/view/Entry/191887 [https://perma.cc/UJM3-6N9Y]. Notice, however, that this definition is disjunctive and contains both the notion “damage to the structure,” as well as damage “sufficient to impair its structural integrity.” As noted below, both of these senses are well attested in common usage.

263. Hegel, 778 F.3d at 1221.

264. Id.

265. See Garcia v. Fed. Ins. Co., 969 So.2d 288, 291–92 (Fla. 2007) (stating that insurance contracts are construed according to their plain meaning and that, when doing so, courts may consult dictionary definitions).

266. Structural, WEBSTER’S THIRD NEW INTERNATIONAL DICTIONARY 2266 (1975).

267. See Jerry Thomas, Historic Bar Jolted in 2-Car Crash: Green Door Stays Open Despite Extensive Damage, CHI. TRIB., Feb. 3, 1998, § 2, at 5 (“The building, which was built in 1872, one year after the Great Chicago Fire, suffered structural damage to the exterior, but business continued as usual . . . .”); Kristen Reed, et al., Car Hits Office; Drug Charges Filed, ORLANDO SENTINEL: CENTRAL FLORIDA, Dec. 23, 2006, at B3 (“The car plowed into a building across the street from the old downtown Sanford Courthouse mid-morning, causing minor structural damage to its exterior.”); Cracks a Concern at Middle School, ELMIRA STAR-GAZETTE, Aug. 27, 2008, at 1 (“The district, facing ‘serious structural damage’ at the façade at Ernie Davis [Middle School], is putting up netting in case bricks fall.”). These examples were located using an online dictionary database, NEWSPAPERS.COM (last visited April 14, 2018).
2. “Structural Damage” and the “Pet Fish” Problem

Even if the court had looked to definitions of both structural and damage, the underlying assumption that the meaning of structural damage is merely the sum of its parts is not defensible. Sometimes the meaning of words and phrases cannot correctly be understood as the sum of their constituent parts. Instead, the meaning of an expression must be understood as crossing the boundary of what we traditionally think of as a word (a string of letters separated by a space, or collection of sounds separated by a pause). While the minimum standalone unit of meaning of an utterance will often be what we traditionally think of as a word. This is the problem that has been characterized by Professor Larry Solan as the “pet fish” phenomenon:

The “pet fish” phenomenon occurs when two words combine to produce a meaning that is not the mechanical composition of the two words separately . . . Thus, the prototypical “pet” is a dog or a cat; the prototypical “fish” is a salmon or cod—but the prototypical “pet fish” is neither a dog nor a salmon, but is clearly the goldfish.

Linguist John Sinclair referred to a similar feature of language when he discussed the “idiom principle,” the notion that “a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices [in the creation of an utterance], even though they might appear analysable into segments.” A speaker cannot break down many linguistic utterances into their constituent parts. These include idiomatic expressions (like, pull someone’s leg, paint the town red, kick the bucket), phrasal verbs (like, get down, make out, carry on about), and collocates or words that commonly occur in the same environment (as in the semantic prosody example above). Such multi-word strings are sometimes referred to as “lexical items,” a minimal standalone unit of meaning that includes the traditional notion of a word, but also includes many multi-word utterances.

---


269. See id.


272. Among these lexical items are words that co-occur so often that they take on a specialized meaning when they appear together in certain contexts. This is phenomenon is referred to as “collocation,” which means “a frequent co-occurrence of words [that] does not have a profound effect on the individual meanings of the words, but there is usually at least a slight effect on the
be able to account for the meaning of phrases like structural damage without accounting for the meaning of the phrase as a whole.273

3. Co-selection, Delexicalization, and the Canon against Surplusage

The court in Problem No. 4 insisted that equating any damage or “physical damage” with “structural damage” would read “structural” out of the policy. In so doing, the court invoked the so-called canon against surplusage (or superfluity), which requires that courts “read the contract to give meaning to each and every word it contains,” and “avoid treating a word as redundant or mere surplusage ‘if any meaning, reasonable and consistent with other parts, can be given to it.’”274 Under the surplusage canon “[w]ords are never to be construed as meaningless if they can be made effective by any reasonable construction of the contract,” and the court should “strive to give meaning to every sentence, clause and word.”275 The surplusage canon is classified as a linguistic canon of interpretation,276 which means, ostensibly, that its legitimacy may “stand or fall by [its] accuracy in reflecting relevant linguistic practices.”277 But, as discussed below, evidence of language use from linguistic corpora gives us reason to be skeptical that the canon against surplusage is a generalizable and descriptive statement about the way language actually works.

There is strong evidence that rather than choosing each word carefully and independently to convey an intended meaning, the choice of a given word often conditions the choice of the next word.278 This process is called co-selection, and there is strong evidence that co-selection is a fundamental part of word choice in natural language.279

A necessary corollary of co-selection is delexicalization—a loss of independent meaning that occurs when a word is subconsciously

meaning, if only to select or confirm the meaning appropriate to the collocation, which may not be the most common meaning.” SINCLAIR, TRUST THE TEXT, supra note 86, at 28.

273. ALAN CRUSE, MEANING IN LANGUAGE 74 (2010).


277. Id. at 1084.

278. SINCLAIR, TRUST THE TEXT, supra note 86, at 19.

279. Id.
selected as a collocate of another word. As Professor John Sinclair observes:

The meaning of words chosen together is different from their independent meanings. They are at least partly delexicalized. This is the necessary correlate of co-selection. If you know that selections are not independent, and that one selection depends on another, then there must be a result and effect on the meaning which in each individual choice is a delexicalization of one kind or another. It will not have its independent meaning in the full if it is only part of a choice involving one or more words . . . . [There] is a strong tendency to delexicalization in the normal phraseology of modern English.

Consider the following pairings of adjectives and nouns, which were found in an examination of corpus evidence. Normally we “understand in grammar that adjectives add something to the noun, or restrict the noun, or add some features to it.” While this is “no doubt true in some cases,” there are many instances where the adjective appears to be the result of co-selection and is delexicalized, in part or in whole:

**Table 9:**

**Delexicalization of Adjectives**

<table>
<thead>
<tr>
<th>The officers stood in close [physical] proximity to the victims.</th>
</tr>
</thead>
<tbody>
<tr>
<td>The chemists conducted a [scientific] experiment.</td>
</tr>
<tr>
<td>The restaurant was at [full] capacity, so they wouldn’t seat us.</td>
</tr>
</tbody>
</table>

Each of the sentences in Table No. 9 above can be read with or without the bracketed term without changing the meaning of the sentence. Moreover, with or without the bracketed term, each of the sentences above is a perfectly well-formed, grammatical English sentence. Still, native English speakers will commonly include such delexicalized adjectives in their communication and will do so subconsciously. This does not mean that adjectives like “physical,”

---

280. *Id.* at 20.
281. *Id.*
282. *Id.*
283. *Id.*
284. At this point, it is possible that some readers will protest that careful writing of the kind they have been trained to perform requires the careful (and highly circumscribed) use of adjectives.
“scientific,” and “full” are meaningless. It means that in certain contexts, through the process of co-selection, such adjectives are delexicalized and add little or no communicative context to an utterance.

With this understanding, we can reexamine the Problem No. 4 court’s reliance on the surplusage canon and the claim that reading “structural damage” as merely “physical damage” would read “structural” out of the contract.285 Consider the following uses of “structural” based on words that are found to commonly co-occur with “structural” in the COCA.

<table>
<thead>
<tr>
<th>Delexicalization of “Structural”</th>
</tr>
</thead>
<tbody>
<tr>
<td>The earthquake impaired the [structural] integrity of the building.</td>
</tr>
<tr>
<td>The contractor lacks sufficient [structural] steel to finish the building.</td>
</tr>
<tr>
<td>The building collapsed due to the failure of [structural] supports.</td>
</tr>
</tbody>
</table>

Here again, each sentence in Table No. 10 above can be read with or without “structural” without changing the meaning of the sentence. But with or without “structural,” each sentence is nevertheless a perfectly well-formed and grammatical English sentence. The reader wouldn’t notice anything unusual with or without “structural.” Native English speakers commonly include delexicalized adjectives like “structural” in sentences because co-selection is part of the human language faculty, and “structural” tends to systematically co-select with words like “integrity,” “engineer,” “steel,” etc. This does not mean that the word “structural” is meaningless. It is easy to think of examples in which a “structural” engineer may be contrasted with other types of engineers, or “structural” steel may be contrasted with other types of steel. The question is whether there are contexts in which structural does not add independent semantic content to a sentence, other than to highlight or

---

underline the noun it modifies. That is, there may be contexts in which “structural” will be paired with some of the word’s common collocates in ordinary speech because of subconscious co-selection, even though “structural” does not add independent meaning to an utterance.

With this in mind, consider the Court in Hegel v. Liberty Insurance Corp.’s complaint that “at oral argument, counsel . . . was unable to identify what type of damage to the property would not be ‘structural’ under their preferred definition.” In the examples above, what type of steel could possibly be implicated, but “structural” steel? What type of building integrity could be impaired by an earthquake if not its “structural” integrity? The court is insisting that “structural” must have some independent meaning, but there is plenty of evidence to suggest that “structural” may simply appear in the contract as a result of subconscious co-selection and, consequently, “structural” is delexicalized and contributes nothing to the meaning of the provision.

4. “Structural Damage” in the Corpus

We can confirm our reading of structural damage by examining the collocation data for structural damage in the COCA. The fifty most common collocates of structural damage are listed as follows:

damage, no, any, major, caused, cause, homes, causing, building, severe, suffered, revealed, serious, significant, reported, sustained, reports, buildings, injuries, brain, trees, due, mri, showed, extensive, 90, heavy, sustain, knee, flooding, suffer, widespread, further, progression, tornado, earthquake, bone, roof, assessment, engineers, improving, suffering, causes, appear, avoid, termites, inhibiting, bruise, hoboken, minimal

These data suggest two broad uses for structural damage, one related to damage to the human body (MRI, injuries, brain, knee, bone, bruise), and the other related to manmade structures (homes, building). The harm to the human body uses of structural damage are not relevant to the question in the case. When we examine structural damage in context, these will be excluded. And at least some of the collocates suggest events of the sort of severity that could harm the load bearing capacity of a building (tornado, earthquake, termites), but for the most part, these

286. 778 F.3d 1214, 1221 (11th Cir. 2015).
287. Id.
288. On the COCA website: (1) Select “Collocates”; (2) Enter “structural damage” in the “Word/phrase” field; (3) Enter an asterisk “*” (a wildcard) in the “Collocates” field; (4) Select “Sections” and select “MAGAZINE” and “NEWSPAPER”; (5) Select “Sort/Limit” and set the “Minimum” to “Mutual Info”; and (6) Click “Find collocates.” COCA Website, supra note 93.
collocates do not indicate whether a “load bearing” or “of or related to structure” sense of structural is implicated. A tornado or earthquake may harm the load bearing integrity of a building or simply cause superficial damage. At the very least, we would need more information to draw conclusions about the meaning of structural damage in the context of Problem No. 4.

We can examine more closely the distribution of senses of structural damage using the concordance feature of the COCA. Using the COCA, we can generate a randomized collection of sample concordance lines showing the variety of contexts in which structural damage occurs. As noted above, the corpus allows us to look at the use of structural damage together rather than looking at both terms in isolation. We can target the time period in which the contract was executed and we can use the corpus evidence to determine the frequency of the different senses of structural damage.

In evaluating the concordance lines, we are looking for the uses of structural damage most similar to those in the contract at issue in Problem No. 4. The concordance evidence reveals four general categories of things that experience structural damage: buildings, bodies, roads, and automobiles. These different contexts are sometimes reflected in the sense division (or macrostructure) of dictionaries. We will exclude from the analysis uses of the phrase that appear in separate contexts and focus only on those contexts that are related to the contractual language, for example, the building contexts.

---

289. The concordance line search in COCA is executed as follows: (1) Select “KWIC” on the COCA homepage; (2) Enter “structural damage” in the “Word/phrase” field; (3) Select “Sections” and select “MAGAZINE” and “NEWSPAPER”; (4) Click “Keyword in Context (KWIC).” Id.
Table 11:
“Structural Damages” in the COCA—1990 to the Present

<table>
<thead>
<tr>
<th></th>
<th>Structural Damage</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>life-threatening and historic storm surge, as well as</td>
<td>structural damage</td>
<td>that could make buildings and homes uninhabitable</td>
</tr>
<tr>
<td>it’s built to sustain no</td>
<td>structural damage</td>
<td>during the most powerful kind of earthquake likely</td>
</tr>
<tr>
<td>Power loss and</td>
<td>structural damage</td>
<td>from the storm are likely to affect other businesses,</td>
</tr>
<tr>
<td>they elicited complaints about</td>
<td>structural damage</td>
<td>from their upstairs neighbors. The real difficulty</td>
</tr>
<tr>
<td>his $48,000 deductible—the amount of</td>
<td>structural damage</td>
<td>his home must suffer before the insurance company</td>
</tr>
<tr>
<td>to wave-battered Atlantic City, the storm caused extensive</td>
<td>structural damage</td>
<td>in cities and towns along the coast of New Jersey.</td>
</tr>
<tr>
<td>the campuses of the University of Houston but no apparent</td>
<td>structural damage</td>
<td>to the buildings, Chancellor Renu Khator told CNBC on</td>
</tr>
</tbody>
</table>

The concordance evidence shows that both of the competing uses of structural damage are attested. There are uses of structural damage that reflect the harmful to load-bearing integrity sense. And there are uses of structural damage that reflect the mere physical damage to a structure sense.

In a review of 100 randomized concordance lines from the COCA, fifty-eight appeared in a building related context. There were twelve instances in which the context clearly indicated the harm to structural integrity use of structural damage and only two instances in which any harm to a structure was indicated. There were forty-four instances in which there was no way to tell whether either sense was intended.

This raises an important question in terms of the theorization of contractual plain meaning. In the few cases in which the context makes clear that one or the other of the senses should apply the harm to load-bearing integrity sense is the most common. And, to the extent that we agree that there is a frequency component to plain meaning, we might take this as evidence that this structural integrity meaning is the plain meaning. However, in most of the instances in which structural damage to buildings appears, there is no way to distinguish between the two
relevant senses. This is true even in the natural disaster context, where wind, flood, or fire may be said to cause *structural damage*, but without more, we cannot tell whether that damage harmed the structural integrity of a building or merely caused superficial damage.

This suggests that the phrase *structural damage* is underspecified with respect to the extent of the damage implied. Under-specification is the “omission of information from linguistic descriptions to capture several alternative realisations of a linguistic phenomenon in one single representation.”290 Under-specification is a common cause of linguistic ambiguity.291 In such cases, in which we are able to distinguish among senses, it is only because the sentence provides us with additional information. This would mean that the phrase *structural damage* is necessarily (linguistically) ambiguous with respect to the extent of damage. This would also mean that the court’s conclusion that there was “no genuine ambiguity” in the agreement was incorrect.292 This evidence also suggests that within the phrase *structural damage*, the adjective structural is the product of co-selection and delexicalization.

This is an important contribution of corpus linguistics when applied to questions of legal interpretation. Corpus linguistics does not simply mechanically identify plain meaning. Instead, corpus-based language evidence can give content to the idea of legal ambiguity. How to address ambiguity in a legal text is a legal question that may be outside of the domain of linguistics. But the identification of meaning and ambiguity can be greatly facilitated with language evidence.

It is important to distinguish the identification ambiguity in Problem No. 4 from similar findings of ambiguity in the experiments on bias in judging ambiguity referenced above.293 The referenced study described the identification of ambiguity based on intuitive judgments about texts, and not based on objective evidence of usage.294 In such circumstances, “[i]f one person says that both proposed readings of a [text] seem plausible, and a colleague disagrees, finding one reading too strained, what is there to do about it but for each to stamp his foot?”295 By contrast, a corpus-based identification of contractual ambiguity is based

291. See id.
294. See id. at 257–60.
295. Id. at 276.
on an evaluation of language evidence—an evaluation that can be replicated and falsified.

V. CONTRIBUTIONS AND POTENTIAL CHALLENGES OF THE CORPUS APPROACH

The problems above illustrate some of the potential contributions that linguistic corpora can make to addressing problems of contract interpretation meaning. Evidence from corpora can be used to test intuitions about the meaning of contractual terms. The corpus can assist in creating a complete picture of the way in which a word is used in a given context by a given speech community and at a given time. The corpus can take account of the syntactic, semantic, and even pragmatic context of an utterance. Unlike the dictionary, the corpus allows its users to examine utterances at the word, phrase (or idiom), or sentence level, and the corpus can take account of multi-word lexical items. Corpus searches can be tailored to the timeframe in which a given text was drafted. And corpus evidence can take account of differences in genre, dialect, register, and speech community. Moreover, the evidence relied upon by a corpus user can be made available to subsequent users to verify (or falsify) prior results. None of these tasks can be performed by a jurist’s unaided intuition or by a dictionary.

Relying on corpus evidence to evaluate the way in which words are used in a contract can remove questions of plain meaning from the black box of a judge’s intuition and compel judges and lawyers to address problems of plain meaning and ambiguity in more concrete, evidence-based terms. Moreover, corpus evidence may be used to reexamine (and possibly reformulate) existing canons of contractual interpretation.

Even with all of these potential contributions of corpus linguistics to contract interpretation, there are still a number of questions with which the law must grapple before the corpus approach can be fully integrated into contract interpretation. These include the relevant reference timeframe for the interpretation of a contract, the relevant register or speech community for the interpretation of a contract, and the appropriate method for validating corpus results. Each of these are addressed below.

A. The Reference Timeframe for Contract Interpretation

The trouble is that in many cases it is not at all clear when a given provision of a contract was initially drafted or whether the linguistic
norms at the time of the contract’s drafting have any bearing on the interpretation of specific provisions. While it is true that the “traditional model of contract interpretation focuses on the ‘meeting of the minds,’” it is likewise true that “[n]ot all contracts and contract terms are individually negotiated.” Specifically, “boilerplate terms are not the product of any actual meeting of the minds,” and are “not representations of the specific intent of the parties to the transaction.”

There are a variety of ways to approach this problem using linguistic corpora. The first is to assume that, regardless of the timeframe in which a given provision first appeared, the language of a contract must be presumed to reflect the intention of the parties at the time it was executed. This is, of course, a legal fiction, but it is not an unprecedented one. Imposing such a legal fiction on the parties to a contract may counsel extra vigilance in contract drafting and negotiation and may lead to more dynamic contract drafting. Existing corpora can be brought to bear to help identify the relevant linguistic conventions. Moreover, as noted above, specialized, industry-specific general corpora or industry-specific contract corpora may, in many cases, be preferable for determining the linguistic conventions of a particular industry.

A second approach would be to identify whether or not a provision at issue in a given case had been separately negotiated, and if so, to interpret the provision consistently with the linguistic norms prevailing at the time of execution. If the provision had not been separately negotiated, the interpreter would then determine when the provision in question first appeared in contracts in the relevant industry and allow the linguistic norms prevailing at the time when the provision was first drafted to inform the interpretation. Such an approach would require the creation of a collection of historical contracts corpora that would track the development and evolution of contracts. While this may sound like a significant undertaking, at least some relevant collections of historical contracts are already in existence. These could be uploaded into a corpus interface without significant challenges. The question, then, is not

298. Id. at 1160.
299. See Tal Kastner, The Persisting Ideal of Agreement in an Age of Boilerplate, 35 L. & SOC. INQUIRY 793, 803 (2010) (“[T]he doctrinal presumption of enforceability of form language reflects an acknowledgement of the limits of knowledge and choice in assent, as well as the de facto operation of boilerplate in establishing the substance of a contractual arrangement.”).
whether it is necessarily possible to interpret the non-negotiated, boilerplate terms of a contract according to the linguistic conventions prevailing at the time when they were drafted, but whether it is appropriate to do so. If—as courts routinely insist—the objective of contract interpretation is to give effect to the parties’ intent, it is difficult to meaningfully ascribe intent to parties that did not draft or negotiate the language in question and who may never have bothered to read it.

Both approaches assume that the interpretation of boilerplate provisions has anything to do with identifying the linguistic conventions of a given speech community. Professor Michelle Boardman has observed that “[b]oilerplate [language] that has repeatedly been construed by courts will take on a set, common meaning, but one that may not be easily understood by reading the language itself.”

Thus, “the meaning ascribed to the language by an innocent first reader will differ markedly from the meaning the language is given in court, the meaning upon which drafters rely.” This circumstance may create an “illusion of understanding,” one that may only be remedied by “knowledge of the subsequent case law and regulatory actions.”

In such a circumstance, courts commonly find themselves construing boilerplate (determining its legal effect) rather than interpreting it (determining its conventional, semantic meaning). In such circumstances, language evidence from the corpus may have less to offer. On the other hand, where prior courts have not yet construed the boilerplate language at issue in a given case, the corpus may provide useful information about the conventional, semantic meaning of the words in the provision.

The corpus is agnostic as to the question of reference timeframe posed above. If the question is whether a provision of a contract was specifically negotiated, or when particular language was first included in similar contracts, or what were the prevailing linguistic conventions at the time of execution or at the time when specific language was first drafted—corpus linguistics can provide helpful, linguistic evidence to address such questions. But the question of whether to apply a legal fiction and interpret contracts according to linguistic conventions prevailing at the time they are executed—regardless of whether or not

301. Id.
302. Id.
303. Id. at 1109–10.
the individual provisions have been separately negotiated and agreed to by the parties—is a jurisprudential question that cannot be addressed with corpus evidence.

B. Register and Speech Community in Contract Interpretation

A further issue arises with the question of speech community. As discussed above, courts often turn to general-use, unabridged English dictionaries when called upon to interpret contracts. This seems to suggest that courts view the task of interpretation as identifying linguistic conventions that are consistent with Standard Written American English. It is not necessarily obvious that all contracts are written with Standard Written American English in mind. As we saw with governable in Problem No. 3, some words take on a specialized sense in certain legal contexts. We would expect contracts to adopt the linguistic conventions of the industry or subject matter for which they are drafted.

The interpretation of a contract should take into account the linguistic speech community for which it was drafted. In the case of consumer contracts, we might adopt a legal fiction that presumes the contract should be interpreted according to linguistic conventions of a particular industry. The construction of a specialized, industry-specific corpus may be necessary to properly account for the linguistic conventions of that industry.

Sometimes it will not be clear whether a given, contested word or phrase takes on a specialized meaning in a particular legal or commercial context. In such circumstances the use of comparative corpora of both Standard Written American English and the relevant legal or commercial speech may be necessary to identify specialized uses of a term, as demonstrated by the identification of the specialized legal use of governable identified in Problem No. 3.

C. Validating (or Falsifying) Corpus-Based Results

One advantage of a corpus-based approach to contract interpretation is the ability of third parties to either validate or falsify corpus-based results. When using a corpus to address an interpretive question leaves a trail of the methods used and the evidence considered. The corpus user may also download the evidence in question and provide it to third parties. A party attempting to verify or falsify the result may repeat the same queries or review the same evidence.

Corpus evidence is not the only linguistic evidence used to evaluate questions of contractual meaning. In their recent article, Interpreting
Contracts via Surveys and Experiments, Omri Ben-Shahar and Lior Jacob Strahilevitz propose to incorporate empirical survey methods from trademark and unfair competition cases into contract interpretation. Ben-Shahar and Strahilevitz argue that survey methods can provide real-world evidence of public perception of contractual plain meaning, as well as real-world evidence of what changes in contractual language would result in less ambiguity. They further argue that surveys can “capture that meaning more accurately than a judge’s imagination.” Surveys may also be able to provide information about the way a contract would be interpreted by different speech communities.

The corpus approach and the survey approach each have their comparative advantages. Sophisticated linguistic corpora are freely available from any computer terminal and can be put to use with adequate training. Because linguistic corpora rely on naturally occurring language (rather than language elicited from survey respondents), corpus linguistics is able to avoid some of the difficult-to-mitigate design challenges and context effects that surveys may be subject to. “Even the best designed elicitation tasks are removed from how people use (and think about) language in everyday life, and people’s reports of their linguistic usage may or may not match up with what they actually do.” However, linguistic corpora face their own design challenges, and because corpus evidence is typically limited to natural language, it is sometimes the case that sufficient texts may not be available to make judgments about the linguistic conventions of the relevant speech community or register at the relevant time. By contrast, survey

305. Id. at 1780.
306. Id. at 1759.
307. See 1 ENCYCLOPEDIA OF SURVEY RESEARCH METHODS, 142 (Paul J. Lavrakas ed., 2008); EDWARD K. CHENG ET AL., MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY § 7:20 (2017) (“The order in which questions are asked on a survey and the order in which response alternatives are provided in a closed-ended question can influence the answers . . . . Although these effects are typically small, no general formula is available that can adjust values to correct for order effects . . . .”); Jon D. Hanson & Douglas A. Kysar, Taking Behavioral Seriously: The Problem of Market Manipulation, 74 N.Y.U. L. Rev. 630, 676–77 (1999) (discussing context effects).
310. This problem was highlighted by the complete absence of examples of the phrase any sport in the context of the verbs participate and practice in the COCA and the paucity of examples of the same in the NOW Corpus. See supra note 191.
methods allow for the particular interpretive question to be narrowly
drafted and specifically targeted to the relevant speech community. This
may allow survey methods to better account for some contextual
information—like the social or spatial context of an utterance—that may
not be easily addressed with corpus evidence.

At bottom, corpus linguistics and survey methods measure different,
but possibly related linguistic phenomena. Corpus linguistics can
identify the linguistic conventions of a speech community based on that
community’s language output or production in natural language settings.
Survey methods, by contrast, may measure language perceptions that
cannot as easily be captured with usage evidence from a corpus.

D. The Meaning of Ambiguity and the Law of Interpretation

One contribution of the corpus approach highlighted above is the
ability to provide objective evidence of the range of potential senses of a
given word in a contractual context. For example, in the cases of sport in
Problem No. 1 and anticipated in Problem No. 2, the courts purported to
identify the plain meaning of the operative term, and concluded that
there was no ambiguity, even though the corpus evidence made clear
that both of the contested senses of sport and anticipated were possible,
attested senses.311 If the standard for what constitutes ambiguity is
“capable or susceptible of different interpretations,” then both courts’
conclusions about plain meaning are in error. The proposition that
disputed contractual terms that are the subject of a litigated dispute could
be susceptible of only one plain meaning is rarely justifiable. After all,
“[p]eople spend the money to come to court only when it is possible to
draw conflicting inferences from the words alone.”312 Moreover, as
noted above, the more common a word is, the more likely it is to have
multiple senses. And the more senses a word has, the more likely
interpreters are to disagree about its meaning in a given context.313

The judicial propensity to gloss over problems of ambiguity and to
arbitrarily select among competing meanings is not new. Arthur L.
Corbin observed:

All through the history of the common law, there is found a very
common assumption of the existence of antecedent rules and
principles, beginning no man knows when, coming from no man

311. See supra section IV.A (Problem 1) & section IV.B (Problem 2).
 PUB. POL’Y 59, 60 (1988).
313. See supra note 37.
knows where, seemingly universal and unchangeable . . . . And yet, at almost all periods, there have been a few jurists who took thought to the matter and who knew better . . . . Among such rules are those indicating that words must have one, and only one, true and correct meaning, [which] must be sought only by poring over the words within the four corners of the paper . . . .

Of course, judges know that contractual language can be ambiguous and have developed rules of interpretation to guide judicial decision-making in the event that such ambiguity is identified. The Plain Meaning Rule is one such rule. Another is the contra proferentem canon, which, as noted above, holds that where the meaning of a contractual provision is found to be ambiguous, courts should “adopt the meaning that is less favorable in its legal effect to the party who chose the words.”

We can well imagine a court applying the contra proferentem canon to both Problem Nos. 1 and 4 because in both cases there were at least two competing uses of the operative terms (sport and structural damage) and because the canon tends to be applied “more rigorously” in insurance contracts, where it is recognized that there is a “difference between the parties in their acquaintance with the subject matter.” This is the role that such canons have historically played in the resolution of contractual ambiguity.

In their recent article, The Law of Interpretation, Professors Will Baude and Stephen E. Sachs argue that “an instrument’s legal effect

314. ARTHUR L. CORBIN, CORBIN ON CONTRACTS, § 535, at 499 (1952); see also id. at 496–97 (“It is true that when a judge reads the words of a contract he may jump to the instant and confident opinion that they have but one reasonable meaning and that he knows what it is. A greater familiarity with dictionaries and the usages of words, a better understanding of the uncertainties of language, and a comparative study of more cases in the field of interpretation, will make one beware of holding such an opinion so recklessly arrived at.”).

315. See KNIEFF, supra note 1, § 24.27, at 282–83. See also RESTATEMENT (SECOND) OF CONTRACTS, supra note 17, § 206 (“In choosing among the reasonable meanings of a promise or agreement or a term thereof, that meaning is generally preferred which operates against the party who supplies the words or from whom a writing otherwise proceeds.”).

316. Gaunt v. John Hancock Mut. Life Ins. Co., 160 F.2d 599, 601–02 (2d Cir. 1947); see also Eagle-Picher Indus., Inc. v. Liberty Mut. Ins. Co., 682 F.2d 12, 23 (1st Cir. 1982) (“[A]ny remaining doubts about interpretation of the policies are properly resolved in favor of the insured, in order to effectuate the policies’ purpose of providing coverage.”). The contra proferentem canon is not limited to cases of linguistic ambiguity but may also be applied in circumstances where the scope of the insurance contract is merely uncertain. Ins. Co. N. Am. v. Forty-Eight Insulations, 633 F.2d 1212, 1219–20 (6th Cir. 1980) (“If there are ambiguities in the policy, or uncertainty over its interpretation, the policy is to be construed against the insurer, and in favor of the insured.”) (emphasis added)).
doesn’t just follow from the meaning of its language . . . .”317 Instead, “the legal system frequently chooses artificial rules of interpretation, and once chosen they’re the law, whether or not they reflect what a given text really meant.”318

This may be correct. But even if the law of interpretation stands poised to provide a substantive (or “artificial”) resolution to problems of ambiguity, it can do little to assist the court with the antecedent identification of ambiguity that is embedded in many contract interpretation canons, including the Plain Meaning Rule and the contra proferentem canon.

Many of the so-called “substantive” canons of interpretation are in fact hybrid canons of interpretation—they require, first, the identification of the presence of ambiguity, and second, a determination of the legal effect of that ambiguity. In the context of criminal law, courts apply the rule of lenity, which requires, first, the identification of an ambiguity in a criminal statute, and second, the resolution of that ambiguity in favor of the criminal defendant. Though the number of steps in the application of Chevron deference is a matter of some dispute,319 a familiar account requires that a court first determine whether a statute is silent or ambiguous on a subject, and second, defer to an agency’s interpretation of a statute if the interpretation of the agency is reasonable.320

Viewed in this light, the law of interpretation often becomes a collection of heuristics that resolves only second-order interpretive problems. It is a law of Step-two, but not of Step-one.

If we accept the common definition of contractual ambiguity as language that is “capable or susceptible of different interpretations,”321 then we would have to conclude that the courts in Problem Nos. 1, 2, and 4 got it wrong. In each case, the court purported to identify the plain meaning of the operative term (sport, anticipated, and structural damage), and in each case, the operative term had more than one attested use, a fact that each court ignored or, at least, failed to meaningfully

317. Baude & Sachs, supra note 276, at 1083.
318. Id. at 1095.
evaluate and explain. We can also imagine that a court applying the same definition of ambiguity to Problem No. 3, and concluding, as a result of courts’ traditional reliance on dictionaries, that governable has only one meaning (because governable has only one definition in several common dictionaries). This result seems unsatisfactory because it fails to take account of a well-attested, specialized legal meaning that appears to do a better job of reflecting the parties’ reasonable economic expectations.

The incongruity between the evidence from the corpus on the one hand and the common judicial understanding of ambiguity on the other suggests two possible outcomes. Courts must either refine their understanding of plain meaning and ambiguity, or they should abandon the Plain Meaning Rule altogether. If we define ambiguity as the existence of two possible interpretations, then virtually every contractual provision contains some ambiguity and the application of the Plain Meaning Rule is always problematic. On the other hand, if we do away with the Plain Meaning Rule entirely, we eliminate the benefits of objectivity and efficiency that the rule purportedly provides.

One possible solution would require reformulating the rule in terms of probability. To the extent that we are confident that we have identified the appropriate syntactic, semantic, and pragmatic context of the interpretive question at issue, and to the extent that we are confident that we have correctly identified the relevant speech community, register, and timeframe for the interpretive task at hand, we might entertain a rebuttable presumption in favor of an overwhelmingly more common use of the word or phrase in the context in question. The strength of this presumption would be correlated to the prevalence of a particular use of a word or phrase, in a particular (semantic, syntactic, and pragmatic) context, in the language use of a particular speech community, and language used at a particular timeframe. The more prevalent the usage in this context, the stronger the contrary evidence would need to be to rebut it. In Problem Nos. 1 and 2 above, the overwhelmingly more common uses of sport (“rule-based competition”) and anticipated (“expected, likely event”) would control unless strong evidence were brought forward to rebut the application of these senses. In Problem No. 4, where the linguistic evidence makes clear that neither sense of structural damage is overwhelmingly more common (or even discernible from context), the court would be free to allow additional extrinsic evidence under the Plain Meaning Rule, or to apply a substantive canon, like the contra proferentem canon.

Problem No. 4 illustrates that the Plain Meaning Rule may not be the only canon of interpretation in need of reexamination (and, possibly,
reformulation). In that case, the court applied the so-called canon against surplusage and insisted that reading “structural damage” as mere “physical damage” to a structure would render “structural” meaningless in violation of the canon against surplusage. But the corpus evidence failed to show any meaningful way to distinguish among the competing, contested uses of “structural damage,” and, indeed, the corpus evidence gives a strong indication that the adjective “structural” adds little to nothing to the meaning of the contractual provision at issue in Problem No. 4 and that the court’s insistence to the contrary is mistaken.

Commenting on the surplusage canon in the statutory context, William Baude and Stephen E. Sachs observed that “[i]f redundancy were actually far more common than we realized among the relevant readers and speakers, then the canon against superfluity might need to be modified or abandoned.” The presence of co-selection and delexicalization in Problem No. 4 (and the prevalence of co-selection and delexicalization as a common linguistic phenomena as demonstrated in language evidence from linguistic corpora) suggest that redundancy in human language is much more common than some jurists believe and that a reformulation of the surplusage canon may be in order. Moreover, Baude and Sachs are undoubtedly correct that “[i]t’s no answer to say, as some defenders of the surplusage canon do, that ‘[contracts] should be carefully drafted,’” or that contract drafters “ought to hire eagle-eyed editors’ to conform . . . to the canon.” Even when engaged in the type of more or less careful use of language that contract drafting entails, there is no evidence that contract drafters are exempt from the subconscious co-selection and delexicalization of words that is one of the hallmarks of word choice by competent English speakers. Nor is there any evidence that “eagle-eyed editors” are capable of identifying and striking redundant words that are subconsciously selected in this way. Instead, the prevalence of subconscious co-selection and delexicalization established by the corpus evidence suggests that the surplusage canon may be in need reformulation to account for common linguistic redundancies.

What is clear from the Problems examined above is that the concepts of Plain Meaning and ambiguity are incompletely theorized. In order for

322. Baude & Sachs, supra note 276, at 1126.
323. Id. at 1126–27 (quoting Antonin Scalia & Bryan A. Garner, Reading Law: The Interpretation of Legal Texts 179 (2012)). As noted, the quotation here is in reference to the use of the surplusage canon in the statutory context. However, the case for reexamination and reformulation of the canon is similar, whether the canon is invoked in the contractual or statutory context.
the Plain Meaning Rule to do the work it is expected to do, courts must begin to take account of a more nuanced and accurate picture of the way language is used and develop interpretive rules and interpretive practices that more accurately reflect relevant language conventions. Linguistic corpora can provide evidence-based content to otherwise vague and poorly defined interpretive concepts like Plain Meaning and ambiguity. Corpus evidence can also provide the tools necessary to evaluate and possibly reformulate canons of contractual interpretation that are failing to give a proper account of the linguistic conventions they are expected to describe and predict.

CONCLUSION

Linguistic corpora will not provide answers to every interpretive question. Judges and lawyers will have to decide the legal effect of the linguistic information corpora provide. But corpora can provide objective evidence of the linguistic conventions of the communities that draft and are governed by the agreements judges and lawyers are called upon to interpret. Corpus evidence can give content to otherwise vague canons of interpretation and provide linguistic evidence to aid in the evaluation of claims about the Plain Meaning (or ambiguity) of a contractual text.