BEYOND THE DICTIONARY: WHY *SUA SPONTE* JUDICIAL USE OF CORPUS LINGUISTICS IS NOT APPROPRIATE FOR STATUTORY INTERPRETATION

INTRODUCTION

Imagine you are Chief Justice John Roberts, and you have been tasked with deciding whether a corporation, AT&T, is entitled to the "personal privacy" exemption of the Freedom of Information Act ("FOIA"),[1] and is therefore protected from having certain corporate documents disclosed for "personal privacy" reasons.[2] Would the fact that the act itself defines "person" to include "corporation" be dispositive to you?[3] After all, the Supreme Court has found corporations enjoy other legal rights enjoyed by natural persons.[4] What other tools would be available to you to interpret the language of the statute?

The Court, in fact, did not find the definition of "person" to correspond directly to the definition of "personal."[5] The decision was a landmark one in the world of statutory interpretation, in that it was the first (and only to this point) in which the Supreme Court relied on the use of corpus linguistics data to interpret the statutory terms.[6]

Corpus linguistics refers to "the study of language function and use by means of large, principled collections of naturally occurring language called corpora [bodies]."[7] The corpora at B.Y.U., which are the most widely used online corpora, and the

---

[1] 5 U.S.C. § 552(b)(7)(C) (2012).

[2] FCC v. AT&T, Inc., 562 U.S. 397, 409-10 (2011).

[3] 5 U.S.C. § 551(2) (2012) (stating that "person" includes an "individual, partnership, *corporation*, association, or public or private organization . . . .") (emphasis added).

[4] *E.g.*, Burwell v. Hobby Lobby Stores, Inc., 134 S. Ct. 2751 (2014) (protecting the free exercise of religion for closely held for-profit corporations by holding that the contraceptives mandate of the Affordable Care Act substantially burdened exercise of religion in violation of the Religious Freedom Restoration Act).

[5] *FCC*, 562 U.S. at 397 ("[A]djectives do not always reflect the meaning of corresponding nouns.").

[6] During oral arguments, Justice Ginsburg references the brief of the Project on Government Oversight, written by Neal Goldfarb in support of the FCC, in which Goldfarb conducts corpus analysis to show that the word "personal" is overwhelmingly used to describe "an individual, not an artificial being." Transcript of Oral Argument at 37, *FCC*, 562 U.S. 397 (2011) (No. 09-1279).

[7] Stephen C. Mouritsen, *Hard Cases and Hard Data: Assessing Corpus Linguistics as an Empirical Path to Plain Meaning*, 13 COLUM. SCI. & TECH. L. REV. 156, 159 (2011) [hereinafter *Hard Cases*] (citing Douglas Biber, *Corpus-based and Corpus-driven Analyses of Language Variation and Use*, *in* THE OXFORD HANDBOOK OF LINGUISTIC ANALYSIS 159-60 (Bernd Heine & Heiko Narrog eds., 2009)).

ones that will be relevant throughout this Comment, draw from different bodies of real-world text, and include the Corpus of Contemporary American English (COCA), Corpus of Historical American English (COHA), and even Wikipedia and TIME Magazine corpora.[8] By examining large databases of language in naturally occurring contexts, many legal scholars and jurists believe a statute's "ordinary meaning" can be reduced to an empirical question.[9] The Supreme Court is not the only tribunal to make use of corpus analysis. In fact, in the past six years, state supreme court justices in Utah and Michigan have cited corpus data directly in opinions, with varying levels of acceptance from the other justices.[10]

But how reliable is a judge's interpretation of corpus data? While judicial use of corpus data for statutory interpretation has been lauded for its superficially objective and empirical methodology, critics have questioned not only its persuasiveness in statutory interpretation, but also judges' ability to accurately interpret the returned data, and apply it to determine the statute's ordinary meaning.[11]

This Comment will argue that judges should not raise corpus analysis *sua sponte* in judicial opinions. Part I will provide background on different judicial conceptions of "ordinary meaning" and outline the use of dictionaries to arrive at the ordinary meaning of statutory terms. It will also introduce corpus linguistics as it has arisen in statutory interpretation. Part II will show the inconsistent methodologies judges have used in

---

[8] Mark Davies, BYU CORPORA, https://corpus.byu.edu [https://perma.cc/N47M-GRVQ] (last visited Jan. 20, 2018).

[9] *See* Thomas R. Lee & Stephen C. Mouritsen, *Judging Ordinary Meaning*, 127 YALE L.J. 788, 795 (2018) [hereinafter *Judging Ordinary Meaning*]; *see also* 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 (2010) [hereinafter *The Dictionary Is Not a Fortress*].

[10] *In re* Adoption of Baby E.Z., 266 P.3d 702, 724-25 (Utah 2011) (Lee, J., concurring in part and concurring in the judgment) (conducting corpus analysis to define the term "custody" in the context of Utah's Parental Kidnapping Prevention Act, and noting both the majority and separate concurrence objected to reliance on this data); People v. Harris, 885 N.W.2d 832, 838-39 (Mich. 2016) (conducting corpus analysis by the majority to determine whether "information" included false or inaccurate statements under Michigan's Disclosures by Law Enforcement Officers Act).

[11] State v. Rasabout, 356 P.3d 1258, 1264-66 (Utah 2015).

analyzing corpus data, as well as how corpus analysis can produce different outcomes from standard methods of statutory interpretation. Finally, Part III will differentiate corpus linguistics from other methods of statutory interpretation, and argue against its *sua sponte* use, focusing on objections grounded in judicial notice, the adversarial process, and the "deceptive empiricism" of corpus use in the context of statutory interpretation.

I. BACKGROUND

A. Judges' Differing Conceptions of "Ordinary Meaning"

While the "ordinary meaning" rule is a cornerstone of judicial interpretation,[12] the methods by which judges have arrived at the ordinary meaning of statutory terms have been criticized for being inconsistent and lacking repeatable methodology. [13] Often, a judge's utilization of this canon of construction will depend on that judge's broader philosophy of statutory interpretation. The "purposivist" judge utilizes ordinary meaning only to the extent which a proposed interpretation is "linguistically permissible" and consistent with that judge's interpretation of the statute's purpose. [14] In contrast, the textualist judge demands a much higher level of textual sufficiency, and when a statutory term is not defined specifically, uses "linguistic intuition" to determine how a word is ordinarily understood.[15]

In essence, judges utilize the ordinary meaning rule on a spectrum of usage exclusivity. On one end of the spectrum are "possible" (or "permissible") meanings of a word, and on the other is the "exclusive" meaning of the word. [16] Another way that scholars and judges have characterized the exclusivity spectrum is in the context of "prototype" analysis.[17] Prototypes are the clearest

---

[12] *Hard Cases*, *supra* note 7, at 159-60.

[13] *Id.*

[14] *Id.* at 160.

[15] *Id.*

[16] *Judging Ordinary Meaning*, *supra* note 9, at 800.

[17] *See* Lawrence M. Solan, *The New Textualists' New Text*, 38 LOYOLA LA L. REV. 2027, 2029-34 (2005) [hereinafter *New Textualists' New Text*]; *see also* McBoyle v. United States, 238 U.S. 25, 27 (1931) (determining whether an airplane was a vehicle for purposes of The National Motor Vehicle Theft Act of 1919, by Justice Holmes'

and most obvious examples within the category that a statutory term represents.[18]

In sum, while judges often agree that the ordinary meaning of a statutory term ought to govern, there is no consensus as to the definition of "ordinary meaning" itself.[19] This Part will use recent Supreme Court decisions[20] on ordinary meaning to show the variety of ways that judges utilize the ordinary meaning rule at different points on the exclusivity spectrum. These cases will be discussed in further detail and in different contexts throughout this Comment.

1. *Muscarello v. United States*

In *Muscarello*, the Court was asked to interpret 18 U.S.C. § 924(c)(1), which imposed a five-year mandatory sentence on anyone who "uses or carries a firearm" in connection to a "drug trafficking crime."[21] The question presented was whether transporting a gun in the locked glove compartment of a car counted as "carrying" within the meaning of the statute.[22]

---

conducting of a prototype analysis in pointing out that the act "evoke[d] in the common mind only the picture of vehicles moving on land"); Lawrence M. Solan, *Judicial Decisions and Linguistic Analysis: Is There a Linguist in the Court?*, 73 WASH. U. L.Q. 1069, 1076 (1995) ("Definitions work from the outside and move inward, and prototypes work from the inside and move outward.").

[18] *See* Lawrence M. Solan, *Why Law Works Pretty Well, But not Great: Words and Rules in Legal Interpretation*, 26 L. & SOC. INQUIRY 243, 258 (2001) [hereinafter *Why Law Works Pretty Well*] ("Prototype analysis tells us that the notion of ordinary meaning has a cognitive basis.").

[19] *New Textualists' New Text*, *supra* note 17, at 2031-32 (citing Holy Trinity Church v. United States, 143 U.S. 457 (1892), as one of the early examples of differentiating between "plain" and "ordinary" meaning in statutory interpretation); *see also The Dictionary is Not a Fortress*, *supra* note 9, at 1952 ("When jurists speak of 'ordinary meaning,' they simply are not always talking about the same thing."); *Judging Ordinary Meaning*, *supra* note 9, at 798 ("[I]ronically, we have no ordinary meaning of 'ordinary meaning.'").

[20] I am not the first to analyze these cases in the context of statutory interpretation, as they represent quintessential examples of contrasting methods used by judges. *See, e.g.*, *Judging Ordinary Meaning*, *supra* note 9, at 803-06; *The Dictionary is Not a Fortress*, *supra* note 9, at 1952-53; Daniel Ortner, *The Merciful Corpus: The Rule of Lenity, Ambiguity and Corpus Linguistics*, 25 B.U. PUB. INT. L.J. 101, 124-35 (2016).

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

[22] *Id.*

In the majority opinion, Justice Breyer seems to blur the line as to which part of the usage spectrum his analysis falls under. Breyer looked at a sample of sentences in Lexis and Westlaw consisting of press usage (*New York Times* and *USA Today*) of the term in sentences that included "carry," "vehicle" and "weapon" in the same sentence.[23] Because more than one-third of these sentences were conveying the idea at issue in this case (carrying a weapon in a vehicle), he concluded that the "in a vehicle" sense of the word "carry" met the threshold frequency to fall under the ordinary meaning of the statute.[24]

However, elsewhere in the analysis, Breyer seems to imply that the "in a vehicle" sense of the word "carry" is not only a common meaning sufficient to meet the ordinariness test, but also the *most* frequent use of the term. This analysis was based on both dictionary definitions, including the Oxford English Dictionary (OED) and the Webster's Third New International Dictionary (WNID3), and characterization of the "in a vehicle" sense of "carry" as "primary," and the "on the person" sense as "special."[25] Even more explicitly, Breyer analyzes the etymological origins of the verb "carry" and concludes that the "in a vehicle" sense is the "first, or basic, meaning of the word."[26] Even though he qualifies his argument by saying that he is excluding other permissible meanings of the verb "to carry" and dichotomizing the "vehicle" and "on the person" senses in his analysis,[27] he is ambiguous as to whether he concludes that the "in a vehicle" sense is primary only among these two definitions, or primary among all definitions.[28] Above all, this case demonstrates that judicial understanding of

---

[23] *Id.* at 129.

[24] *Id.*

[25] *Id.* at 128-30.

[26] *Id.* at 128.

[27] *Id.*

[28] *See The Dictionary Is Not a Fortress*, *supra* note 9, at 1952 ("[T]he claim that the ordinary sense of *carry* 'includes' carrying in a car, is a far cry from the claim that this is the word's first, primary, and 'ordinary English' meaning."); *see also id.* (noting that the "linguistically permissible" analysis seems to be the basis of *Muscarello*'s reasoning).

"ordinary meaning" is not always consistent with the concept of the most "common usage."[29]

2. *Smith v. United States*

In *Smith*, the Court was interpreting the same statute as in *Muscarello*.[30] This time, however, the Court was tasked with deciding whether the defendant's attempt to trade a firearm for cocaine constituted using "a firearm 'during and in relation to . . . [a] drug trafficking crime[.]'"[31]

The majority opinion, written by Justice O'Connor, acknowledged that when a word in a statute is not given a particular meaning, any nontechnical word is to be given its "ordinary or natural meaning."[32] Citing WNID3 and Black's Law Dictionary for support, the majority concluded that the verb "use" in this context was wide enough to encompass the actions of the defendant in this case.[33] The majority also recognized that statutory language cannot be analyzed devoid of context, and argued that while "uses a firearm" may *include* using a firearm as a weapon, as this is the intended purpose of a firearm, this fact does not *exclude* other uses of a firearm that may fall within the category covered by the verb.[34] In essence, the majority found the use of the term "linguistically permissible" in this context, even though it acknowledged the use was not the most ordinary meaning.

On the other hand, the dissent, written by Justice Scalia, utilized a prototype analysis to argue that while the verb "use" may have been linguistically permissible to describe the defendant's conduct, it is far from the ordinary meaning of "uses a firearm."[35] In the context of an instrumentality such as this, to

---

[29] *Id.* at 1953-54 ("[T]he Court does not appear to grasp the distinction between how a word *can be* used and how it *ordinarily is* used.") (quoting Smith v. United States, 508 U.S. 223, 242 (1993) (Scalia, J., dissenting) (emphasis in original)).

[30] Smith v. United States, 508 U.S. 223, 225 (1993) (citing 18 U.S.C. § 924(c)(1)).

[31] *Id.*

[32] *Id.* at 228.

[33] *Id.* at 228-29.

[34] *Id.* at 230.

[35] *Id.* at 242 (Scalia, J., dissenting); *see also Why Law Works Pretty Well*, *supra* note 18, at 258 (using *Smith* as an example of a "battle[] among the justices over definitions versus prototypes"). *But see* MCI Telecomms. Corp. v. American Tel. & Tel.

"use" ordinarily means to use in the capacity for which it was intended.[36]

3. *Taniguchi v. Kan Pacific Saipan*

In a more recent case, the Court interpreted a statute allowing the prevailing party in federal litigation to recover certain costs, including costs incurred by an "interpreter."[37] When the case was dismissed, the defense submitted a request for expenses incurred from having documents translated from Japanese to English.[38]

In the majority, written by Justice Alito, the Court again proved consistent at least in its first step of the "ordinary meaning" analysis: noting that any nontechnical term not defined in a statute is to be given its "ordinary meaning."[39] In finding that "interpreter" ordinarily meant one who translates oral, rather than written language, the majority cited a number of dictionaries current with the passage of the statute in 1978, including general use dictionaries and the then-current edition of Black's Law Dictionary.[40]

The dissent, written by Justice Ginsburg, relied on the definition in WNID3, which defined "interpreter" as "one that translates; *esp*[41]: a person who translates orally for parties conversing in different tongues."[42] The dissent also relied on the 2009 edition of Black's Law Dictionary (rather than the 1968 edition, relied on by the majority, which was current with the statute's passage).[43] Most notably, the dissent concedes that the "written translator" definition is not the most common, but argues

---

Co., 512 U.S. 218, 225-26 (1994) (writing for the majority, Scalia cites various dictionaries, rather than conducting a prototype analysis, to argue that the FCC's power to "modify" a filing requirement for common carriers did not give it the power to eliminate the filing requirement altogether).

[36] *Why Law Works Pretty Well*, *supra* note 18, at 258.

[37] Taniguchi v. Kan Pac. Saipan, Ltd., 566 U.S. 560, 567-68 (2012).

[38] *Id.* at 563.

[39] *Id.* at 566.

[40] *Id.*

[41] Sense divider, meaning "especially."

[42] *Id.* at 576 (Ginsburg, J., dissenting).

[43] *Id.*

that either of the two senses of the term should fall under its "ordinary meaning."[44]

This case raises two important questions. Firstly, should the currentness of the dictionary be selected based on the statute's passing, contemporaneous with the case at hand, or by some other standard? Secondly, when two senses of a term are "common," but one is slightly more so, is the less common sense excluded from the statutory term's ordinary meaning?[45]

B. The Rise of Dictionary Usage as a Tool for Ordinary Meaning

The Court's use of dictionaries has experienced an exponential rise in recent years. At the historical outset, citation to dictionaries was very sparse.[46] At different times in the Court's jurisprudence, jurists have pointed to a few different historical and interpretive developments to explain the rise in dictionary usage.

The first explanation came from the historical decrease in the Court's role in legislating common law, as more and more cases became tied to statutory language and interpretation.[47]

The most recent explanation, which grew out of this phenomenon, was the rise of "new textualism" as the Court's primary interpretive methodology.[48] Coupled with this movement was the preference of "original plain meaning'" of the statutory terms over "'original intent' . . . of the drafters."[49] This movement, championed by Justice Scalia, called for examination of the

---

[44] *Id.*

[45] *Judging Ordinary Meaning*, *supra* note 9, at 804-05 ("[The dissenters] do not expressly disagree with Justice Alito's assertion that the *oral translator* notion is most common; they are simply saying that [either of two] common senses of a term should count as ordinary.") (citing *Taniguchi*, 566 U.S. 560) (Ginsburg, J., dissenting).

[46] *The Dictionary Is Not a Fortress*, *supra* note 9, at 1920 (noting that "[p]rior to 1864, the Supreme Court cited dictionaries in only three cases.").

[47] Felix Frankfurter, *Some Reflections on the Reading of Statutes*, 47 COLUM. L. REV. 527 (1947) (noting the decrease in "common-law litigation" from 40% in 1875 to almost zero in the middle of the twentieth century, concluding that the Court had greatly diminished its legislative capacity in the context of common law).

[48] William N. Eskridge, Jr., *The New Textualism*, 37 UCLA L. REV. 621 (1990).

[49] Phillip A. Rubin, *War of the Words: How Courts Can Use Dictionaries in Accordance with Textualist Principles*, 60 DUKE L.J. 167, 171 (2010) [hereinafter *War of the Words*] (describing this form of originalism as a "corollary of textualism").

statutory text alone, and only limited reference to extrinsic material. [50] The movement is evidenced by the exponential increase in dictionary usage beginning in the 1970s.[51] A recent study has found that this rise in dictionary usage correlates to the increased use of the term "ambiguous" or a variation thereof in the Court's language usage in opinions, noting that between 1970 and 2005, the term "ambiguous" increased to an occurrence of over one-hundred words per million, from only eleven to thirteen words per million from its inception until 1969.[52]

In updating this study of the Court's use of the term "ambiguous," a search in the Corpus of U.S. Supreme Court Opinions reveals that while the decade average so far in the 2010s is close to as high as it has ever been (65.49 words per million), the term's frequency was only 21.38 words per million in the year 2016, which is the lowest frequency since the 1970s.[53]

1. How Courts Use (and Misuse) Dictionaries

Initially, dictionaries played a minimal role in the Court's interpretive analysis. Specifically, dictionary definitions served as a memory aid for judges, for those terms that were not technical in nature, but that judges had forgotten or did not understand.[54] In other cases, judges used dictionaries as a source for a selection of definitions from which to choose, as a vehicle for context-based arguments, stemming from the intent or purpose of the statute in question.[55]

---

[50] Rickie Sonpal, *Old Dictionaries and New Textualists*, 71 FORDHAM L. REV. 2177, 2192-93 (2003).

[51] Jeffrey L. Kirchmeier & Samuel A. Thumma, *The Lexicon Has Become a Fortess: The United States Supreme Court's Use of Dictionaries*, 47 BUFF. L. REV. 227, 251-52 (1999) [hereinafter *The Lexicon Has Become a Fortress*].

[52] *The Dictionary Is Not a Fortress, supra* note 9, at 1920.

[53] Word search in Corpus of U.S. Supreme Court Opinions for "ambiguous", BYU CORPUS, https://corpus.byu.edu/scotus [https://perma.cc/P95D-J49T] (last visited Oct. 20, 2017).

[54] Note, *Looking It Up: Dictionaries and Statutory Interpretation*, 107 HARV. L. REV. 1437, 1439 (1994) [hereinafter *Looking It Up*]; *The Dictionary Is Not a Fortress, supra* note 9, at 1921 (referring to this function of dictionaries as the "definitional function").

[55] *Looking It Up, supra* note 54, at 1439-40. (citing Colony, Inc. v. Commissioner, 357 U.S. 28, 33 (1958)).

Another early function of dictionaries was in an "instantiating" capacity.[56] In this context, dictionaries were used by judges whose proffered meanings had been challenged, and who looked to dictionaries to prove that the proffered sense of a term had been employed in the everyday lexicon.[57]

This latter function is consistent with both the "linguistically permissible" ordinary meaning philosophy discussed above (in Part I.A),[58] as well as the intended use of dictionaries. Almost all modern dictionaries are "descriptive," meaning they are historical records of a term's existing usage, and do not purport to lay out how a term should be used or what meaning it must always have in a given context.[59]

Recently, however, judges have misused dictionaries in ways that take them outside of their descriptivist function. The first way judges have misused dictionaries is by attempting to make persuasive arguments about a term's "primary" meaning by the definition's placement in the hierarchy of definitions, called the "Sense-Ranking Fallacy."[60] An example of an instance when this flawed analysis was dispositive was in *Smith v. United States,* in which the Court was interpreting the term "foreign country" in the Federal Tort Claims Act.[61] The question before the Court was whether Antarctica, which had no recognized government, fell within the statutory meaning of "foreign country." In finding that Antarctica was within the category covered by the term "country," the Court gave weight to the first dictionary definition of "country" as a "region or tract of land," to the exclusion of the second definition, "a political state or nation."[62]

---

[56] *The Dictionary Is Not a Fortress, supra* note 9, at 1922.

[57] *Id.*

[58] *Id.*

[59] LEXICOGRAPHY: PRINCIPLES AND PRACTICE 5 (R. R. K. Hartmann ed., 1985); *see also* WNID3 preface ("[A] definition, to be accurate, must be written only after an analysis of usage"); Hart & Sacks, THE LEGAL PROCESS: BASIC PROBLEMS IN THE MAKING AND APPLICATION OF LAW 1190 (Eskridge & Frickey eds., 1994) (dictionaries provide "an historical record, not necessarily all-inclusive, of the meanings which words in fact have borne.").

[60] *The Dictionary Is Not a Fortress, supra* note 9, at 1928.

[61] Smith v. United States, 507 U.S. 197 (1993).

[62] *Id.* at 201 (citing Webster's New International Dictionary 609 (2d ed. 1945)); *see also* Ellen P. Aprill, *The Law of the Word: Dictionary Shopping in the Supreme Court,* 30 ARIZ. ST. L.J. 275, 298 (1998) [hereinafter *The Law of the Word*] (describing the

To understand why "sense-ranking" is a fallacy, one must look at the purpose of dictionaries as well as how they are compiled. Today, lexicographers compile dictionaries from a corpus, which is "a collection of whole or partial texts or recorded speech stored and indexed electronically so that individual words can be found quickly."[63] Since it is impossible to compile every real-world use of a term for the corpus, lexicographers have discretion in choosing the samples, as well as in choosing how many to include, in order to have a collection of samples that is accurately representative of the lexicon.[64] It is through this lens that the purpose of dictionaries in statutory interpretation must be viewed. In addition to limitations imposed by their descriptivist function as historical records of language usage,[65] dictionaries are also limited in that they are merely a "proxy for the lexicon" since they are merely representative.[66] Furthermore, lexicographers do not necessarily list the most common (or "primary") sense of a word as the first definition, and instead often elect the "most literal and typical meaning of the word."[67] For these reasons, ranking definitions by placement in the hierarchy is an overstep on the proper function of dictionaries.

A second common misuse of dictionaries in the context of statutory interpretation is analysis of the word's etymological roots. *Muscarello* also provides an example of this fallacious analysis, by examining etymological dictionary entries to argue that the ordinary meaning of "carry" encompasses to "convey in a

---

latter definition as "far more appropriate to the subject matter of the statute"); Muscarello v. United States, 524 U.S. 125, 128 (1998) (listing the first definition of the term "carry" in multiple dictionaries, and concluding that this meaning was "primary").

[63] *War of the Words*, *supra* note 49, at 179.

[64] *Id.* at 179-80.

[65] *See supra* note 59 and accompanying text.

[66] *War of the Words*, *supra* note 49, at 189 ("In invoking a dictionary to define a word, one is not really searching for what the dictionary *says*, but rather what the word *means* within the lexicon.").

[67] Angus Stevenson, *Tricks of the Trade: How Do You Decide the Meaning of a Word?*, THE GUARDIAN (Dec. 1, 2007), https://www.theguardian.com/money/2007/dec/01/workandcareers.work7 [https://perma.cc/W6GL-DAHR] (noting that while the most "common" use of the word "run" is in the sense of to "run a company," this is obviously not the most literal sense of the word, and would not be listed as the first definition).

car."[68] Both legal scholars and linguistic experts have pointed out that, in the analysis of ordinary meaning, a term's earlier or original meaning offers no assistance in determining present usage. [69] Citing to the etymological origins of a term (often involving examining its roots in other languages, such as Latin) is a "much-practiced folly" that can lead to deceptive and anachronistic evidence of word meaning.[70]

A third issue with the use of dictionaries is the lack of consistency with which judges cite specialized legal dictionaries, as opposed to general use dictionaries. While judges have acknowledged that the ordinary meaning analysis begins with the assertion that nontechnical words are to be given their "ordinary meaning," judges often include technical dictionaries in their analysis of nontechnical terms. [71] In doing so, the Court "undermine[s] the textualists' claim that they seek to find the meaning of ordinary language for the ordinary speaker."[72] While judges may have some basis for analysis of two different dictionary types in the ordinary meaning analysis of a nontechnical term, such a basis has consistently gone unstated, and seemingly amounts to nothing more than an unchecked aggregation of definitions.[73]

---

[68] Muscarello v. United States, 524 U.S. 125, 128 (1998) (citing the Barnhart Dictionary of Etymology 146 (1988) and the Oxford English Dictionary); *see also* United States v. Lopez, 514 U.S. 548, 586 (1995) (examining the etymological origins of the term "commerce" to support the contention that regulating firearm possession in a school zone was outside Congress's authority to regulate interstate commerce).

[69] MICHAEL STUBBS, WORDS AND PHRASES: CORPUS STUDIES OF LEXICAL SEMANTICS (2001); *The Dictionary Is Not a Fortress*, *supra* note 9, at 1939-41; *New Textualists' New Text*, *supra* note 17, at 2051.

[70] Kenneth G. Wilson, THE COLUMBIA GUIDE TO STANDARD AMERICAN ENGLISH 178 (using the term "dilapidated" as an example of this fallacy: even though the term comes from the Latin "dapis," meaning "stone," in Modern English, the term is not limited to only those structures made of stone).

[71] Smith v. United States, 507 U.S. 197, 228-29 (1993) (citing both Black's Law Dictionary and Webster's International Dictionary to define "use").

[72] *The Law of the Word*, *supra* note 62, at 311.

[73] *Id.*; *see also Looking It Up*, *supra* note 54, at 1439 n.12 (noting that while a distinction can be made between specialized and general use dictionaries, judges "have given no indication that they find non-legal dictionaries less useful.").

C. The Rise of Corpus Linguistics as the Answer

1. Corpus Linguistics: What Is It?

In response to judges' seemingly standard-less ordinary meaning methodologies, scholars and a few judges have turned to a linguistic method called "corpus linguistics" as a way to interpret a statute's ordinary meaning.[74] Corpus Linguistics is the "study of language function and use by means of an electronic collection of naturally occurring language called a corpus."[75] Because corpus analysis is simply a method for data collection and analysis, rather than a "theory of language," it is distinct from other branches of linguistics.[76] While originally created by lexicographers and linguists to examine speech patterns, corpora are now available to the public, and have been transplanted into other areas of research.[77] Because corpora attempt to include a representative sample of the target demographic, they include texts from a variety of media and genres. For the COCA, these texts are "evenly divided" among the five genres of "spoken, fiction, [popular] magazine[s], newspaper[s], and academic [journals]."[78]

Corpus analysis has a few defining characteristics: first, because it analyzes real world language usage in naturally occurring contexts, it is an empirical approach to linguistic

---

[74] *See* Brief for the Project on Government Oversight, the Brechner Ctr. for Freedom of Info., and Tax Analysts as Amici Curiae in Support of Petitioners, FCC v. AT&T, Inc. 562 U.S. 397 (2011) (No. 09-1279) [hereinafter FCC v. AT&T Amicus Brief]. *See also In re* Adoption of Baby E.Z., 266 P.3d 702, 715-32 (Utah 2011) (Lee, J., concurring in part and concurring in the judgment); State v. Rasabout, 356 P.3d 1258, 1271-90 (Utah 2015) (Lee, J., concurring in part and concurring in the judgment).

[75] *Hard Cases*, *supra* note 7, at 190.

[76] Friederike Müller & Birgit Waibel, *Corpus Linguistics—An Introduction*, UNIVERSITY OF FREIBURG, https://www.anglistik.uni-freiburg.de/seminar/abteilungen/sprachwissenschaft/ls_mair/corpus-linguistics/corpus-linguistics-an-introduction?set_language=en [https://perma.cc/2L6D-AQXJ] (last visited June 6, 2018).

[77] FCC v. AT&T Amicus Brief, *supra* note 74, at 14-15 ("Until recently, the use of corpora was limited to lexicographers, linguists, and other researchers. But these sophisticated tools are now available to anyone with internet access.").

[78] Mark Davies, *Corpus of Contemporary American English*, BYU CORPUS, https://corpus.byu.edu/coca [https://perma.cc/TGQ2-HK69] (last visited June 6, 2018) (noting that spoken language samples only account for 20% of the corpus data).

analysis. Second, computers are central to a corpus-based approach, and the results retrieved by the database require "both quantitative and qualitative analytical techniques."[79]

While any collection of texts in a database may be categorized as a corpus[80], corpora generally fall into one of several categories. A "tagged" corpus is one in which each word is annotated with metadata, including the word's part of speech and every variation of that word within that part of speech, such that results may be filtered so as to exclude noun or adjective forms of a word when one is analyzing only the verb form.[81] Corpora can further be separated into "monitor" corpora, which are continually updated to track evolution in language usage, and "sample" corpora, which contain a fixed and unchanging body of texts.[82]

The recent use of sophisticated corpora, such as COCA, in legal opinions has grown out of the use of more "informal" corpora in recent years in response to the perceived inadequacy of dictionary usage.[83] One of the most notable examples of this was in *United States v. Costello*, a Seventh Circuit case in which the court was tasked with deciding whether a girlfriend allowing her immigrant boyfriend to live with her constituted "harboring" in the context of illegal aliens.[84] Judge Posner conducted a series of Google searches, compared the number of results, and found that

---

[79] DOUGLAS BIBER, SUSAN CONRAD & RANDI REPPEN, CORPUS LINGUISTICS: INVESTIGATING LANGUAGE STRUCTURE AND USE 4 (1998).

[80] *Hard Cases*, *supra* note 7, at 191.

[81] *Id.* at 192. For more discussion of the tagging process, see Mark Davis, *The 385+ Million Word Corpus of Contemporary American English (1990-2008+): Design, Architecture, and Linguistic Insights*, 14 INT'L J. CORPUS LINGUISTICS 159, 164 (2009), and Roger Garside, *The CLAWS Word-Tagging System*, *in* THE COMPUTATIONAL ANALYSIS OF ENGLISH: A CORPUS-BASED APPROACH 30, 33 (Roger Garside et al. eds., 1987).

[82] *Id.* The most widely available public corpus, the COCA, is a tagged/monitor corpus. Mark Davies, *Corpus of Contemporary American English*, BYU CORPUS, https://corpus.byu.edu/coca/ [https://perma.cc/TGQ2-HK69] (last visited June 6, 2018).

[83] Lauren Simpson, *#OrdinaryMeaning: Using Twitter as a Corpus in Statutory Analysis*, 2017 BYU L. REV. 487 (2017) [hereinafter *#OrdinaryMeaning*] (delineating formal and informal corpora, defining formal as those created for the purpose of corpus analysis); United States v. Costello, 666 F.3d 1040, 1044 (7th Cir. 2012) ("The selection of a particular dictionary and a particular definition is not obvious and must be defended on some other grounds . . . If multiple definitions are available, which one best fits the way an ordinary person would interpret the term?") (quoting *Looking It Up*, *supra* note 54, at 1445).

[84] *Costello*, 666 F.3d at 1043.

the verb "harbor" in this context required some level of intentional concealment from authorities.[85] In simply comparing the number of results from each search, Judge Posner conducted only quantitative analysis in support of his conclusion, illustrating one of the limitations of informal corpora.

Similarly, in *State v. Canton*, the Utah Supreme Court, which has been a driving force in the movement toward corpus linguistics in legal interpretation, was interpreting the phrase "out of the state" for the purposes of tolling the statute of limitations in a criminal case.[86] So that the court could analyze the phrase in its entirety, the court conducted a Google News search of all the instances of "out of the state" used in news stories in May 2013.[87] Of the 150 results, the court found that 27 referenced a person-state relationship, and none of these 27 referenced the abstract concept of "legal presence" (as opposed to physical presence, which is the meaning the court adopted).[88] This analysis is an improvement on the search conducted in *Costello*, in that it takes into account qualitative analytical methods (filtering results and analyzing context), even if not as fully as the capabilities of "formal" corpora allow.[89]

Recently, corpus analysis has been suggested as a method of analyzing language usage in the context of patent interpretation. This idea grew out of a standard set forth by the Supreme Court in *Nautilus, Inc. v. Biosig Instruments, Inc.*[90] In that case, the Court stated: "a patent is invalid for indefiniteness if its claims . . . fail to inform, with reasonable certainty, those skilled in the art about the scope of the invention."[91] Whether patent language defines the patent with "reasonable certainty" is a matter that can be settled by analyzing usage patterns in a specific field.[92]

---

[85] *Id.* at 1044.

[86] State v. Canton, 308 P.3d 517 (Utah 2013).

[87] *Id.* at 523 n.6.

[88] *Id.*

[89] *#OrdinaryMeaning*, *supra* note 83, at 496 ("Though perhaps not as methodologically sound as formal corpora such as COCA, informal corpora appear to have been slightly better received by courts . . .").

[90] Nautilus, Inc. v. Biosig Instruments, Inc., 134 S. Ct. 2120 (2014).

[91] *Id.* at 2124.

[92] Joseph Scott Miller, *Reasonable Certainty & Corpus Linguistics: Judging Definiteness After Nautilus & Teva*, 66 KAN. L. REV. 39, 45 (2017) ("Treating all prior art U.S. patents . . . as texts in a single corpus for computer-based analysis, the patent

2. Navigating the Corpus: Search Queries and Analytical Steps

Because the COCA is a tagged corpus, search queries are constructed around the word's part of speech and filter preferences. First, in order to "lemmatize" search results for a given word (search for every version of the word), the word is surrounded by brackets. Next, if the researcher wishes to narrow the search based on part of speech, this is done by adding COCA's designated expressions for each part of speech tag: .[n*] for nouns, .[v*] for verbs, _j* for adjectives, and _r* for adverbs. For example, the search [lovely]_*j will return every adjectival instance of any form of the adjective lovely (lovely, lovelier, loveliest).[93]

The COCA provides two basic functions that the researcher can use in analysis: (1) "collocation" data, in which the researcher can search for words that tend to occur in conjunction with the "node" word[94]; and (2) Kew Word In Context (KWIC) display, which allows the researcher to analyze a word in its natural context, rather than filtered by collocate data.[95]

The first of these, collocation, can further be divided into two categories for the purposes of corpus searching: (1) "situational" collocation data, which involves searching generally for a term's most common collocates; and (2) "collocational preference," a more dichotomous method of searching, which involves analyzing the linguistic preference of certain modifiers to collocate with one term over another similarly-situated term.[96] For both of these methods, results are displayed in "concordance lines," or "twenty-eight-word snapshots" that contain the search term in the middle of one line of context.[97]

---

system can use the tools of corpus linguistics . . . to establish an art's prevailing usage patterns as of a given date.").

[93] *See also Hard Cases*, *supra* note 7, at 195 (formatting the search for every nominal version of the word "enterprise" as "[enterprise].[n*]").

[94] SUSAN HUNSTON, CORPORA IN APPLIED LINGUISTICS 68 (2002) ("Collocation is the tendency of words to be biased in the way they co-occur").

[95] *Hard Cases*, *supra* note 7, at 197.

[96] Neal Goldfarb, *A Lawyer's Introduction to Meaning in the Framework of Corpus Linguistics*, 2018 BYU L. REV. (forthcoming 2018) (manuscript at 7-9), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2907485 [https://perma.cc/FAV7-PCCL] [hereinafter *Lawyer's Intro to Corpus Linguistics*] (using the different collocates of "strong" and "powerful" to demonstrate collocational preference).

[97] John D. Ramer, *Corpus Linguistics: Misfire or More Ammo for the Ordinary-Meaning Canon?*, 116 MICH. L. REV. 303, 317 (2017) (citing *Hard Cases*, *supra* note 7,

In structuring a corpus search, legal scholars have outlined multiple steps and considerations. For example, Stephen C. Mouritsen has described corpus analysis as a three-step process: (1) structuring the query (without including "outcome determinative" language); (2) reviewing concordance lines, paying particular attention to the competing uses that the parties proffer; and (3) reviewing collocation data.[98]

Another consideration for legal researchers is choosing the correct corpus. [99] The correct corpus must be accurately representative of the "linguistic community" which it represents, meaning the law should be interpreted so that it "reflect[s] the common usage of those it attempts to regulate."[100]

However, somewhat inconsistent with the Mouritsen method, others have suggested that the methodology should depend on the statutory term in question. Under this method, whether to conduct a collocation search, in addition to a simple KWIC search, depends on the part-of-speech of the word. While collocation searches will shed light on the ordinary meaning of adjectives, they will be ineffective on nouns that "ordinarily appear[] alone."[101]

II. PROBLEMS WITH CORPUS DATA: INCONSISTENT METHODOLOGY AND RESULTS

A. Problems with Corpus Methodology

Legal scholars have laid out a number of considerations in structuring search queries for corpus searching. For example, Stephen Mouritsen laid out some principled methods for filtering results when analyzing the use of "enterprise" in the context of the RICO statute. He used the corpus features that returned the use

---

at 197). For more in-depth COCA instruction, see the COCA instructional video series. TheGrammarLab, *COCA 101: Introduction to Using the Corpus of Contemporary American English*, YOUTUBE (Jul. 12, 2012), https://www.youtube.com/watch?v=sCLgRTlxG0Y [https://perma.cc/E5GQ-JGNR].

[98] *Hard Cases*, *supra* note 7, at 203.

[99] *The Dictionary Is Not a Fortress*, *supra* note 9, at 1956; *see also* Ramer, *supra* note 97, at 326-27.

[100] *Id.* at 1956.

[101] Ramer, *supra* note 97, at 327 (referencing *People v. Harris* as an example of this phenomenon, which is discussed further *infra* Section II.A.1).

of the nominal form of "enterprise" in the context of "concordance lines," or single lines with the node word in the middle and its context on either side. He then excluded those results that used enterprise as a "non-count" noun (NC), Organizational or Proper Name (OPN), General Effort noun (GE), and uses that were vague and indecipherable (VAG). [102]

Mouritsen also conducted analysis of the *Muscarello* case, by using corpus data to determine which of two definitions of the verb "carry" was more common.[103] However, Mouritsen's approach is only one of multiple ways of structuring search queries and analyzing corpus data. For example, Neal Goldfarb, author of the influential amicus brief in the *FCC* case, has recently conducted his own corpus analysis of *Muscarello*. However, his methodology is vastly different. Goldfarb goes far beyond looking at two possible definitions, and looks at the use of the verb "carries" in a general sense as it is used in everyday speech.[104] Furthermore, Goldfarb uses his own method of collecting and analyzing the data, called Corpus Pattern Analysis, which "focuses on multiword patterns rather than on individual word meanings."[105] Goldfarb's method requires compilation which goes far beyond that which is conducted by the database itself, which illustrates that more than a rudimentary knowledge of the corpus is required to conduct an exhaustive analysis.

While exclusion and filtering of results is a common thread in many of the corpus searches conducted by legal scholars, there is no consistency in determining which results are excluded, and almost no searches utilize the same principled steps laid out by Mouritsen or Goldfarb.[106] This leads to skewed results and variety in analysis methods, even among corpus analyses done on the

---

[102] *Hard Cases*, *supra* note 7, at 198.

[103] *The Dictionary Is Not a Fortress*, *supra* note 9, at 1958-70.

[104] *Lawyer's Intro to Corpus Linguistics*, *supra* note 96, at 3 ("[W]hen viewed without pre-conceptions, what does the corpus data tell us about how the word *carry* behaves?"); *contra The Dictionary is Not a Fortress*, *supra* note 9, at 1962 (pointing out that the issue in *Muscarello* "is not 'carries' at large, but 'carries a firearm.'").

[105] Lawyer's Intro to Corpus Linguistics, supra note 96, at 3.

[106] Carissa Hessick, *Corpus Linguistics and Criminal Law*, PRAWFSBLAWG (Sept. 6, 2017), http://prawfsblawg.blogs.com/prawfsblawg/2017/09/corpus-linguistics-and-criminal-law.html [https://perma.cc/5JY7-H8PB] ("There does not appear to be a single, correct way to conduct a database search and analysis.").

same statutory provision.[107] The following cases will show how this inconsistent methodology plays out in judicial use of corpus data.

1. *People v. Harris*

Perhaps the most famous example of the inconsistent methodology associated with judicial use of corpus linguistics comes from the Michigan Supreme Court case *People v. Harris*.[108] In that case, both the majority and dissent relied on corpus analysis in interpreting whether the term "information" encompassed false or inaccurate statements under the Michigan Disclosures by Law Enforcement Act. However, they each came to the opposite conclusion.

The majority utilized a "collocation" search in COCA to find the words that co-occurred with the term "information" most commonly within a four-word span.[109] They analyzed these results by searching through the list of collocates, and finding those that related to "truth or falsity." Through this approach, the majority found that the term "accurate" was the adjective most commonly used, but "false" and "inaccurate" were also commonly used.[110] Based on this information, the majority found that the term "information" was meant to encompass false or inaccurate statements.[111]

The dissent used the corpus in a completely different way. Justice Markman, joined by Justice Viviano, did a much simpler search, searching the COCA's sources for the noun "information." They found that it occurred 168,187 times, in the COCA's

---

[107] *See* Derek Sinko, *The Use of "Use": Legislative Intent, Plain Meaning, and Corpus Linguistics* (Feb. 4, 2015), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2560305 [https://perma.cc/GF4N-M92G]. *But see* Zachary D. Smith, *United They Hold, Divided They Might Fail: A Corpus Linguistics Analysis of the U.S. Supreme Court's Recent Ordinary Meaning Cases* 15-20 (Dec. 18, 2015), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2781672 [https://perma.cc/DD63-D9MQ] (using inconsistent methodologies in analyzing whether the use of "use" in a federal statute prohibiting the use of a gun in relation to drug trafficking included trading the firearm for drugs).

[108] 885 N.W.2d 832 (2016).

[109] *Id.* at 348.

[110] *Id.* at 348 n.33.

[111] *Id.* at 348.

sources.[112] They then did a similar collocate search as the majority, but interpreted the results differently. They found that the vast majority of the time that "information" occurred in the corpus (167,250 of the 168,187 times), it was unmodified by any adjective that indicated truth or falsity of the statements.[113] The dissent also went further than the majority by analyzing concordance lines. In analyzing "information" in its natural context, the dissent found that it is ordinarily used to connote accurate statements.[114]

However, notably, the dissent failed to explicitly state how many concordance lines it analyzed, or if results were filtered by any principled method. Above all, this case demonstrated that "reasonable judges trying their best to use the COCA can reach opposite conclusions."[115]

2. *The Utah Cases*

Justice Lee of the Utah Supreme Court, in two recent opinions, utilized corpus linguistics in concurring opinions. However, he utilized the COCA in two different capacities.

In the first case, *In Re Adoption of Baby E.Z.*,[116] the court was interpreting the term "custody" in the state Parental Kidnapping Prevention Act (the PKPA). Justice Lee, in his concurrence, conducted a thorough corpus analysis. He analyzed 500 concordance lines of the noun "custody" in the COCA, and determined that "custody" is ordinarily used in the context of divorce, not adoption.[117] After filtering out 202 results because they were used in a criminal context, he found that the concordance lines explicitly referenced divorce 146 times, and only referenced adoption 12 times.[118]

Justice Lee went on to complete a collocate analysis. He found that "custody" collocated with "divorce" 129 times, but with

---

[112] *Id.* at 850 n.14 (Markman, J., dissenting).
[113] *Id.*
[114] *Id.*
[115] Ramer, *supra* note 97, at 316.
[116] 266 P.3d 702 (Utah 2011).
[117] *Id.* at 724-25 (Lee, J., concurring in part and concurring in the judgment).
[118] *Id.* at 724 n.21.

"adoption" only 13 times.[119] He concluded that because "[t]he word 'custody' is some ten times more likely to collocate with the word 'divorce' than with the word 'adoption' in contemporary usage[,]"[120] the ordinary meaning of "custody" in this context referenced divorce.

However, Justice Lee left out an important step in his collocation analysis, making a similar mistake to the majority in *Harris*. In updating this search, I found that the term "custody" appears 9,130 times in the COCA.[121] Of these, it collocates with "divorce" 156 times, and "adoption" 18 times.[122] This means that the vast majority of the time that "custody" appears in the COCA (almost 98%), it is not used with either of these competing collocates.[123] While this may not affect the outcome of the corpus analysis as a whole, it could very well affect the reliability of a collocation search. Justice Lee narrowed the inquiry to "divorce" and "adoption," since these were the meanings proffered by the parties.[124] However, this is not to say that analysis of the overall "ordinary meaning" of the statutory term would not shed light on which of two meanings is more common, or that the researcher should skip a broader analysis of the statutory term in context.

Four years later in 2015, Justice Lee utilized corpus analysis again in *State v. Rasabout*. The case involved twelve separate felony convictions of unlawful discharge of a firearm, after the defendant fired twelve shots at a house in a gang-related drive-by shooting.[125] The question before the Utah Supreme Court was whether the verb "discharge" in the relevant Utah Code section meant firing of a single shot, or emptying of the magazine.[126]

However, in this case, Justice Lee did not conduct a general concordance analysis of the statutory term in question before

---

[119] *Id.* at 724 n.23.

[120] *Id.* at 725.

[121] *See infra* Appendix, Figure 1.

[122] *See infra* Appendix, Figure 2.

[123] *But see* People v. Harris, 885 N.W.2d 832, 850 n.14 (2016) (Markman, J., concurring in part and dissenting in part) (finding that the term "information" rarely collocated with an adjective relating to the veracity of the statements).

[124] *Baby E.Z.*, 266 P.3d at 725 (calling the interpretation of the PKPA a "contest between probabilities of meaning") (quoting Felix Frankfurter, *Some Reflections on the Reading of Statutes*, 47 COLUM. L. REV. 527, 527-28 (1947)).

[125] State v. Rasabout, 356 P.3d 1258, 1260 (2015).

[126] *Id.* at 1263.

analyzing collocates. This time, he searched for collocates within five words of the verb "discharge," and only analyzed concordance lines for "firearm" and its synonyms ("firearms," "gun," and "weapon").[127] He found 86 instances of the verb "discharge" occurring with one of these four nouns.[128] He also found that twelve of these instances clearly indicated firing of a single bullet, while only one indicated firing of multiple shots.[129] Furthermore, of the 86 instances, 36 provided "insufficient detail to indicate whether the *discharge* at issue had reference to a single shot or to the emptying of a magazine."[130]

Based on this data, Justice Lee drew a negative inference from the fact that only one instance clearly indicated the firing of multiple shots. He concluded that because "almost every conclusive instance of *discharge* of a weapon involve[d] a single shot,"[131] this was the ordinary meaning of the statutory term.

However, this analysis leaves out inferences to be drawn from those instances that were not conclusive. Almost 42% of the concordance lines that Justice Lee analyzed were inconclusive as to whether "discharge" meant a single shot or emptying the magazine. This raises important questions of the threshold required to deem one sense of a term its "ordinary meaning," especially considering the notice requirement in the context of criminal statutory law.[132]

B. Corpus Data has Produced Different Results from Standard Methods of Statutory Interpretation

A much larger problem with judicial use of corpus analysis is that it creates results that are inconsistent with traditional canons of statutory construction. For example, Stephen Mouritsen conducted an analysis of the Supreme Court's interpretation of

---

[127] *Id.* at 1281-82 (Lee, J., concurring in part and concurring in the judgment).

[128] To provide perspective on the rate of growth in the COCA, an updated search for the verb "discharge" collocating within five words of the same four nouns reveals 101 occurrences.

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

[130] *Id.*

[131] *Id.*

[132] *See* Carissa Byrne Hessick, *Corpus Linguistics and the Criminal Law*, 2018 BYU L. REV. at 5-6 (forthcoming 2018) (noting the problems that corpus frequency analysis creates for notice and accountability in the context of criminal law).

"carry a firearm" in *Muscarello v. United States*, in which the Court concluded that a statute imposing a minimum sentence for those found guilty of "carrying" a firearm in relation to a drug-trafficking crime included carrying the firearm in a vehicle, not just on the person.[133] Mouritsen's corpus analysis came to the conclusion that the Court's definition of the term was the less "ordinary" meaning, and criticized the court's unprincipled use of dictionary definitions.[134] However, as is pointed out above, one can make the same criticism in the context of corpus analysis, as its use is even less principled than dictionaries. Mouritsen argues that this inconsistency is evidence that corpus analysis comes closer to the empirical truth of the "ordinary" meaning of a statutory provision than judges' intuition.[135] However, while dictionary usage is far from principled or consistent, adding an even more inconsistent tool for analysis serves only to further ambiguate the process of statutory interpretation, rather than clarify it.

1. *Smith v. United States*: My Own Corpus Findings[136]

In this Part, I will go through my own corpus analysis (in the COCA) of the term "country" as used in *Smith v. United States*.[137] In that case, the Court, using dictionaries, based its conclusion on the presumption that the term "country," for purposes of the Federal Tort Claims Act, encapsulated areas without any form of central government.[138]

---

[133] Muscarello v. United States, 524 U.S. 125 (1998).

[134] *The Dictionary Is Not a Fortress*, *supra* note 9, at 1926-28, 1951-66 (describing the "Sense-Ranking Fallacy", in which the Court argues that the *first* definition in a dictionary indicates the ordinary sense of a word, because of its placement as the first definition).

[135] *Id.* at 1970 (describing a debate over corpus analysis an "empirical one," while judges' debate over intuitions is "metaphysical").

[136] I, by no means, claim to be an expert in conducting corpus searches for the purposes of statutory interpretation. I am presenting these brief findings to show that (1) corpus analysis has the potential to produce results inconsistent with standard methods of statutory interpretation; and (2) the amount of discretion required at each step of the analysis creates unstable variety in results.

[137] 507 U.S. 197 (1993). For a brief summary of the case, see *supra* Part I.B.1.

[138] *Id.*

In conducting my analysis, I first analyzed 100 randomly selected concordance lines of the term "country."[139] Of these 100, I discovered that 74, when viewed in context, unambiguously implied some form of central government. I further divided these 74 concordance lines into three main categories: (1) those that reference "government" specifically; (2) those that reference governmental positions or governmental bodies; and (3) those that make a miscellaneous reference, from which it could be inferred that a government existed.

I analyzed approximately ten concordance lines that fit into the first category, those that reference "government" specifically. For example:

(39) "The price of rice has risen 30 percent in the country since January. In response, the government has flooded markets with subsidized rice . . . ."

I analyzed approximately thirty-three concordance lines that fit into the second category, those that reference governmental positions or governmental bodies. For example:

(55) "By midweek President Obama came here to Tucson to try to heal the fresh wounds of a grieving and confused community, to honor the fallen and try to unify the country."

The remainder of the 74 concordance lines required some form of inference that a government existed, such as reference to citizens, civil rights, or lawsuits. Here are two examples:

(93) "In New Orleans and in other places, many of our fellow citizens have felt excluded from the promise of our country."

(75) "The supreme irony is that China, a country with well-documented civil rights abuses, intolerance for free speech and persecution of religion, will hold the keys to this country's economy for decades.

Of the twenty-six concordance lines that did not make reference to government, there were two broad categories. The first was the use of "country" to reference a specific area with a certain defining characteristic, such as (1) "cactus country," or (25) "hog country." The second was the use of "country," outside the context of government, to define the boundaries of the topic being

---

[139] Mark Davies, *Corpus of Contemporary American English*, BYU CORPUS, https://corpus.byu.edu/coca [https://perma.cc/QTB6-6ZTP] (last visited Mar. 9, 2018).

discussed. For example: (23) "Boeheim's team was placed in the South Region, a sort of miniature tournament for some of the best foul differentials in the country."

My collocate analysis of the term "country" revealed a different story. First I searched for the most common collocates within four words of any noun form of "country" (by using the search "[country]_*nn"). The results indicated that, of the twenty most common collocates, only two ("communist" and "illegally") unambiguously suggested a form of government.[140]

I then conducted collocate analysis of the term "nation," a commonly cited synonym for "country," using the same method.[141] This collocate search indicated that, of the twenty most common collocates for any noun form of "nation," six suggested a form of government.[142]

The lack of government-related collocates for the term "country" can either be analyzed as supporting or undermining the Court's holding in *Smith*. While one side would argue it shows that "country" is not ordinarily used in the context of government (supporting *Smith*), the other side would argue that the entirety of this corpus analysis indicates that government is included in the category covered by "country" itself, and therefore no modifiers are necessary.

Lastly, I utilized the chart feature of the COCA to analyze how the term "country" (in all of its noun forms) is distributed among the different genres of speech.[143] It is worth noting that the results indicated the term overwhelmingly occurs most in spoken word, having a higher frequency and "per million" number than any other genre.[144] This step in the analysis is consistently overlooked by researchers utilizing corpus linguistics, but it can have a major impact on corpus findings, as language use is often

---

[140] *See infra* Appendix, Figure 3 (showing collocates of the term "country").

[141] Ethan J. Herenstein, *The Faulty Frequency Hypothesis: Difficulties in Operationalizing Ordinary Meaning Through Corpus Linguistics*, 70 STAN. L. REV. ONLINE 112, 121 (2017) [hereinafter *The Faulty Frequency Hypothesis*] (suggesting the searching of synonyms as a "methodological adjustment" as a way to improve on the flaws of the frequency analysis, and conduct a more robust search).

[142] *See infra* Appendix, Figure 4 (showing collocates of the term "nation").

[143] Analyzing chart data is a way to "incorporat[e] prevalence and newsworthiness of the underlying phenomena into [the] corpus analysis." *The Faulty Frequency Hypothesis, supra* note 141, at 120.

[144] *See infra* Appendix, Figure 5 (showing the Chart results for "country").

vastly different in oral communication than in written.[145] The researcher also must consider this information in light of the fact that 80% of the COCA comes from printed language, and only 20% comes from spoken conversations via television and radio.[146]

This rudimentary corpus search was designed to point out the extremely high level of discretion at each stage of the corpus analysis, and the unpredictable variety in results that such a level of discretion can cause.

III. CORPUS LINGUISTICS SHOULD NOT BE RAISED *SUA SPONTE* IN JUDICIAL OPINIONS

A. Why Corpus Linguistics is Different: "Deceptive Empiricism"

When courts use dictionaries, they generally go through several steps, each of which involves some level of discretion: (1) choosing the word to be defined; (2) selecting the proper type of dictionary; (3) selecting a specific dictionary within that type; (4) selecting the appropriate edition; and (5) choosing the correct definition.[147] While judges' use and analysis of dictionaries is far from consistent, a judge conducting a robust analysis of dictionary definitions is still working within a relatively contained linguistic universe. For this reason, judges, as well as litigators and the public, have an idea of the limitations imposed by dictionaries.[148]

---

[145] *See* Diane L. Schallert, Glenn M. Kleiman & Ann D. Rubin, *Analyses of Differences Between Written and Oral Language*, UNIV. OF ILL. AT URBANA-CHAMPAIGN at 1 (Apr. 1977), https://www.ideals.illinois.edu/bitstream/handle/2142/17970/ctrstreadtechrepv01977i00029_opt.pdf [https://perma.cc/T7WN-E8VB] (noting differences between written and oral speech, including "the greater precision and detail found in writing, the greater amount of repetition found in speech, and differences caused by the availability of prosody (intonation, stress and rhythm) in speech but not writing").

[146] *#OrdinaryMeaning*, *supra* note 83, at 500-01 (noting the bias created by the COCA in favor of "speakers who have higher education, who write as part of their employment, or who are interviewed by the media").

[147] *The Lexicon Has Become a Fortress*, *supra* note 51, at 264; *see also Lawyer's Intro to Corpus Linguistics*, *supra* note 95, at 31 ("[D]istinguishing between word senses often depends at least in part on the exercise of judgment and discretion by the lexicographer.").

[148] *See War of the Words*, *supra* note 49, at 191, 198; *see also* Antonin Scalia & Bryan A. Garner, *A Note on the Use of Dictionaries*, 16 GREEN BAG 2d 419, 422-23 (2013) (outlining limiting principles to use when consulting dictionaries).

On the other hand, as my analysis in Part II.B.1 indicates, a corpus search involves literally hundreds of levels of discretion, and involves an almost unending, and deceptively empirical, linguistic universe. Not only must the researcher analyze hundreds of concordance lines, and make inferences therefrom, but also must decide when to use certain functions available in the corpus. For example, of the 74 concordance lines of "country" that suggested some form of government, about ten required that I use the "expanded context" feature to look at the entire paragraph of text (rather than just one line), while the others did not.[149]

This "deceptive empiricism" is the core of the danger of corpus analysis, largely because frequency analysis does not necessarily provide the researcher with a term's ordinary usage. Indeed, "a word may be invoked more frequently in one sense than another for reasons that have little to do with the common understanding of that word."[150] For example, Ethan J. Herenstein laid out alternative reasons for a usage's high frequency within a corpus that have nothing to do with that term's ordinary usage: (1) the prevalence of the "underlying phenomenon that the term denotes"; and (2) the newsworthiness of the underlying phenomenon. [151] Thus, corpus-users often make the wrong inferences from the frequency data provided by the corpus: just because there is more opportunity and incentive for news outlets and other genres to cover one sense of a term, making that sense more frequent within the corpus, this is not necessarily reflective of how an ordinary American speaker of English ordinarily *understands* the term.

---

[149] Another example of a step in the analysis I could have added would be to do the same search in another corpus (e.g., the Corpus of U.S. Supreme Court Opinions), and compare the results.

[150] *The Faulty Frequency Hypothesis*, *supra* note 141, at 117; *see also* Lawrence B. Solum, *Triangulating Public Meaning: Corpus Linguistics, Immersion, and the Constitutional Record* (Apr. 26, 2017), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3019494 [https://perma.cc/288S-YGRR] ("Corpus data may tell us something about the relative frequency of the various meanings, but the most frequent meaning is not necessarily the meaning in context.").

[151] *The Faulty Frequency Hypothesis*, *supra* note 141, at 116-20 (using *Rasabout* as an example of these two related alternative reasons: if it is more common for a person discharging a firearm to fire one bullet rather than empty the chamber, there are fewer opportunities for the news outlets/genres reflected in a corpus to use the latter form).

Lastly, the researcher also has discretion in deciding how to structure the search from the start. The researcher can either (1) narrow the search by focusing on the most common of two collocates; (2) do a broader search that seeks to find how a term is generally used in everyday speech (focusing on KWIC searching); or (3) both.[152] Which one of these options the researcher chooses will determine how to conduct the process going forward. Because of the amount of discretion involved at each stage of reviewing corpus data, a judge who raises corpus data *sua sponte* in an opinion is going well beyond the interpretive task.

B. Judicial Notice

In addition to problems arising from inconsistency in corpus methodology, and the "deceptive empiricism" of statistical frequency data, the use of corpus linguistics in legal opinions is foreclosed by the rules governing facts that are properly judicially noticed. The Model Code of Judicial Conduct, which many states have adopted in relevant part[153], states that "[a] judge shall not investigate facts in a matter independently, and shall consider only the evidence presented *and any facts that may properly be judicially noticed.*"[154] This rule is supplemented by the Federal Rules of Evidence, which limit a judge's judicial notice power to "adjudicative fact[s] only, not legislative fact[s]."[155]

The core of the argument advocating for judicial use of corpus data is that "[i]ndependent investigation is foreclosed only as to 'facts,' not law."[156] Similarly, proponents of corpus data employ the

---

[152] *See supra* note 103; *see also supra* Part II.A.1 (citing *People v. Harris* as an example of the different ways judges can utilize corpus linguistics).

[153] *Comparison of ABA Model Judicial Code and State Variations*, AM. BAR ASS'N (Dec. 7, 2017), https://www.americanbar.org/content/dam/aba/administrative/professional_responsibility/2_9.authcheckdam.pdf [https://perma.cc/5DDN-VS4N] (comparing state proposals or rules as adopted to Canon 2.9(c) of the ABA Model Code).

[154] MODEL CODE OF JUDICIAL CONDUCT, r. 2.9(c) (AM. BAR ASS'N 2011) (emphasis added).

[155] FED. R. EVID. 201.

[156] State v. Rasabout, 356 P.3d 1258, 1284 (2015) (Lee, J., concurring in part and concurring in the judgment).

same rationale that courts use for dictionary usage: judicial notice of "the meaning of words in the vernacular language."[157]

Therefore, the core of the question about corpus linguistics is whether it may be judicially noticed by the judge. While Federal Rule of Evidence 201 does not deal directly with judicial notice of legislative facts, judges have liberally taken notice of such facts in their capacity as statutory interpreters.[158] Legislative facts have been defined as "facts, truths, and assumptions a judge considers when faced with the task of creating law."[159] There is no question that judges may act upon "investigation of the pertinent general facts, social, economic, political, *or scientific*," given they provide the "factual grounds therefor."[160]

However, corpus linguistics goes far beyond the category of information that is properly relied upon by judges in interpreting statutes. While judges have considered scientific studies in the capacity of interpreting doctrine,[161] these rest on consideration of facts proffered by scientists. There is no precedent for a judge's reliance on legislative, scientific data that she conducted herself.

While some scholars have argued in favor of independent judicial research to aid in accurately assessing scientific evidence (in the context of *Daubert* admissibility decisions)[162], this is a wholly different proposition from judges conducting scientific inquiry for the purpose of statutory interpretation.

---

[157] Ramer, *supra* note 97, at 323 ("The justification for using dictionaries sua sponte justifies using the COCA sua sponte.") (quoting Brown v. Piper, 91 U.S. 37, 42 (1875)).

[158] FED. R. EVID. 201(a). The advisory committee's note states: "In determining the content or applicability of a rule of domestic law, the judge is unrestricted in his investigation and conclusion . . . . This is the view that should govern judicial access to legislative facts." *Id.* (quoting Edmund M. Morgan, *Judicial Notice*, 57 HARV. L. REV. 269, 270-71 (1944)); *see also Rasabout*, 356 P.3d at 1284 (Lee, J., concurring in part and concurring in the judgment) ("In performing the core function of deciding what the law is or should be, we cannot properly be restricted from consulting sources that inform our understanding.").

[159] Charles Edward Suffling, *Judicial Notice*, 48 MISS. L.J. 919, 920 (1977)).

[160] MCCORMICK ON EVIDENCE § 331 (7th ed. 2013) (emphasis added).

[161] *See, e.g.*, Durham v. United States, 214 F.2d 862, 872 (1954) (considering scientific inquiries on psychiatric learning to analyze the scientific soundness of the right-and-wrong test of criminal insanity, noting "the fact finder should be free to consider all information advanced by relevant scientific disciplines.").

[162] Edward K. Cheng, *Independent Judicial Research in the Daubert Age*, 56 DUKE L.J. 1263, 1280-82 (2007).

Furthermore, corpus data falls under a unique category of information, in that while corpus analysis is a scientific endeavor[163], many judges attempt to claim an extent of expertise in the area. In their mind, corpus analysis is a means to accomplish an analysis in a field in which they *are* experts: interpreting the law.[164] While a judge may have an expertise in interpreting a statutory provision, she does not have the expertise to decide how a related branch of scientific analysis contributes to that task. Conflating the task of judges to interpret the law with analysis of scientific corpus data can lead to judges "proffer[ing] data that has only the appearance of careful empiricism."[165]

Lastly, judicial reliance on corpus data also raises concerns about the independence and accountability of the judiciary. The first canon in the Code of Conduct for United States Judges requires judges to uphold "the integrity and independence of the judiciary." [166] A judge who relies on corpus data "skirt[s] responsibility for [her] interpretations," by shifting the interpretive inquiry to statistical frequency analysis. [167] While

---

[163] State v. Rasabout, 356 P.3d 1258, 1265 (2015) (stating that "[l]inguistics is a scientific field of study that uses empirical research to draw findings") (citing TONY MCENERY & ANDREW HARDIE, CORPUS LINGUISTICS: METHOD, THEORY AND PRACTICE (2012) (calling corpus linguistics an "empirical, scientific enterprise")).

[164] *Rasabout*, 356 P.3d at 1285 (Lee, J., concurring in part and concurring in the judgment) ("So I concede the point that judges will not bring to bear the kind of training possessed by 'linguistic experts' . . . . But that, respectfully, is not the point. We judges *are* experts on one thing—interpreting the law."); *see also* Stephen C. Mouritsen, *Corpus Linguistics in Legal Interpretation—An Evolving Interpretive Framework*, 6 INT'L J. LANGUAGE & L. 67, 85 (2017) ("Though judges are generalists with respect to many of the issues that come before them, they are expected to be specialists, even experts, with respect to interpretive tasks.").

[165] *Judging Ordinary Meaning*, *supra* note 9, at 866.

[166] JUDICIAL CONFERENCE, U.S. COURTS, CODE OF CONDUCT FOR UNITED STATES JUDGES, Canon 1 cmt. (2014), http://www.uscourts.gov/judges-judgeships/code-conduct-united-states-judges [https://perma.cc/XW4Y-9H4G]
("Deference to the judgments and rulings of courts depends on public confidence in the integrity and independence of judges.").

[167] Carissa Hessick, *Corpus Linguistics and Criminal Law*, PRAWFSBLAWG (Sept. 6, 2017), http://prawfsblawg.blogs.com/prawfsblawg/2017/09/corpus-linguistics-and-criminal-law.html [https://perma.cc/E92L-UBTW] (noting that, in the context of criminal law, corpus analysis "reframe[es] the question as a dispute over database searches rather than a decision about punishment").

dictionaries, when used correctly,[168] do not compromise the integrity of independent judicial interpretation, reliance on deceptively empirical corpus data lifts too much of the interpretive burden off the judge's shoulders.

C. Adversarial Process

The adversarial process is one of the most sacred protections of the American scheme of justice. At the crux of the argument for adversarial process is the proposition that conflicting proofs presented by adversaries are more likely to yield reliable information on which the decision-maker can rely in making a decision.[169] The adversarial theory is grounded in a tripartite system of (1) the "neutral and passive fact finder," (2) "party presentation of the evidence," and (3) "highly structured forensic procedure" consisting of procedural, evidentiary, and ethical rules.[170] *Sua sponte* use of corpus linguistics leads to the very outcome the adversarial system was designed to prevent: "[losing] on appeal on a ground that [the defendant] had no opportunity to address."[171]

A recent decision in the Utah Supreme Court seems to align with this view of corpus linguistics. In *Fire Insurance Exchange v. Oltmanns*, the Utah Supreme Court was reviewing a declaratory judgment brought by the insurer in an insurance dispute.[172] The insured's brother had wrecked a jet ski, and the court of appeals ultimately held the claim to be covered under the policy.[173] When the insured then brought a counterclaim for attorney's fees, arguing the declaratory judgment had been brought in bad faith, the court was tasked with deciding whether the claim was "fairly

---

[168] *See Looking It Up*, *supra* note 54, at 1452 (describing the proper role of dictionaries as the beginning of the interpretive process, and used only to the extent they "further particular interpretive goals").

[169] Stephan Landsman, *A Brief Survey of the Development of the Adversary System*, 44 OHIO ST. L.J. 713, 714 (1983); *see also* Jerold H. Israel, *Cornerstones of the Judicial Process*, 2-SPG KAN. J.L. & PUB. POL'Y 5, 13 (1993) ("[S]elf-interested adversaries will uncover and present more useful information . . . than would be developed by the judicial officer in an inquisitorial system.").

[170] Landsman, *supra* note 169, at 714-17.

[171] State v. Rasabout, 356 P.3d 1258, 1264-66 (2015).

[172] Fire Ins. Exch. v. Oltmanns, No. 20160304, 2017 WL 5623415 at *1 (Utah Nov. 21, 2017).

[173] *Id.*

debatable" in the first place.[174] The Supreme Court ultimately affirmed summary judgment in favor of the insurer, but in a concurring opinion, Justice Durham criticized the insurer's lawyer for not providing "substantial evidence as to how a layman reading the contract would interpret 'jet ski.'"[175] Justice Durham suggested that corpus data should be presented *by the parties* so that the court can have "meaningful tools" at its disposal for interpretive tasks.[176] If the parties present this sort of empirical evidence to the court, the adversarial process is not compromised, because the judge is now performing one of her established tasks: assessing the reliability of conflicting proofs brought before the court.[177]

The response to this argument is that judges using corpus data are addressing the same issue as parties addressed in their briefs: the ordinary meaning of statutory language. They further argue for corpus data by drawing an analogy to the common practice of citing authority in opinions that neither party addressed in their briefs.[178]

However, the fact that corpus analysis is conducted for the broader purpose of interpreting statutory language, a task that both adversaries conduct in their briefs, is immaterial in this instance. The real issue is the *category* of argument the judge puts forth through corpus analysis.[179]

While the Code of Conduct for United States Judges does not squarely address the issue of *sua sponte*/ex parte reliance on scientific evidence, it does address the analogous issue of ex parte communications with "disinterested experts on the law." The Code provides that if a judge consults such an expert, he or she must

---

[174] *Id.*

[175] *Id.* at 16 n.9 (criticizing the use of only websites, boat reviews, and Wikipedia articles).

[176] *Id.* ("[L]awyers should provide courts with meaningful tools using the best available methods when the court is tasked with determining ordinary meaning. . . . These tools for empirical analysis are readily available *to lawyers* and should be used when appropriate.") (emphasis added).

[177] *See* Landsman, *supra* note 169.

[178] Ramer, *supra* note 97, at 323 (citing State v. Rasabout, 356 P.3d 1258, 1284 n.35 (2015)) (Lee, J., concurring in part and concurring in the judgment).

[179] *Rasabout*, 356 P.3d at 1264 ("But because [the judge's] rationale is so different in kind from any argument made by the parties, Mr. Rasabout has never had a reasonable opportunity to present a different perspective.").

"giv[e] . . . notice to the parties of the person . . . consulted and the subject matter of the advice[,] and afford[] the parties reasonable opportunity to . . . respond . . . ."[180] In the context of *sua sponte* judicial use of corpus analysis, even if one were to treat the corpus as a "disinterested expert," the judge would still have a responsibility to give the parties a reasonable opportunity to respond. Such an opportunity is not afforded if the first mention of corpus data is in the opinion itself.

Some scholars have moved for an expansion of this rule to give judges an opportunity to consult not only "disinterested experts," but also to "search for and read research material and other literature, not presented or cited by the parties, concerning issues of science or technology directly applicable or relevant to a pending or impending proceeding before the judge."[181] However, even if the rule were expanded to include this type of research, there would still be a requirement that the judge "gives notice to the parties of the material and literature consulted and, in a manner within the judge's discretion, affords them reasonable time to comment and submit other relevant material."[182]

CONCLUSION

While judges should be free to carry out their judicial duties without unnecessary hindrances, to allow *sua sponte* use of corpus linguistics in judicial opinions would be to unduly stretch the interpretive task. There is no doubt that analysis of corpus data, at least to some extent, is a scientific endeavor. My analysis of its judicial use up to this point, as well as my own brief corpus findings, show that the variables involved create a rate of deviation that is simply inconsistent with the American adversarial process.

---

[180] JUDICIAL CONFERENCE, U.S. COURTS, CODE OF CONDUCT FOR UNITED STATES JUDGES, Canon 3A(4)(c) (2014), http://www.uscourts.gov/judges-judgeships/code-conduct-united-states-judges [https://perma.cc/XW4Y-9H4G].

[181] George D. Marlow, *From Black Robes to White Lab Coats: The Ethical Implications of a Judge's* Sua Sponte, Ex Parte *Acquisition of Social and Other Scientific Evidence During the Decision-Making Process*, 72 ST. JOHN'S L. REV. 291, 333, 334 (1998) (laying out Judge Marlow's proposed amendment).

[182] *Id.*

This is not to say that there is no place for corpus linguistics in the field of law. Indeed, the development of the marriage between law and corpus linguistics is inevitable.[183] However, to be consistent with the adversarial process, this development must start with lawyers.[184] Furthermore, the adversarial use of corpus linguistics should be accompanied by other methodologies for a more holistic approach, with each approach "confirming or questioning" the others.[185] In sum, the future of corpus linguistics depends on lawyers and scholars developing processes that will minimize the variables involved, and allow for data that "satsif[ies] the highest values of the scientific method."[186]

*Daniel C. Tankersley**

---

[183] Neal Goldfarb, *Some Comments on Hessick on Corpus Linguistics (Updated)*, LAWNLINGUISTICS (Sept. 13, 2017), https://lawnlinguistics.com/2017/09/13/some-comments-on-hessick-on-corpus-linguistics/#more-1399 [https://perma.cc/5H49-PLAF].

[184] *See generally Lawyer's Intro to Corpus Linguistics*, *supra* note 96; *see also Rasabout*, 356 P.3d at 1283 n.32 (Lee, J., concurring in part and concurring in the judgment) ("There is a bit of a chicken-and-egg problem in the complaint about lack of briefing. Until judges convey openness to analysis using a corpus like COCA, lawyers will not present it. My opinion is aimed at opening the door to better briefing.").

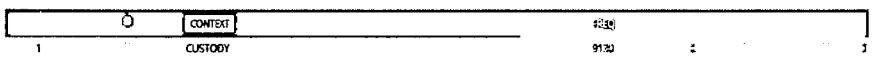
[185] Solum, *supra* note 150 (describing a method of "triangulating" ordinary meaning using three methods: (1) corpus linguistics, (2) immersion in the "linguistic and conceptual world," and (3) "studying the constitutional [or legislative] record").

[186] Stephen C. Mouritsen, *Corpus Linguistics in Legal Interpretation*, 6 INT'L J. LANGUAGE & L. 67, 86 (2017) ("[T]he promise of the LCL movement is that when such answers come, they will be grounded . . . in empirical data gathered through experiments that are both replicable and falsifiable.").

* Staff Editor, *Mississippi Law Journal*; J.D. Candidate 2019, University of Mississippi School of Law. The author would like to thank Professor Matthew Hall for his advice and guidance. He would also like to thank his friends and family for their support and encouragement during the writing process.

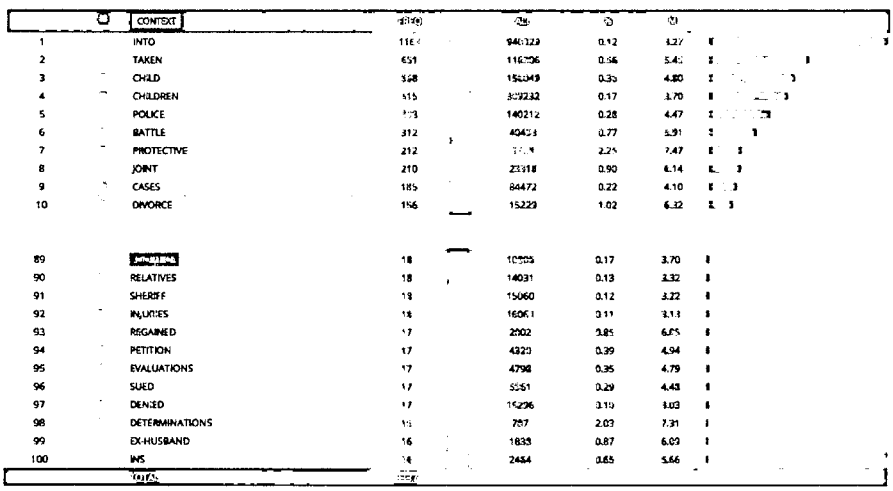
APPENDIX: CORPUS RESULTS

Figure 1



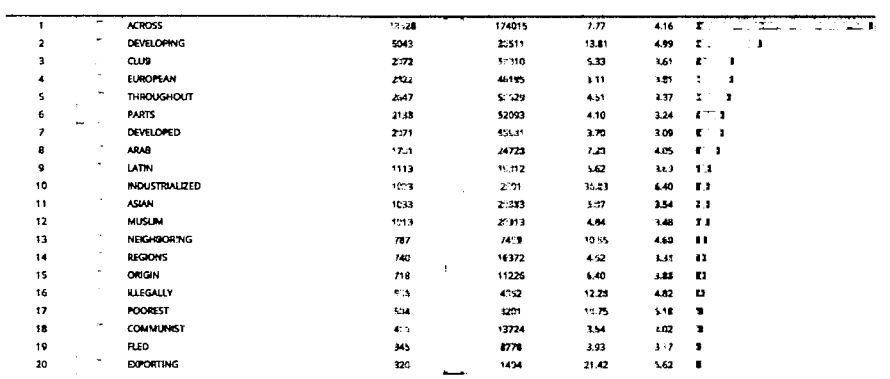
		CONTEXT			
1		CUSTODY		9130	

Figure 2



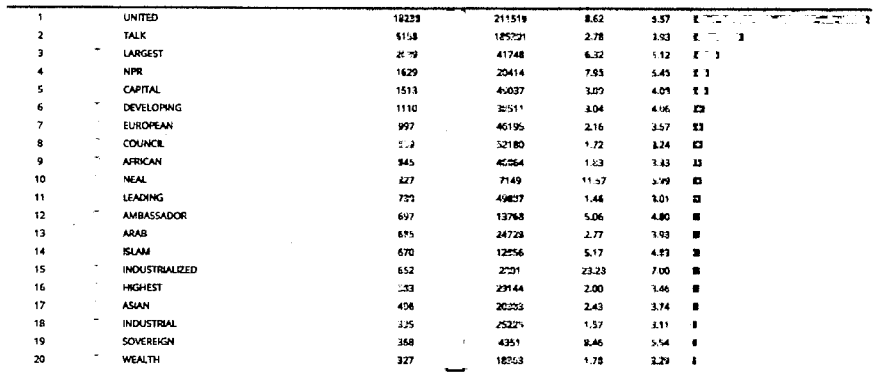
		CONTEXT				
1		INTO	11[illegible]	946329	0.12	3.27
2		TAKEN	651	116706	0.56	5.48
3		CHILD	568	156349	0.36	4.80
4		CHILDREN	515	309232	0.17	3.70
5		POLICE	393	140212	0.28	4.47
6		BATTLE	312	40423	0.77	5.91
7		PROTECTIVE	212	[illegible]	2.25	7.47
8		JOINT	210	23318	0.90	6.14
9		CASES	185	84472	0.22	4.10
10		DIVORCE	156	15229	1.02	6.32
89		[illegible]	18	10505	0.17	3.70
90		RELATIVES	18	14031	0.13	3.32
91		SHERIFF	18	15060	0.12	3.22
92		INJURIES	18	16061	0.11	3.13
93		REGAINED	17	2002	0.85	6.05
94		PETITION	17	4320	0.39	4.94
95		EVALUATIONS	17	4798	0.35	4.79
96		SUED	17	5551	0.29	4.43
97		DENIED	17	15436	0.10	3.03
98		DETERMINATIONS	15	757	2.03	7.31
99		EX-HUSBAND	16	1833	0.87	6.09
100		INS	16	2454	0.65	5.66

Figure 3



1		ACROSS	[illegible]28	174015	7.77	4.16
2		DEVELOPING	5043	25511	13.81	4.99
3		CLUB	2772	[illegible]310	5.33	3.61
4		EUROPEAN	2522	46195	3.11	3.81
5		THROUGHOUT	2447	[illegible]429	4.51	3.37
6		PARTS	2138	52093	4.10	3.24
7		DEVELOPED	2071	65531	3.70	3.09
8		ARAB	1721	24723	7.21	4.05
9		LATIN	1113	[illegible]312	5.62	3.83
10		INDUSTRIALIZED	1093	2701	35.23	6.40
11		ASIAN	1033	21383	5.07	3.54
12		MUSLIM	1013	20913	4.84	3.48
13		NEIGHBORING	787	7459	10.55	4.69
14		REGIONS	740	16372	4.52	3.51
15		ORIGIN	718	11226	6.40	3.85
16		ILLEGALLY	615	4752	12.28	4.82
17		POOREST	514	3201	16.75	5.18
18		COMMUNIST	495	13724	3.54	4.02
19		FLED	345	8778	3.93	3.17
20		EXPORTING	320	1494	21.42	5.62

Figure 4



1		UNITED	18233	211519	8.62	9.57	
2		TALK	6158	185721	2.78	3.93	
3	-	LARGEST	2[illegible]9	41748	6.32	9.12	
4		NPR	1629	20414	7.93	5.45	
5		CAPITAL	1513	45037	3.09	4.09	
6	-	DEVELOPING	1110	32511	3.04	4.06	
7		EUROPEAN	997	46195	2.16	3.57	
8	-	COUNCIL	[illegible]	32180	1.72	3.24	
9	-	AFRICAN	945	46064	1.83	3.83	
10		NEAL	827	7149	11.57	[illegible]	
11		LEADING	732	49857	1.46	3.01	
12	-	AMBASSADOR	697	13768	5.06	4.80	
13		ARAB	695	24728	2.77	3.93	
14		ISLAM	670	12556	5.17	4.83	
15	-	INDUSTRIALIZED	652	2791	23.23	7.00	
16		HIGHEST	633	29144	2.00	3.46	
17		ASIAN	406	20253	2.43	3.74	
18	-	INDUSTRIAL	335	25227	1.57	3.11	
19		SOVEREIGN	368	4351	8.46	5.54	
20	-	WEALTH	327	18263	1.78	3.29	

Figure 5

SECTION (CLICK FOR SUB-SECTIONS) (SEE ALL SECTIONS AT ONCE)	FREQ	SIZE (M)	PER MIL	CLICK FOR CONTEXT (SEE ALL)
SPOKEN	833,920	109.4	7,142.87	
FICTION	135,592	104.9	1,211.48	
MAGAZINE	436,312	110.1	3,717.91	
NEWSPAPER	550,984	106.0	4,876.16	
ACADEMIC	527,064	103.4	4,730.83	
1990-1994	540,048	104.0	5,192.81	
1995-1999	417,896	103.4	4,039.70	
2000-2004	448,328	102.9	4,355.21	
2005-2009	414,328	102.0	4,060.40	
2010-2104	421,288	102.9	4,093.71	
2015-2017	241,984	62.3	3,883.60	
TOTAL	2,483,872			SEE ALL TOKENS