Searching for Law in All the Wrong Places

Evan C. Zoldan

Recommended Citation

Zoldan, Evan C., "Searching for Law in All the Wrong Places" (2021). *Minnesota Law Review: Headnotes*. 80.

https://scholarship.law.umn.edu/headnotes/80

Response

Searching for Law in All the Wrong Places

Evan C. Zoldan[†]

INTRODUCTION

In a recent issue of the University of Chicago Law Review, Justice Thomas R. Lee and Stephen C. Mouritsen defend their use of corpus linguistics techniques for interpreting legal texts against a number of serious critiques.[1] For squarely addressing criticism of their work, they should be lauded: no matter whether they or their critics are convinced during the course of this debate, the exchange of views itself has the potential to clarify important points of disagreement.

It is in the spirit of this scholarly exchange that I respond to Lee & Mouritsen's article, *The Corpus and the Critics*,[2] including answering some objections they raise about my previous work on corpus linguistics and legal interpretation.[3] But, rather than rehash arguments I previously have made, or arguments ably made by Professor Tobia in his response to Lee & Mouritsen,[4] I will offer a few thoughts that I hope will move the conversation forward. Specifically, this Essay responds to Lee & Mouritsen's suggestion that legal interpreters can use a corpus of legislative history to cure any deficiencies in the composition of a general corpus—that is, a corpus that includes a variety of types of texts. Contrary to Lee & Mouritsen's suggestion, a corpus of legislative history does not allow an interpreter to learn how statutory terms are

1. *E.g.*, Kevin Tobia, *Testing Ordinary Meaning*, 134 Harv. L. Rev. 726 (2020); Anya Bernstein, *Democratizing Interpretation*, 60 Wm. & Mary L. Rev. 435 (2018); Carissa Byrne Hessick, *Corpus Linguistics and the Criminal Law*, 2017 BYU L. Rev. 1503 (2017).

2. Thomas R. Lee & Stephen C. Mouritsen, *The Corpus and the Critics*, 88 U. Chi. L. Rev. 275 (2021).

3. Evan C. Zoldan, *Corpus Linguistics and the Dream of Objectivity*, 50 Seton Hall L. Rev. 401 (2019).

4. Kevin Tobia, *The Corpus and the Courts*, U. Chi. L. Rev. Online (Mar. 5, 2021), https://lawreviewblog.uchicago.edu/2021/03/05/tobia-corpus/ [https://perma.cc/B8SJ-48DD].

used in legal texts because legislative history is not written in statutory language. Moreover, Lee & Mouritsen's concession—that searching different bodies of text may be appropriate in different circumstances—raises, but does not resolve, a crucial predicate issue: how to choose the appropriate corpus. Lee & Mouritsen's suggestions about how to choose a corpus are in tension with one another and, ultimately, unsatisfactory.

A. A CORPUS OF LEGISLATIVE HISTORY

Corpus linguistics is a set of procedures or methods that help researchers analyze linguistic information in bodies of text.[5] Although the study of linguistic information is not new,[6] corpus linguistics has taken on new importance as electronic storage and retrieval systems allow users to search data sets, or corpora, that can be searched manually only with great difficulty.[7] This wealth of electronic data, which has opened up new possibilities for researchers, has been particularly enticing to legal interpreters. A growing number of scholars,[8] and even some judges,[9] have turned to corpus linguistics techniques when searching for the original public meaning of the text of the Constitution or the ordinary meaning of statutory language.

In their work on corpus linguistics, Lee & Mouritsen have searched for the meaning of statutory language nearly exclusively in general corpora, that is, corpora that contain a balance of different types of texts that are meant to approximate "ordinary" speech.[10] For

---

5. For a description of corpus linguistics, see GRAEME KENNEDY, AN INTRODUCTION TO CORPUS LINGUISTICS 1 (2014); K. Kredens & M. Coulthard, *Corpus Linguistics in Authorship Identification*, in THE OXFORD HANDBOOK OF LANGUAGE AND LAW 504, 504–05 (Peter M. Tiersma & Lawrence M. Solan eds., 2016) (describing different definitions of "corpus" and "corpus linguistics").

6. KENNEDY, *supra* note 5.

7. Peter Tiersma & Lawrence Solan, *Introduction*, in OXFORD HANDBOOK OF LANGUAGE AND LAW, *supra* note 5, at 3.

8. Stefan Th. Gries & Brian G. Slocum, *Ordinary Meaning and Corpus Linguistics*, 2017 BYU L. REV. 1417, 1441 (2017); Lee J. Strang, *How Big Data Increases Originalism's Methodological Rigor: Using Corpus Linguistics to Recover Original Language Conventions*, 50 U.C. DAVIS L. REV. 1181 (2017) (using corpus search to determine meaning of constitutional text); James C. Phillips, Daniel M. Ortner & Thomas R. Lee, *Corpus Linguistics & Original Public Meaning: A New Tool to Make Originalism More Empirical*, 126 YALE L.J. FORUM 21, 24–26 (2016); Jennifer L. Mascott, *The Dictionary as a Specialized Corpus*, 2017 BYU L. Rev. 1557 (2017).

9. *In re* Adoption of Baby E.Z., 266 P.3d 702, 727 (Utah 2011) (Lee, J., concurring); People v. Harris, 885 N.W.2d 832, 833–34 (Mich. 2016); American Bankers Ass'n v. Nat'l Credit Union Admin., 306 F. Supp. 3d 44, 68 (D.D.C. 2018).

10. *E.g.*, Lee & Mouritsen, *supra* note 2, at 356–57 (describing corpus search in

example, they often rely on *The Corpus of Contemporary American English*, which contains more than a billion words from a variety of sources, including novels, magazines, newspapers, and transcriptions of spoken words.[11] In my previous work on law and corpus linguistics, I considered whether statutory terms should be interpreted to conform to the meaning of language found in a general corpus.[12] Because of the pervasive differences between nonlegal language and statutory language, I concluded that it is not appropriate for legal interpreters to interpret statutory language as if it were nonlegal language.[13]

In *The Corpus and the Critics*, Lee & Mouritsen acknowledge the many differences between statutory language and the language found in nonlegal sources; indeed, they concede that these differences "sometimes call into question the probity of evidence from a general corpus."[14] They contend, however, that when it is not appropriate to search a *general* corpus for statutory meaning, an interpreter can instead search a corpus of *legislative history*.[15] The corpus of legislative history to which Lee & Mouritsen refer was still hypothetical at the time of their writing.[16] Nevertheless, they predict that using corpus linguistics tools to search a corpus of legislative history will allow interpreters to learn the specialized legal meaning, if any, of a statutory term.[17] But, even when this hypothetical corpus comes to fruition, it will not help interpreters find the meaning of statutory language. Because language found in a corpus of legislative history is unlike statutory language, examining a corpus of legislative history will not reveal the specialized legal meaning of statutory language.

---

the Corpus of Contemporary American English).

    11. CORPUS CONTEMPORARY AM. ENG., https://www.english-corpora.org/coca/ [https://perma.cc/88G4-5RLU] (last visited Sept. 13, 2021).

    12. Zoldan, *supra* note 3, at 444–45.

    13*. Id.*

    14. Lee & Mouritsen, *supra* note 2, at 301.

    15*. Id.* at 294. Lee and Mouritsen do not describe what such a corpus might contain.

    16*. Id.* at 301 ("BYU Law is in the initial stages of developing a corpus of legislative history.").

    17*. Id.*

When they equate statutory language to the text in a corpus of legislative history, Lee & Mouritsen confuse *legal language* with *language about law*. Language can be used to talk about law without itself being legal language. To take an obvious example of this difference, consider the following excerpt from a well-known folk poem, *The Goose and the Common*, which criticizes English enclosure laws:

> The law locks up the man or woman
> Who steals the goose from off the common
> But leaves the greater villain loose
> Who steals the common from off the goose.[18]

This poem is *about* law, to be sure—it refers to property rights, theft, and punishment—but it is not written *in* legal language. It is well-documented that legal language has different characteristics than nonlegal language, including differences in word choice, syntax, and other conventions.[19] And of all legal language, statutory language is the most distinct from nonlegal language. Perhaps most importantly, statutory language prescribes behavior; that is, it provides rules of decision for courts, agencies, and individuals to follow.[20] Because of this unique function, statutory language exhibits a number of distinguishing features. For instance, statutes contain words and phrases uncommon in nonlegal speech.[21] These include both archaic words and phrases[22] and formulaic constructions, like enacting clauses.[23] Statutory language also contains common words used in uncommon ways,[24] like phrases that derive from the common law and near synonyms recited in formulaic pairs. These coupled synonyms

---

18.   James Boyle, *The Second Enclosure Movement and the Construction of the Public Domain*, 66 L. & COMTEMP. PROBS. 33 (2003). For other versions of this poem, see *The Goose and the Common*, UNION SONGS, https://unionsong.com/u765.html [https://perma.cc/YU4A-K8FL] (last visited Sept. 13, 2021).

19.   Yon Maley, *The Language of Legislation*, 16 LANGUAGE IN SOC'Y 28, 40 (1987).

20.   Andrei Marmor, *The Pragmatics of Legal Language*, 21 RATIO JURIS 423, 425 (2008).

21.   DAVID MELLINKOFF, THE LANGUAGE OF THE LAW 11 (1962) (noting that legal language is a "specialized tongue that distinguishes lawyer and nonlawyer").

22*.   Id.*

23*.   E.g.*, 1 U.S.C. § 101 ("The enacting clause of all Acts of Congress shall be in the following form: 'Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled.'").

24.   MELLINKOFF, *supra* note 21, at 11–12 (noting that legal language is characterized by the "frequent use of common words with uncommon meanings").

include the well-known doublets "aid and abet,"[25] "null and void,"[26] "safe and sound,"[27] along with many others. In addition, because legislation is (nearly always)[28] written to classify future conduct or events, it is written in general, prospective, impersonal language.[29] And to minimize unforeseen results, legislation is often crafted with unusually explicit or redundant language meant to reduce unintended consequences.[30] Finally, reflecting the purpose for which they are consulted, statutes normally contain numerous visual cues, like paragraph numbering, headings, and unique indentation.[31] Some of the special characteristics of statutory language are set out in Table 1, below.[32]

Table 1. Conventions of Statutory Language

Conventions of Statutory Language
Prescriptive language
Words uncommon in nonlegal speech
Archaic words
Formulaic language
Unusual word choices and phrases
Coupled synonyms
General, impersonal, prospective language
Explicit or redundant language to cover unforeseen circumstances

---

25.   22 U.S.C. § 2712 (permitting Secretary of State to impose regulations related to services that would aid and abet international terrorism).

26.   30 U.S.C. § 28i (deeming a claim null and void for failure to pay certain fees).

27.   12 U.S.C. § 371c (requiring that certain banking transactions be made only if they are consistent with safe and sound practices).

28.   Relatively rarely, statutes are written to target an identifiable individual for treatment that is not generally applicable. *See*, *e.g.*, Evan C. Zoldan, *Legislative Design and the Controllable Costs of Special Legislation*, 50 MD. L. REV. 415, 422–23 (2019) (defining and describing special legislation); Evan C. Zoldan, *The Equal Protection Component of Legislative Generality*, 51 U. RICH. L. REV. 489, 497–500 (2017) (providing modern examples of special legislation).

29.   Maley, *supra* note 19, at 40 ("Rules use linguistic forms that realise the meanings asymmetrical, formal, and impersonal.").

30*.   Id.* at 41 (noting that legislative drafters tend to "employ communicative strategies for producing rules that have precise and explicit meanings. This will involve technicality on the one hand and specification and repetition on the other.").

31*.   Id.* at 39–40.

32.   A more complete accounting of the special characteristics of statutory language is found in Evan C. Zoldan, *The Conversation Canon* (forthcoming 2021).

Language found in nonlegal sources lacks these distinctive features of statutory language. This observation may well be obvious when considering a poem like *The Goose and the Common*. As a poem, its style, meter, and other characteristics follow the conventions of poetry rather than the conventions of statutory language.[33] But, the differences between legislative history and statutory language are no less stark. Despite the fact that floor statements, hearing testimony, and committee reports all include language *about* law or prospective law, legislative history is not written in statutory language. As a result, analyzing language found in a corpus of legislative history will not, *contra* Lee & Mouritsen, reveal how statutory language is used.

To illustrate the fact that legislative history, like other language found in nonlegal sources, lacks the distinctive features of statutory language, we can compare the language of a particular statute with the legislative history that was generated during its creation. By comparing these different types of texts, it becomes clear how dramatically the language in a corpus of legislative history would differ from statutory language. Indeed, we can make this comparison with statutory language that Lee & Mouritsen themselves cite as an example of language that would benefit from the insights of corpus linguistics. Lee & Mouritsen often cite *Taniguchi v. Kan Pacific*,[34] in which the Supreme Court construed the word "interpreter," found in the Court Interpreters Act.[35] A close look at the Act reveals that it exhibits the most common characteristics of statutory language highlighted above.

To begin with, it prescribes behavior, creating rights and vesting authority in an agency ("The Director of the Administrative Office of the United States Courts shall establish a program to facilitate the use of interpreters in courts of the United States.").[36] It includes words and phrases uncommon in nonlegal speech ("relator"),[37] archaic words ("writ"),[38] and formulaic constructions ("Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled").[39] The Act includes common words used in unusual

---

33. *See* JOHN HOLLANDER, RHYME'S REASON 3 (1981) (introducing conventions of poetry, including structures of lines of verse, patterns of rhyme, and schemes of syntax and word order).

34. Taniguchi v. Kan Pac. Saipan, Ltd., 566 U.S. 560 (2012).

35. P.L. 95-539, 92 Stat. 2040 (Oct. 24, 1978).

36. *Id.*

37. *Id.*

38. *Id.*

39. *Id.*

ways ("taxed as costs"),[40] common law terms ("reasonably")[41] and even the recitation of coupled synonyms ("rules and regulations").[42] Moreover, it is written in general, prospective, impersonal language ("The presiding judicial officer…shall utilize the services of the most available certified interpreter.").[43] And it is written with unusually explicit and redundant language, ostensibly to cover unforeseen situations ("the presiding judicial officer shall not establish, fix, or approve compensation and expenses").[44] Finally, it also includes the expected visual cues, like paragraph numbering, headings, and unique indentation. Some of the special statutory characteristics of the Court Interpreters Act are set out in Table 2, below.

---

40. *Id.* at 2041.
41. *Id.*
42. *Id.* at 2043.
43. *Id.*
44. *Id.* at 2041.

Table 2. Statutory Conventions Present in the Court Interpreters Act

Convention of Statutory Language	Example from the Court Interpreters Act
Prescriptive language	"The Director…shall establish a program to facilitate the use of interpreters"
Words uncommon in nonlegal speech	"relator"
Archaic words	"writ," "moneys," "thereof"
Formulaic language	"Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled . . ."
Unusual word choices and phrases	"taxed as costs," "clerk of court"
Coupled synonyms	"rules and regulations," "between or among"
General, impersonal, prospective language	"The presiding judicial officer…shall utilize the services of the most available certified interpreter . . ."
Explicit or redundant language to cover unforeseen circumstances	"establish, fix, or approve compensation"

   Comparing the features of the Court Interpreters Act itself with the legislative history generated during its creation demonstrates just how different statutory language is from legislative history. Consider two different sources of Court Interpreters Act legislative history: hearing testimony in support of the bill[45] and a floor statement made by one of the bill's sponsors.[46] Most obviously, unlike the Act itself, this

---

   45. *E.g.*, *Court Interpreters Act: Hearing on H.R. 10228, H.R. 10129, and S. 1315 Before the Subcomm. on Civ. and Const. Rts. of the H. Comm. on the Judiciary*, 95th Cong. 4–5 (1978) (statement of Rep. Fred Richmond).
   46. *E.g.*, 124 CONG. REC. S.36685 (Oct. 13, 1978) (statement of Mr. DeConcini).

legislative history lacks the visual cues, archaic language, coupled synonyms, prospective, prescriptive language, and other features of statutory language. Rather than vesting authority or providing a rule of decision for anyone to follow, the legislative history describes, in narrative form, the reasons for the introduction of the bill ("40 million individuals...presently do not enjoy the full benefits of our legal system.").[47] Rather than prescribing rules of conduct, it describes events; and rather than applying prospectively, it refers to past and present events ("Our legal system has not lived up to the basic American ideal of equal justice and fairness to all.").[48] And, rather being limited to impersonal, general language, it includes personal reflections ("I am very pleased by the imminent passage of S. 1315.").[49]

There are other differences, too, like differences in readability,[50] between legislative history and statutory language. Even without a full accounting of the differences between these two types of texts, however, it is apparent how Lee & Mouritsen's proposal to use a database of legislative history misses the mark. Lee & Mouritsen have already conceded that the differences between statutory language and nonlegal language "call into question the probity of evidence from a general corpus."[51] But, legislative history, too, differs from statutory language—and in the same ways as nonlegal language more generally. Just like other nonlegal language, legislative history lacks the distinctive word choice, syntax, grammar, and other conventions of statutory language. As a result, if and when a corpus of legislative history is created, it would stand on the same footing as a general corpus. And just like a general corpus, a corpus of legislative history will not allow Lee & Mouritsen to learn, contrary to their claim, how a statutory term is used in legal language.

II. CHOOSING A CORPUS OF LEGISLATIVE HISTORY

Lee & Mouritsen's concession that it is appropriate to search different corpora under different circumstances highlights another thorny question: how do we know when to search a particular corpus? The answer to this question is crucial. Because different corpora, by

---

47*. Id.*

48*. Court Interpreters Act: Hearing*, *supra* note 45.

49*. Id.*

50*. See* Robert Benson, *The End of Legalese: The Game is Over*, 13 N.Y.U. J. L. & Soc. Change 519, 547–50 (1984) (describing Flesch and Gunning Fog calculations of readability).

51. Lee & Mouritsen, *supra* note 2, at 301.

design, have different bodies of text, searching one corpus will yield different data than searching another.[52] As a result, Lee & Mouritsen first must demonstrate why searching a particular corpus is appropriate before their use of corpus linguistics tools can plausibly be said to provide predictable, replicable answers to interpretive questions. If they cannot demonstrate a persuasive way to choose one corpus over another, there is little hope of corpus linguistics tools returning "evidence-based answers"[53] to interpretive questions, as Lee & Mouritsen assert.

Lee & Mouritsen suggest two ways to choose between a corpus of legislative history and a general corpus. In some parts of their work, they argue that the decision to use a corpus of legislative history should be based on the interpreter's ideological commitments. By contrast, elsewhere, they argue that an interpreter's choice of corpus should be based on how a term is used. There are two main problems with these suggestions. First, they are in tension with one another. And second, even taken separately, these two suggestions do not appear to justify an interpreter's choice of one corpus over another.

A.   Choice of Corpus: A Theoretical or Empirical Question?

Lee & Mouritsen offer two suggestions about when to use a corpus of legislative history to search for the meaning of statutory text. However, these suggestions appear to be in tension with one another, if not irreconcilable. On one hand, Lee & Mouritsen appear to believe that the choice between a corpus of legislative history and a general corpus properly depends on the interpreter's ideological commitments. They offer the following dichotomy: "Judges who prioritize fair notice in statutory interpretation may want to consult a modern, general corpus—especially for criminal statutes—while those who view themselves as 'faithful agents of the legislature' may wish for a corpus of congressional speech."[54] On this view, the characteristics of a particular statutory term should not matter to the choice of corpus. Rather, the choice of corpus should be a function of (or shibboleth indicating) the interpreter's ideological commitments.

But, if this is Lee & Mouritsen's view, it is hard to square with their other statement about how to choose a corpus: that the choice should depend on how a particular statutory term is used in practice. Specifically, when considering the term "carry," the term interpreted

---

52.   Zoldan, *supra* note 3, at 420.

53.   Lee & Mouritsen, *supra* note 2, at 294.

54*.   Id.*

in *Muscarello v. United States*,[55] they note the possibility that "we use the terminology of 'carry' a firearm differently when we are prohibiting it in a criminal law…than we do when we are merely speaking descriptively."[56] And, if a word does have a different legal and nonlegal meaning, Lee & Mouritsen concede that it "could call into question the utility of a general corpus for assessing the meaning of the language of this statute."[57] Elsewhere, they reaffirm this view, arguing that the choice of corpus should depend on how frequently a word is used in a particular corpus.[58] On this view, the choice of corpus should *not* depend on an interpreter's ideological commitments. Rather, it properly depends on some characteristic of the term being interpreted.

I will not hazard to guess which of these two opinions properly states Lee & Mouritsen's position on corpus choice. But, I do note the tension between these two positions. For example, on the first view, an interpreter committed to fair notice would use a general corpus irrespective of whether the word she is interpreting is used differently in legal speech and nonlegal speech. By contrast, under the second view, an interpreter of any ideological preference would search in a specialized corpus for a word used in a specialized way. As they further refine their work, it is essential for Lee & Mouritsen to clarify their opinion about when to choose a general corpus, as opposed to some other kind of corpus. In the absence of a test for choosing among different corpora, it is impossible for Lee & Mouritsen's corpus searches to be described as objective; and it is impossible for them to consider their interpretations "evidence-based."[59]

B.   ADDITIONAL DIFFICULTIES FOR THE CHOICE OF CORPUS

Until and unless they set out their views in more detail, Lee & Mouritsen's suggestions about how to choose a corpus cannot be fully evaluated. Moreover, as they clarify their position, Lee & Mouritsen should keep in mind some difficulties they will encounter justifying either of their two suggestions. First, *contra* their suggested dichotomy, choosing a corpus of nonlegal language does not promote fair notice; likewise, choosing a corpus of legislative history does not promote faithful agency. And second, choosing a corpus based on whether a word is "ordinary" or "legal" in nature fails to reflect the fact that

---

55.   Muscarello v. United States, 524 U.S. 125 (1988).
56.   Lee & Mouritsen, *supra* note 2, at 301.
57*.   Id.*
58*.   Id.* at 303.
59*.   Id.* at 294.

these categories are fuzzy at best.

1.   Neither Fair Notice nor Faithful Agency

Lee & Mouritsen argue that interpreters who value fair notice should favor interpretations based on the meaning of words found in a general corpus. However, interpreting statutory terms according to their nonlegal meaning does not provide notice of statutory obligations. A statute—even a simple one—is more than just a string of "ordinary" words with the occasional "legal" term mixed in.[60] Rather, as our analysis of the Court Interpreters Act revealed, the entire structure of a statute is different than nonlegal speech, including its grammar, syntax, word choice, and other conventions.[61] It is for this reason that statutes, even simple ones, tend to be unintelligible to the general public.[62] Because of the array of differences between legal and nonlegal language, isolating one term from the rest of the statutory text, interpreting it as "ordinary" text, and reinserting it into the statute, does not make the statute as a whole intelligible to the public. As a result, searching a general corpus for a term embedded in a statute will do little to promote notice of a statute's obligations.

What's more, Lee & Mouritsen's methodology will often make the meaning of statutory language *less* predictable compared with the use of traditional interpretive methods. When interpreted through traditional methods, a term that creates a legal obligation often acquires a specialized meaning that is more precise than the nonlegal meaning of that same term.[63] As a corollary, resorting to a general corpus to interpret a term will return a broader range of possible senses than would be uncovered by engaging in traditional methods of interpretation. The greater the number of possible meanings of a term, the less likely it is that a member of the public will have notice of the meaning that a legal interpreter, like a judge, would apply in any given situation. Consider, for example, a simple legal command that most of us encounter frequently: the single word "stop." When used on a stop sign, this word has a precise meaning; and it is one that can easily be learned through traditional legal analysis, including by looking it up in

---

60.   Peter M. Tiersma, *Some Myths About Legal Language*, *in* Speaking of Language and Law 29 (Lawrence M. Solan et al. eds., Oxford Univ. Press, 2015).

61*.   See* Tables 1–2 and accompanying text.

62.   John Gibbons, Language and the Law 25 (John Gibbons ed., 1994).

63.   David A. Strauss, *Why Plain Meaning?*, 72 Notre Dame L. Rev. 1565, 1568 (1997) ("Terms like 'witness,' 'zoning,' and even 'speed limit,' when used in a legal context, can mean something quite different from what they might mean when used in other contexts.").

a driver's manual[64] or by reading case law.[65] By contrast, if an interpreter attempts to interpret this language according to the nonlegal meaning of the word "stop," including by searching for it in a general corpus, she will quickly find that the word has many different nonlegal senses, including: prevent, plug up, cease permanently, and others.

As the stop sign example shows, interpreting statutory language according to the meaning of words found in a general corpus can multiply the number of possible meanings compared with using traditional tools of interpretation. By increasing the number of possible meanings, a search in a general corpus makes it less likely that a member of the public will know which meaning a judge will choose. As a result, reliance on a general corpus for statutory meaning not only fails to provide fair notice, it is apt to provide *less* notice to a member of the public than would be uncovered by following traditional methods of interpretation.

As an alternative to fair notice, Lee & Mouritsen also argue that interpreters (presumably judges) who view themselves as faithful agents of the legislature will favor a corpus of legislative history. As noted above, the language of legislative history is not legal language, but rather nonlegal language about law. Interpreting statutory language as if it were nonlegal language is not faithful to the will of the legislature because it undermines the role that the legislature plays in policymaking. As I explain more fully in a forthcoming piece, *The Conversation Canon*,[66] the legislature sets policy, in large part, by using terms that have well-known meanings to the primary audiences of statutes, that is, judges, agency officials, and other individuals who interpret statutes in their professional or official capacities. For example, when Congress granted the FDA authority to regulate tobacco in the Tobacco Control Act,[67] it used terms, like "adulterated" and "misbranded,"[68] that judges and agency officials knew and understood because of their long experience with the Food, Drug, and Cosmetic Act.[69]

---

64. *E.g.*, Ohio BMV Manual 59, https://driving-tests.org/wp-content/uploads/2020/08/OH_en.pdf [https://perma.cc/7KPF-DNAT] (last visited Sept. 13, 2021) (noting that a stop sign requires drivers to "stop before proceeding").

65. *E.g.*, Grossman v. Andros, 735 N.E.2d 499, 502 (Ohio Ct. App. 1999) (interpreting statutory obligation to stop at a stop sign).

66. Zoldan, *supra* note 32.

67. Family Smoking Prevention and Tobacco Control and Federal Retirement Reform Act, Pub. L. No. 111-31, § 103(d), 123 Stat. 1776 (2009).

68. *Id.*

69. Kordel v. United States, 335 U.S. 345, 348 (1948) (interpreting "misbranded" under the FDCA); United States v. Universal Mgmt. Servs., Corp., 191 F.3d 750, 754 (6th Cir. 1999) (interpreting "adulterated" under the FDCA).

By using this familiar language (rather than, say, "stale" or spoiled"), Congress ensured that prior interpretations of the statutory language by agency officials and judges would constrain future interpretations of the same language in the new statute. To the contrary, when judges interpret statutory text according to the *nonlegal* meaning of its words, they give statutory language an interpretation that does not reflect the specialized knowledge of the statute's authors and audiences and that is not constrained by previous judicial interpretations and agency practices. A judge who gives statutory language meanings that do not reflect the specialized knowledge of the statute's authors and audiences is in no way a "faithful agent" of the legislature. Instead, this kind of judge all-but-guarantees that the resulting interpretation will fail to reflect the legislature's will, as expressed through the language that it chose.

2.   The "Ordinary" or "Legal" Nature of a Term

In significant tension with their arguments about fair notice and faithful agency, Lee & Mouritsen also argue that the choice of corpus should depend, not on the ideology of the interpreter, but rather on some characteristic of the term being interpreted.[70] For example, they suggest that some terms are legal in nature while others are ordinary or nonlegal in nature.[71] To take an extreme example, perhaps we could all agree that the statutory term "*per stirpes*" is easily identified as a legal term that ought to be interpreted as such.[72] But, to the extent that Lee & Mouritsen's choice of corpus requires determining that there is something inherently "legal" or "nonlegal" about particular terms, their methodology faces a significant challenge: most statutory terms cannot be labelled legal or nonlegal simply by looking at them. Rather, the determination that a word is used in a nonlegal rather than in a legal sense is a conclusion rather than an objective fact that can be discovered. As a result, a choice of corpus that depends on distinguishing legal from nonlegal terms will be rooted in exactly the type of subjective determination that Lee & Mouritsen claim to be avoiding.

The line between legal and nonlegal terms, if one exists, is indistinct at best. Not every word that "has the sound of the law" is a legal term.[73] Moreover, and more importantly for this Essay, countless words sound "ordinary" because they are used in nonlegal settings,

---

70.   Lee & Mouritsen, *supra* note 2, at 301.

71*.   Id.* at 303.

72*.   See* MICH. COMP. LAWS § 700.2718.

73.   MELLINKOFF, *supra* note 21, at 17.

but in fact have specialized legal meanings.[74] Consider just a few examples. Whatever the nonlegal meaning of the commonly used word "rule," it surely means something different in the context of Administrative Law.[75] Or, take the word "consideration;" its well-known meanings in contract law differ markedly from its nonlegal uses.[76] For a constitutional law example, we need search no further than the word "search." If a police officer helps me look for my lost keys in a public park, a user of nonlegal English could well conclude that the officer and I searched for my keys; but the officer almost surely did not perform a "search" within the meaning of the Fourth Amendment.[77] By contrast, if a police officer attaches a GPS device to my car to monitor its location, the officer likely did perform a "search" within the meaning of the Fourth Amendment, although one would not describe the conduct this way in nonlegal English.[78]

As these examples demonstrate, terms used in legal texts are often commonly used in nonlegal speech as well. But, simply because a word is commonly used in nonlegal speech does not tell us that it is used in a nonlegal sense in a particular statute. Because there is nothing either inherently legal or nonlegal about most statutory terms, an interpreter's conclusion that a statutory term is legal or nonlegal is just that—a conclusion rather than an objective assessment of an inherent characteristic of the word itself. As a result, a choice of corpus based on whether a word is legal or nonlegal will, similarly, be a value-laden conclusion. The fact that their choice of corpus rests on the interpreter's subjective determination about whether a term is "legal" or "nonlegal" in nature refutes Lee & Mouritsen's claim that interpretations based on the data they retrieve from a corpus search are objective.[79]

---

74. Strauss, *supra* note 63, at 1568 ("Terms like 'witness,' 'zoning,' and even 'speed limit,' when used in a legal context, can mean something quite different from what they might mean when used in other contexts.").

75. 5 U.S.C. § 551(4) ("'[R]ule' means the whole or a part of an agency statement of general or particular applicability and future effect designed to implement, interpret, or prescribe law or policy.").

76. 3 WILLISTON ON CONTRACTS § 7:2 (4th ed.) (describing modern meanings of "consideration" in the context of contract law).

77. Oliver v. United States, 466 U.S. 170, 178 (1984) ("[A]n individual may not legitimately demand privacy for activities conducted out of doors in fields.").

78. United States v. Jones, 565 U.S. 400, 404–05 (2012) (holding that a Fourth Amendment "search" occurs when the "Government physically occupied private property for the purpose of obtaining information").

79. Lee & Mouritsen, *supra* note 2, at 294.

CONCLUSION

There is much to admire in Lee & Mouritsen's latest contribution to the corpus linguistics literature. Their responses to their critics demonstrate an admirable willingness to engage with tough questions about their enterprise and even to make some concessions when faced with persuasive arguments. However, their responses, while welcome, do not answer the significant objections that have been raised to their use of corpus linguistics tools to interpret legal texts. In this Essay, I raise two related objections: Lee & Mouritsen's proposed corpus of legislative history fails to capture the way that statutory words are used in a legal context; and they have failed to articulate how interpreters should choose between a legal corpus and a general corpus.

But, lurking behind these seemingly fine points stands a more fundamental question, and it is one that goes to the heart of Lee & Mouritsen's project. Their stated reason for using corpus linguistics tools to interpret legal texts is to elicit, in their words, "objective evidence" of meaning. Nevertheless, their ambivalence about the role of different corpora—what these corpora should contain and when they should be searched—undermines this goal. Because different corpora contain different texts, corpus analysis can be no more objective than the choice of corpus. And without a better-supported view of the role of different corpora in the interpretive process, a corpus search cannot lead to interpretations that can fairly be described as objective.