THE ORDINARY LAWYER CORPUS: THE FEDERALIST PAPERS APPROACH

SAMANTHA THORNE*

INTRODUCTION

Corpus linguistics ("CL") is gaining steam in courts and in journals. Though neither a *Harry Potter* incantation[1] nor a *CSI* forensic investigative procedure,[2] CL helps courts place ordinary language usage in context. By examining large corpora of written and transcribed words, judges can more objectively determine how people understand legal texts and, in turn, how they as judges ought to interpret ambiguous legal texts.

Yet we are not to the point of Supreme Court justices declaring "we are all corpus linguists now."[3] CL is well acquainted with criticism. But rather than disparaging the tool itself, many critics really

---

* J.D. Candidate, Harvard Law School Class of 2023. Adam Thorne (Deeplearning Scientist at PhysicsAI; B.S. in Biomechanical Engineering and M.S. in Mechatronics, Robotics, and Machine Learning from Stanford University) built the "ordinary lawyer" corpus.

1. Josh Jones, *The "Weaponization" of Corpus Linguistics: Testing Heller's Linguistic Claims*, 34 BYU J. PUB. L. 135, 144 (2019) ("'Corpus linguistics' might sound strange and intimidating—not unlike a *Harry Potter* incantation . . . .").

2. John S. Ehrett, *Against Corpus Linguistics*, 108 GEO. L.J. ONLINE 50, 51 (2019) ("'Corpus linguistics' may sound like a forensic investigative procedure on *CSI* or *NCIS* . . . .").

3. *But cf.* Elena Kagan, *The Scalia Lecture: A Dialogue with Justice Kagan on the Reading of Statutes*, HARV. L. TODAY (Nov. 17, 2015), http://today.law.harvard.edu/in-scalia-lecture-kagan-discusses-statutory-interpretation [http://perma.cc/3BCF-FEFR] ("[W]e're all textualists now.").

object to how a particular CL analysis is structured based on the critic and corpus linguist having competing interpretive theories.[4] Interpretive approaches undergird the CL analysis and come to light when answering whether CL should value a general speech community over a more specialized one, public meaning over speaker intent, and (if prioritizing public meaning) whether it is public meaning today or at the time of enactment.[5] These issues lie at the heart of selecting the relevant speech community and thus the proper corpus for a CL analysis.

Many CL analyses in the literature have prioritized contemporary public meaning based on a general speech community.[6] This Note does the same. But rather than choosing the broader public or the "ordinary person" as the relevant speech community, this Note focuses on lawyers (the "ordinary lawyer"). Despite lawyers' large role in connecting ordinary people to courts, their voices are silent in CL analyses. No corpus consisting of words written entirely by lawyers exists.

While the lawyer voice has been silent in CL, it has not been silent in other methods of discerning ordinary meaning. First, publications that employ survey methods to find ordinary meaning import a legal perspective. Professor Kevin Tobia surveys law students and U.S. judges in *Testing Ordinary Meaning*,[7] while authors of *Statutory Interpretation from the Outside* survey 1L law students.[8]

---

4. *See* Thomas R. Lee & Stephen C. Mouritsen, *The Corpus and the Critics*, 88 U. CHI. L. REV 275, 300 (2021) ("One critique of the use of corpus tools concerns the selection of the relevant language community. . . . [W]e think that the arguments about language community ultimately reinforce rather than undermine the need for and utility of corpus linguistic analysis . . . .").

5. *See id.* at 294 ("The answers to these questions will determine which corpus a judge should use in a particular case.").

6. *See, e.g.*, Thomas R. Lee & Stephen C. Mouritsen, *Judging Ordinary Meaning*, 127 YALE L.J., 788, 847 (2018) [hereinafter Lee & Mouritsen, *Judging Ordinary Meaning*].

7. Kevin P. Tobia, *Testing Ordinary Meaning*, 134 HARV. L. REV. 726, 734 (2020).

8. Kevin Tobia et al., *Statutory Interpretation from the Outside*, 122 COLUM. L. REV. 213, 245 (2022).

Second, textualism has considered a lawyer perspective. Then-Professor Amy Coney Barrett explained in *Congressional Insiders and Outsiders* that Justice "Scalia was not always clear about whether the prototypical reader is an ordinary member of the public or a lawyer."[9] At times, "he treated lawyers as the relevant linguistic community" since "one can hardly claim that the ordinary guest at a cocktail party would be aware of the ancient principles of common law . . . ."[10] Justice Barrett highlighted that "[m]ore should be said about whether and when a court should interpret statutes through the eyes of an ordinary lawyer rather than an ordinary person."[11] The fact that these methods of finding ordinary meaning value the lawyer perspective raises the question of why CL has not followed their lead.

Even *prioritizing* the lawyer voice in CL analyses may be worthwhile. Professors John McGinnis and Michael Rappaport address this.[12] They argue that the relevant interpretive *rules* that existed at a document's time of enactment bind how judges interpret the document in the future.[13] In the case of the Constitution, this rule was a "legal interpretive rule": looking to the *lawyer's* understanding of legal documents. This rule seems to extend to statutes.[14] Provided this phenomenon, why does the CL literature not prioritize ordinary meaning based on a general ordinary lawyer corpus?

---

9. Amy Coney Barrett, *Congressional Insiders and Outsiders*, 84 U. CHI. L. REV. 2193, 2202 (2017).

10. *Id.*

11. *Id.* at 2210.

12. John O. McGinnis & Michael B. Rappaport, *Original Methods Originalism: A New Theory of Interpretation and the Case Against Construction*, 103 NW. U. L. REV. 751 (2009).

13. *Id.* at 764 ("[I]nterpretive rules that were regarded as applying to the Constitution are binding.").

14. *See id.* at 757 (articulating legal interpretive rules govern how to interpret legal documents generally).

Two plausible explanations exist. First, no one has made a strong enough normative argument for choosing lawyers as a relevant speech community yet. Second, a lawyer-focused corpus does not appear to exist. This Note addresses both blind spots.

Part II details the legal theory underpinning the "ordinary lawyer" CL analysis. I argue that statutory register, modern practice, and history support a law of interpretation of prioritizing a lawyer perspective in finding ordinary meaning. Part III highlights the goals and methods of constructing an "ordinary lawyer" corpus. Part IV summarizes the three cases undergoing an ordinary lawyer CL analysis (*McBoyle v. United States*, *United States v. Muscarello*, and *Taniguchi v. Kan Pacific Saipan, Ltd.*). This section discusses the cases' interpretive questions and the outcomes of Justice Lee and Professor Mouritsen's 2018 ordinary person CL analyses. Part V details the results of the three cases' ordinary lawyer CL analysis and compares them to Justice Lee and Professor Mouritsen's ordinary person corpus results of the same cases. Finally, Part VI concludes the Note, considering some implications of the CL ordinary lawyer analysis and potential next steps.

I. LEGAL THEORY: WHY A LAWYER SPEECH COMMUNITY?

A. Law of Interpretation

Deferring to lawyers' understandings of legal documents is a background law of interpretation. When ambiguity stumps courts, Professors William Baude and Stephen Sachs argue that a general "law of interpretation" should fill "gaps that would otherwise be filled by the interpreter's normative priors."[15] This is the idea that "a system of established rules of construction might make the process of statutory interpretation more predictable, effective, and

15. William Baude & Stephen E. Sachs, *Law of Interpretation*, 130 HARV. L. REV. 1079, 1097 (2017).

even legitimate."[16] Interpretive rules bind legal arguments since they "allo[w] us to agree on what our rules are precisely so that we can debate whether to change them."[17]

A law of interpretation requires normative support, which Professors McGinnis and Rappaport provide. They advocate that the determinative law of interpretation in constitutional contexts is an "original methods" approach, defined as looking to how individuals at the time of enactment expected courts and readers to resolve interpretive questions that arose from the Constitution.[18] The original method for interpreting the Constitution is to adopt a "legal interpretive rule": deferring to the lawyer community's linguistic practice and prioritizing its understanding of legal documents.[19] Professors McGinnis and Rappaport state that statutes and other legal documents at the time of the Constitution's enactment share the same legal interpretive rules as the Constitution.[20] They also state that the legal interpretive rule applies whether one prioritizes original intent or original meaning.[21] For original intent, enactors

16. Thomas R. Lee et al., *A Linguistic Approach to Linguistic Canons*, at 269 (forthcoming).

17. Baude & Sachs, *supra* note 15, at 1097.

18. *See* McGinnis & Rappaport, *supra* note 12, at 754–55 ("[T]he meaning of language requires reference to the interpretive rules and methods that were deemed applicable to that language at the time it was written.").

19. *Id.* at 765.

20. *See id.* at 769 ("These rules were applied generally to legal documents."); *see also id.* at 770 ("[T]he enactors assumed the interpretive rules that were applied to statutes would also be a model for constitutional interpretive rules."). It is worth noting what the authors say at *id.* at 791 n.140 ("[W]e do not have space to determine the amount of evidence needed to establish the interpretive rules that are binding today.") This Note assumes CL can apply original methods to more modern statutes because features of modern statutes do not differ significantly from early statutes and because Professors Rapport and McGinnis qualify their statement at *id.* ("If the interpretive rules derive from a constitutional provision, the evidence required is not likely to differ from that needed to establish any constitutional norm as a matter of original meaning.").

21. *See id.* at 765.

expected their words to be understood based on background interpretive rules: how one with formally endowed legal knowledge would approach finding meaning.[22] For original meaning, the broader public understood that a lawyer's interpretation of the document would trump the layperson's.[23] This does not mean that deferring to lawyers' understandings of legal texts requires words to be read as legal terms of art: lawyers often interpret language ordinarily.[24] Interestingly, Professors McGinnis and Rappaport argue that using original methods tethers statutory analyses to the time of enactment and originalist principles.[25] This tether could mean that CL analyses are not beholden to corpora that only capture Framer or congressional intent at the time of enactment, or the public's understanding of words at that time.

If original methods trump more minute originalist inquiries into public understanding or speaker intent at the time of enactment, prioritizing legal interpretive rules has strong implications for CL. First, focusing on the lawyer perspective overcomes temporal concerns with CL corpora. Choosing a corpus to find ordinary meaning relies just as much on answering "meaning as of when?" as answering "whose meaning?"[26] One may advocate looking only at word usage surrounding a statute's time of enactment and avoiding modern corpora like the NOW Corpus (assuming that the relevant enactment time is before NOW's start date).[27] Because a lawyer corpus is an application of the original method of legal interpretive rules, originalists have more flexibility in choosing an adequate CL corpus when using a lawyer-focused corpus. Second, a lawyer corpus and its original method application help bridge an enduring

22. *See id.* at 760.

23. *See id.*

24. *See id.* at 765 n.49.

25. *Id.* at 772 ("[T]he meaning of the Constitution [was] derived from original methods [and] will be continuous with the ordinary meaning of the document.").

26. Lee & Mouritsen, *Judging Ordinary Meaning*, *supra* note 6, at 818.

27. The NOW Corpus is a database of more than 15 billion words from web-based newspapers and magazines from 2010 to the present.

division perennial to statutory interpretation: speaker intent versus reader understanding. This Note prioritizes ordinary meaning over intent because it aligns well with "our understanding of what the rule of law entails," and collective intent and original expectations are nebulous concepts.[28] Ordinary meaning also better assuages concerns related to public notice, reliance interests, consistent application, and respect to legislative will while yielding "greater predictability than any other single methodology."[29] But it is important to note that this law of interpretation stands regardless of prioritizing intent or meaning. Applying Professors McGinnis and Rapport's original methods theory to statutes since the Founding Era, statutory speakers intend, and statutory readers understand, that people read statutes in accordance with background legal interpretative rules.

B. Statutory Register

Register analysis supports this law of interpretation of prioritizing the lawyer perspective. Justice Thomas Lee and Professor Stephen Mouritsen assert that a search for ordinary meaning must "address the differences in various linguistic registers."[30] They connect ordinary meaning to linguistic register because choosing an appropriate speech community depends on the register at issue. To analyze register, one looks at the functional relationship between language use and its situational context. Looking at this relationship helps capture extralinguistic variables essential to understanding language. In a forthcoming article, Justice Lee and his coauthors "define the key extralinguistic variables that define the statutory register" by "identify[ing] the mode of communication and the nature of the participants."[31] They land on three features: Statutory

28. WILLIAM N. ESKRIDGE, JR., INTERPRETING LAW: A PRIMER ON HOW TO READ STATUTES AND THE CONSTITUTION 36 (2016).

29. *Id.*

30. Lee & Mouritsen, *Judging Ordinary Meaning*, *supra* note 6, at 827.

31. Lee et al., *supra* note 16, at 278.

language "(a) is highly consequential; (b) may be produced only by a limited few who are constitutionally invested with legislative power; and (c) is intentionally difficult to create and to revise, such that cancellation of an anticipated implicature must be made in advance."[32] I add two more features, both focusing on statutory participants. First, lawyers have an overrepresented role in the legislative process as statutory producers. As Professors Adam Bonica and Maya Sen highlight in their book:

> American history points us to a pivotal and often-overlooked player: lawyers . . . [S]ince the nation's founding, lawyers have occupied politically outsized roles. For example, more than half of the men who signed the Declaration of Independence in 1796 were lawyers or trained in law (twenty-nine out of fifty-six) while twelve out of the first sixteen presidents were lawyers. In contemporary times, lawyers—a group that today comprises just 0.4 percent of the voting-age population—are extraordinarily overrepresented in Congress and the Executive Branch, with nearly 42 percent of congressional representatives as of 2018 coming from the legal profession.[33]

Second, legislators write statutes knowing that lawyers will be the key readers: "they interpret the law on behalf of clients" from the general public and "are themselves agents of the people they represent."[34] In other words, lawyers have an overrepresented role in statutory consumption. While Justice Lee and others argue "there is very little opportunity for interaction between the producers and the consumers of statutory language," the interaction that exists seems to come from lawyers being present in both production and consumption roles.[35] These two situational features of statutes—that they are written and read by an overrepresented community of lawyers—point to choosing lawyers as the relevant speech community.

---

32. *Id.* at 280.

33. ADAM BONICA & MAYA SEN, THE JUDICIAL TUG OF WAR 5 (2020).

34. Barrett, *supra* note 9, at 2209.

35. Lee et al., *supra* note 16, at 279.

C. Practice

Modern lawyer-client practices indicate that CL analyses should focus on the lawyer community. Lawyers advocate for their clients and help them access legal interpretation. In the CL literature, authors prioritize finding ordinary meaning based on contemporary language use by the broader public predominantly for fair notice purposes. And many may worry that prioritizing an "ordinary lawyer" corpus over an "ordinary person" one evades that fair notice goal. But "in reading a statute as a lawyer would, a court is not betraying the ordinary people to whom it owes fidelity, but rather employing the perspective of the intermediaries on whom ordinary people rely."[36] The constructive notice fiction "does not depend on the proposition that the language of the law is accessible to all people" but rather "assumes that the people are capable of deciphering language"—usually by relying on lawyers as a language deciphering tool.

Lawyer-generated revenues support relying on the lawyer community as the relevant speech community.

> [M]illions of ordinary people each year . . . consult with lawyers about the meaning of legal documents[.] [This] provides strong evidence that legal interpretive rules and concepts are employed in legal documents. The billions of dollars spent through these visits are not wasted, because it is important that statutory or contractual terms be more precise and less ambiguous than terms of ordinary language.[37]

And according to the 2021 Am Law 100 Report, the largest law firms in the U.S. earned $111 billion in total revenue in 2020 and the

36. Barrett, *supra* note 9, at 2209-10.

37. McGinnis & Rappaport, *supra* note 12, at 765 n.51.

average revenue per lawyer was $1.05 million.[38] This information all suggests that lawyers bring value when it comes to interpreting legal documents. Whether costs for legal services align with the value that they provide is another question, but these inflows support the background legal interpretive rule argument.

D. History

U.S. history defends prioritizing lawyers as the relative speech community. Consider the representative nature of American governance and the constitutional ratification process. "[T]he people did not vote directly on the Constitution, just as they did not vote directly on the passage of statutes."[39] Instead, they "relied on their representatives—who were more likely to be schooled in legal understanding or able to consult more learned colleagues."[40] Representation existed at the time of ratification and exists today.[41] Looking to the lawyer perspective neither takes the "We the People" out of the Constitution[42] nor removes the people's voice from the legislative process. U.S. government (through elected officials) and judicial practice (through lawyers who "draft documents that speak in the client's name") rely on representation.[43]

Next, consider the *Federalist Papers*. "As with other legal documents, the people decided whether to ratify the Constitution based on an explanation of its meaning by those with legal knowledge."[44] Alexander Hamilton, James Madison, and John Jay (collectively "Publius") wrote articles and essays to connect the larger public to

38. *2021 Am Law Report*, AMERICAN LAWYER, law.com/americanlawyer/2021/04/20/the-2021-am-law-100-report/ [https://perma.cc/NN3L-NGQM] (last visited June 24, 2022).

39. *Id.* at 771.

40. *Id.*

41. *See* BONICA & SEN, *supra* note 33, at 5.

42. McGinnis & Rappaport, *supra* note 12, at 770.

43. *Id.* at 771.

44. *Id*.

a vital legal document. Hamilton and Jay were both lawyers. Although Madison did not practice law, he studied it. More recently recovered manuscripts of his notes (lost for over a century) reveal his "significant grasp of law and his striking curiosity about the problem of language."[45] According to Professor Mary Sarah Bilder, Madison approached "the problem of legal interpretation as a *student* of law, never from the secure status of *lawyer*."[46]

Like the States and the public, the Court has also relied on these lawyers' understandings of the Constitution. Countless cases from *Marbury v. Madison*[47] to *Printz v. U.S.*[48] have relied on the *Federalist Papers* to interpret the Constitution. Some of the most controversial Court opinions like *Bush v. Gore*[49] also cite the *Federalist Papers* to answer interpretive questions. Considering Court usage of *The Federalist* is important. The *Federalist Papers* seem to apply Professors McGinnis and Rappaport's original methods law of interpretation directly.[50] This phenomenon strongly evinces that courts consider lawyer interpretations of legal texts when resolving legal questions.

E. Addressing Concerns

The above arguments evoke pushback. Does prioritizing ordinary lawyer meaning over ordinary public meaning ossify lawyers' disproportionate political and judicial power and create a quasi-aristocracy? This question is beyond the scope of this Note and depends more on policies to make law school and legal services more accessible. I only argue that courts and society already seem to focus on lawyer interpretations in their interpretive processes. An "ordinary lawyer" corpus aligns theory with on-the-ground practices and makes lawyers' extensive role in interpretation explicit.

---

45. Mary Sarah Bilder, *James Madison, Law Student and Demi-Lawyer*, 28 LAW & HIST. REV. 389, 390 (2010).

46. *Id.*

47. 5 U.S. (1 Cranch) 137, 170 (1803).

48. 521 U.S. 898 (1997).

49. 531 U.S. 98 (2000) (per curiam).

50. McGinnis & Rappaport, *supra* note 12, at 770

Bringing this practice to light is important not only for resolving ambiguities in legal texts but also for down-the-road policy decisions.

I also do not argue that interpreters and CL users should always prioritize the lawyer voice, particularly a modern lawyer voice. I echo Justice Barrett's call for more to "be said about whether and when a court should interpret statutes through the eyes of an ordinary lawyer rather than an ordinary person."[51] But I also answer Justice Barrett's call, in part. I hope this Note makes a case for why one should consider prioritizing lawyer perspectives and at least match CL practices with other ordinary meaning methods, which consider the lawyer voice at times. As I will discuss after conducting my own "ordinary lawyer" CL analyses, utilizing such a corpus can yield interpretive benefits.

Another concern from the above arguments relates to the *Federalist Papers* analogy. Many have hailed the essays as an important source of evidence of the Constitution's original meaning. But maybe the *Federalist Papers* do not embody the law of interpretation that Professors McGinnis and Rappaport advance. Instead, the essays may be more of an analog to legislative history since two of the three authors (Hamilton and Madison) were delegates at the 1787 Constitutional Convention and helped write the Constitution. One could argue that the newly minted American public and later courts did not refer to the *Federalist Papers* because they were following the original methods interpretation approach. Instead, they consider the *Federalist Papers* because men who were a part of the drafting and ratification process wrote them.

I disagree. First, the public likely did not rely on the *Federalist Papers* due to most of the writers' being Constitutional Convention delegates—Hamilton, Madison, and Jay wrote anonymously under a single pseudonym. And the public did not know the *Federalist*'s

51. Amy Coney Barrett, *supra* note 9, at 2210.

authorship until after Hamilton's death.[52] Second, while Supreme Court opinions have highlighted the writers' constitutional contribution status when relying on the *Federalist Papers*, I argue that relying on the *Federalist Papers* despite the authors' status still falls under the original methods law of interpretation. The key tenet of the legal interpretive rules is that lawyers have an educational, training, or knowledge advantage in interpreting certain types of documents. Because the Constitution was not well-circulated at this point and was a unique document, most people at the time of ratification—even lawyers—could not meet the requisite level of added interpretive value to a document with which they were unfamiliar. Third, the *Federalist Papers* are distinct from legislative history. Publius wrote them to persuade a larger public and they are not a transcript of Constitutional Convention proceedings.

II. CREATING AN "ORDINARY LAWYER CORPUS"

At the outset, it is important to note that creating a full-fledged "ordinary lawyer" corpus requires more attention than was allotted here, despite the time and labor already dedicated. The "ordinary lawyer" corpus is an initial corpus. But searching for ordinary meaning consistently and methodologically is like a relay: what matters most is that you advance the literature and allow people to grab the baton to build off prior progress. The work of determining ordinary meaning is even more like a "relay marathon"[53] (think

52. *About the Authors—Federalist Essays in Historic Newspapers*, LIBRARY OF CONGRESS, https://guides.loc.gov/federalist-essays-in-historic-newspapers/authors [https://perma.cc/3HZY-H3DC] (last visited June 24, 2022).

53. *See* James Salzman, Professor of Law, Harvard Law School and UCLA School of Law, COP26—What happened, What didn't happen, and What happens next? At Harvard Law School (Nov. 18, 2021) Harvard Law School lecture. Professor Salzman used this analogy to describe COP26 and long-term international efforts in the environmental policy and legal arena.

something akin to a *Ragnar* relay)[54] since courts and scholars have sought to resolve ambiguity in legal texts by discerning ordinary meaning since our nation's founding. It is an endurance event holistically even if the individual contributions are less so. Although this corpus did not run the full marathon, I believe it ran a strong relay leg.

A. Goals

The goal in constructing an "ordinary lawyer" corpus was to follow the *Federalist Papers* approach. This meant finding lawyer-written articles, essays, and blogs accessible to and intended for a larger public that provide lawyerly insights into legal texts and legal issues.[55] JD Supra and Lawyers.com meet these criteria. JD Supra offers articles written by large law firms that cover issues ranging from labor and employment to constitutional law.[56] Reading these articles does not require a subscription. The same is true for Lawyers.com.[57] We only used articles from Lawyers.com to fill in an area of law absent from JD Supra: criminal law. Because criminal proceedings mainly involve state actors, Lawyers.com is an important source for finding articles that covered criminal topics. The website includes blog posts written by lawyers from smaller, more local, and state-focused firms. The Lawyers.com blogs are more informal than the JD Supra articles. Individuals from the public can even submit criminal law questions for lawyers to answer.

Another goal was to make a lawyer equivalent to NOW's broader public-focused corpus. But because this is a from-scratch

54. The Ragnar Relay Series is a series of long-distance running relay races organized and orchestrated by Ragnar Events, LLC, which is based in Salt Lake City, UT. Teams of twelve run about 200 miles in a relay fashion from start to finish.

55. The broader public includes fellow lawyers, akin to the *Federalist Papers* largely being circulated to state representatives.

56. JD SUPRA, https://www.jdsupra.com/browse/legal-news.aspx [https://perma.cc/3D7U-3SVC] (last visited Feb. 14, 2022).

57. *Criminal Law*, LAWYERS.COM, https://www.lawyers.com/legal-info/criminal/ [https://perma.cc/96BL-8EFF] (last visited June 9, 2022).

corpus, it is raw and not tagged or monitored like the NOW Corpus. It is also much smaller than the NOW Corpus since it has only two contributory streams: JD Supra and Lawyers.com. But the "ordinary lawyer" corpus is a general corpus that looks to language use that is less formal than lawyer briefs to courts. The corpus extends beyond the adversarial process, which would likely advocate for two senses of the word. The "ordinary lawyer" corpus also provides more novel machine-learning techniques compared to more traditional corpora.

B. Methodology

I recruited the help of a deep learning scientist to construct the corpus using machine learning. He used Python web scraping libraries to extract word usage from various articles taken from two websites: www.jdsupra.com/browse/legal-news.aspx and blogs.lawyers.com/attorney/criminal-law/. He extracted all JD Supra articles from all the listed topics: Labor and Employment, Finance and Banking, General Business, Civil Procedure, Science, Computers and Technology, International Trade, Health, Securities, Business Organization, Elections and Politics, Intellectual Property, Administrative Agency, Privacy, Tax, Consumer Protection, Communications and Media, Environmental, Civil Rights, Energy and Utilities, Insurance, Residential Real Estate, Civil Remedies, Antitrust and Trade Regulation, Constitutional Law, and Government Contracting. These articles ended up being from 2020 to 2021. He scraped all Lawyers.com criminal law articles from 2018 to 2021. In total, he extracted 1028 articles and created a corpus text file of all articles appended together. Most of the articles came from JD Supra despite the longer time horizon accounted for in Lawyers.com. The dates do not align between the two websites because only articles dating to 2020 could be extracted from JD Supra. There would be too insignificant a sample of criminal law articles if we extracted only articles from 2020 to 2021 from Lawyers.com.

III. THE CASES & PREVIOUS "ORDINARY PERSON" CORPUS LINGUISTICS OUTCOMES

We performed a CL analysis for three cases using the "ordinary lawyer" corpus: *McBoyle v. U.S.*,[58] *U.S. v. Muscarello*,[59] and *Taniguchi v. Kan Pacific Saipan, Ltd*.[60] Justice Lee and Professor Mouritsen use these three cases in their Article, *Judging Ordinary Meaning*, to conduct a CL analysis through the NOW Corpus—the predominant modern general "ordinary person" corpus. The interpretive questions that I analyze are the same in substance and scope that Justice Lee and Professor Mouritsen analyze. Keeping these factors the same helps us compare the ordinary lawyer corpus to the ordinary person corpus.

The paragraphs below outline the three cases' legal questions and Justice Lee and Professor Mouritsen's accompanying NOW CL results. Collocation and concordance line analyses combine to create a CL analysis. Collocation shows us "the words that are statistically most likely to appear in the same context" as a searched term for a given period.[61] It gives a "*snapshot* of the semantic environment" that the subject word appears in which concordance lines later confirm.[62] Concordance lines "provide the crucial context in which different instantiations of the searched term have occurred."[63] Concordance lines require corpus linguists to pull randomized sample uses of the searched term and display it in its semantic environment. This CL prong relies on more context than collocation.

58. 283 U.S. 25 (1931).
59. 524 U.S. 125 (1998).
60. 566 U.S. 560 (2012).
61. Lee & Mouritsen, *Judging Ordinary Meaning*, *supra* note 6, at 837.
62. *Id.* (emphasis added).
63. Jones, *supra* note 1, at 147.

A. McBoyle v. U.S.

In *McBoyle v. U.S.*, the Court answered whether an "airplane" was a "vehicle" under the National Motor Vehicle Theft Act of 1919.[64] The justices concluded that "[w]hen a rule of conduct is laid down in words that evoke in the common mind only the picture of vehicles moving on land, the statute should not be extended to aircraft."[65] In other words, an airplane is not a vehicle under the statute.

Based on Justice Lee and Professor Mouritsen's collocation analysis, the NOW Corpus strongly indicates that "automobile" is the most common use of "vehicle."[66] The concordance lines confirm this. Indeed, 91 percent of "vehicle" uses were automobile-related, and none related to "airplanes." "To the extent that our notion of ordinary meaning has a frequency component, this data suggests that *automobile* is overwhelmingly the most common use of the word *vehicle*" in the NOW Corpus.[67]

B. U.S. v. Muscarello

Muscarello asked whether "carrying a firearm" encompassed the conveyance of a gun in a glove compartment when interpreting a statute that called "for a five-year mandatory prison term for a person who 'uses or carries a firearm' 'during and in relation to' a 'drug trafficking crime.'"[68] Applying a more sense-ranking approach, the Court held that "carries a firearm" broadly includes one who knowingly possesses and conveys firearms in a vehicle, including in the locked glove compartment which the person accompanies. Worrying about the statute's purpose to combat the dangerous combination of drugs and guns, Justice Breyer's majority

64. McBoyle, 283 U.S. at 27.
65. *Id.*
66. Lee & Mouritsen, *Judging Ordinary Meaning*, *supra* note 6, at 837.
67. *Id.* at 842.
68. *Id.* at 803.

opinion focused on the term "carry" (not "carries a firearm"), arguing its usage is not limited to "on-the-person application." Justice Ginsburg's dissent disagreed with Justice Breyer's isolation of "carries" and noted that the issue presented "is not 'carries' at large but 'carries a firearm.'"[69]

Justice Lee and Professor Mouritsen follow the majority and isolate "carries" in their CL analysis.[70] But they do factor in the larger "carries a firearm" syntactic structure and semantic relationships: controlling for the "syntax of a transitive verb, with the semantic features of a human subject and a weapon object."[71] The carry collocation analysis is less straightforward than the vehicle collocation. Various "carries" uses did not fit within the object and subject features that Justice Lee and Professor Mouritsen prioritize. Instead, there are instances of the inverse structure (inanimate objects carrying the human subject), metaphorical uses (felonies carrying certain penalties), and references to carrying out plans and executing. Justice Lee and Professor Mouritsen argue that the search still "reveals common collocates of *carry* that have similar semantic features to *firearm*" that "help us better evaluate the contexts in which *carry a firearm* occurs."[72]

After controlling for the desired structure and relationship of "carries a firearm," the concordance analysis yielded 104 instances indicating a sense of "carry a firearm" on one's person and only five suggesting "carry" in the car sense. "To the extent that we view the question of ordinary meaning as involving statistical frequency," Justice Lee and Professor Mouritsen's analysis "tells us that carry on one's person is overwhelmingly the most common use, while carry in a car is a possible but far less common use."[73]

---

69. Muscarello, 524 U.S. at 143 (Ginsburg, J., dissenting).
70. Lee & Mouritsen, *Judging Ordinary Meaning*, *supra* note 6, at 799.
71. *Id.* at 823.
72. *Id.* at 846.
73. *Id.* at 847–48.

C. Taniguchi v. Kan Pacific Saipan, Ltd.

Taniguchi involved an interpretive question about the ordinary meaning of the term "interpreter." Namely, is a litigation expert who translates written documents from one language to another an interpreter under a statute authorizing an award of costs for prevailing parties that use such services? The majority held that the statute did not encompass written translation since the more common sense of interpreter is one "engaged in simultaneous oral translation."[74] The majority and dissent mainly disagreed over whether the statute covers only the most common sense or permissible senses too.[75]

Justice Lee and Professor Mouritsen's "interpreter" collocates largely support the majority's position that interpreter most commonly refers to an "interpreter of spoken language" given collocates of "speaking, spoke, and listen."[76] The concordance analysis was slightly less straightforward since multiple senses of "interpreter" referenced interpretation of works of art, documents written in a primary language, and interpretation from a primary language to a second language.[77] Various concordance lines included transcripts of spoken interviews facilitated by an interpreter too. Importantly, however, there was not "a single instance of anyone referred to as an *interpreter* performing a text-to-text translation."[78] This finding (or more aptly lack of finding "interpreter" as a text-to-text translator) supports the majority's holding.

---

74. Lee & Mouritsen, *Judging Ordinary Meaning*, supra note 6, at 799.
75. *Id.*
76. *Id.* at 848 (emphasis omitted).
77. *Id.* at 849.
78. *Id.* at 850.

IV. APPLYING THE "ORDINARY LAWYER" CORPUS: CL RESULTS

We automated the collocation analysis by creating Python Word Clouds. For every article read, the machine learning identified instances of "vehicle," "vehicles," "carry," "carries," "carried," "carrying," "interpreter," and "interpreters." We stored the four words on either side of the searched term and created Word Clouds from these stored surrounding words. Word Clouds visually represent word frequencies by presenting more frequent words in proportionally larger text.

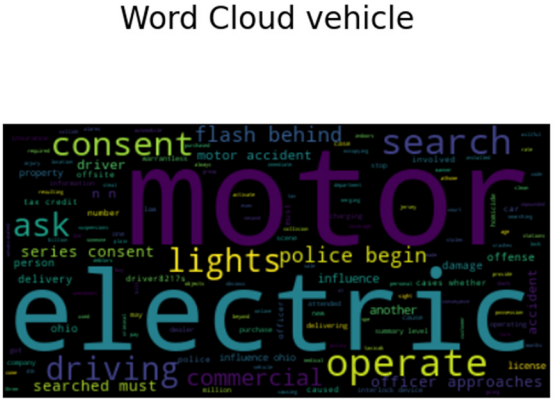
The concordance analysis was partially automated. We stored the ten words on both sides of each searched term to create a concordance text file with twenty-one-word phrases (with the searched term at the center). Having these concordance lines all listed, I also read directly from the website articles to ensure that the extracted language fit with what the public reading the articles would see and chiefly read in this manner. In manually reviewing, I mainly read the title, the sentence with the searched term, and the two sentences the term was sandwiched in between.

A. McBoyle v. U.S.

1. Collocation

McBoyle interprets the word "vehicle." Based on the Word Cloud below, the surrounding word "snapshot" that collocation provides indicates "vehicle" (or "vehicles") is used most in "motor vehicle" or "electric vehicle" contexts. These uses seem to "evoke in the common mind only the picture of vehicles moving on land."[79] But "motor vehicle" and "electric vehicle" also seem too broad. Justice Lee and Professor Mouritsen's collocation "vehicle" analysis provided more specific indicators of on-the-ground vehicle usage.

79. McBoyle, 283 U.S. at 27.



2. Concordance Lines

The "meat and potatoes" part of the CL analysis saved the day. 167 out of 170 uses of "vehicle" referred to automobiles, with the vast majority being a direct substitute to "car" used in a surrounding sentence. A few cases introduced some uncertainty about whether "electric vehicle" refers to "electric cars." Reading more surrounding text and seeing that it refers to ground transportation largely resolved these uncertainties. Here are some examples of "vehicle" in an automobile sense:

1. *In September 2021, North Carolina became the first state to amend its dealer statute to expressly permit at-home vehicle delivery in connection with the sale of new and used cars*.[80]
2. *Armando was pulling into a gas station and came very close to hitting another vehicle that was not in its lane. He had heated*

80. Jessica Berk & Alison Eggers, *Parameters for At-Home Vehicle Delivery by Dealers Codified in North Carolina*, JD SUPRA (Nov. 24, 2021), https://www.jdsupra.com/legalnews/parameters-for-at-home-vehicle-delivery-6536006/ [https://perma.cc/HJ2J-9HG7].

words with the driver and then parked his car to go into the convenience store. The two men in the vehicle he almost hit parked, exited their car, and started walking over towards Armando.[81]

3. *As cars evolve, users increasingly demand different functionalities, shifting the value from hardware to software. Based on the rapidly-changing consumer preferences, software may soon make up over 30 percent of vehicle content.*[82]
4. *Stealing a car is a property crime that comes with additional penalties. In New Jersey, an individual who steals a motor vehicle is required to pay a $500 penalty and has their license suspended for a year*.[83]
5. *Under the bill, Washington and Oregon are projected to receive billions in funds . . . $71 million for Washington to expand its electrical-vehicle charging network . . . .*[84]

When reading through the "electric vehicle" uses, it is important to note that charging stations refer to stations for electric *cars*. The three uses that did *not* use "vehicle" in the automobile sense used "vehicle" to describe a "freight vehicle" or as a legal term of art ("contract vehicle" and "special purpose vehicle"). At times, I expanded the context scope to resolve enduring uncertainties but

---

81. Christina Tsirkas, *When Accused of Aggravated Assault, Stand Your Ground*, LAWYERS.COM (Apr. 7, 2019), https://blogs.lawyers.com/attorney/criminal-law/when-accused-of-aggravated-assault-stand-your-ground-54495/ [https://perma.cc/UR9Y-H7HP].

82. Amir El-Aswad, *OESA Strategic Insights Executive Briefing Series 2021—Preparing Suppliers for Market Disruptions, Searching the Deep Sea to Electrify Vehicles, and Introducing the Lucid Air*, JD SUPRA (Dec. 9, 2021), https://www.jdsupra.com/legalnews/oesa-strategic-insights-executive-6708056/ [https://perma.cc/S672-9A4B].

83. Herbert Ira Ellis, *What Are the Different Types of Property Crimes In New Jersey?*, LAWYERS.COM (Nov. 12, 2020), https://blogs.lawyers.com/attorney/criminal-law/what-are-the-different-types-of-property-crimes-in-new-jersey-65566/ [https://perma.cc/C6L6-44F4].

84. Amber Novack & Tara O'Hanlon, *What the Infrastructure Bill Means for Transportation, Construction, and Real Estate in the Northwest*, JD SUPRA (Nov. 17, 2021), https://www.jdsupra.com/legalnews/what-the-infrastructure-bill-means-for-5329656/ [https://perma.cc/A4E8-J95U].

mainly kept the context to what I stated previously. People might disagree with my decision to expand the degree of context. But following Professor Nourse, context provides pragmatic enrichment essential to interpretation.[85]

These collocation results align with Justice Lee and Professor Mouritsen's results. The distinguishing feature is context. The searched word in the lawyer corpus relates to a statute or a legal issue (e.g., the potential crime in the *Armando* example). Here, lawyers couch their use of "vehicle" in a legal context more like an interpreting court will likely confront. "The need to consider context is a staple element of the judicial inquiry into ordinary meaning:"[86] Everyone "takes nonsemantic context—pragmatics—into account in deriving meaning from language."[87] Thus, relying on lawyer-written articles "pragmatically enriches meaning."[88]

B. U.S. v. Muscarello

1. Collocation

Muscarello interprets iterations of "carry" ("carry," "carries," "carried," and "carrying"). The snapshot provided by the Word Cloud does not control for a "syntax of a transitive verb, with the semantic features of a human subject and a weapon object."[89] But I account for this in the concordance analysis.

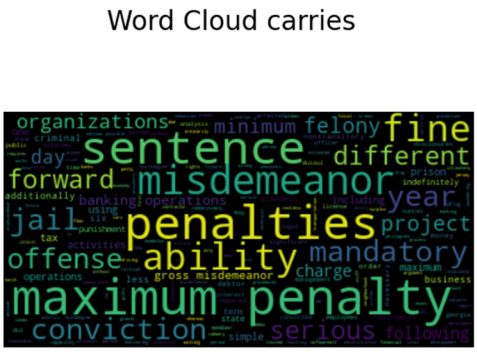
85. Victoria Nourse, *Picking and Choosing Text: Lessons for Statutory Interpretation from the Philosophy of Language*, 69 FLA. L. REV. 1409, 1412 (2017).

86. Lee & Mouritsen, *Judging Ordinary Meaning*, *supra* note 6, at 821.

87. *Id.* at 816.

88. Nourse, *supra* note 85, at 1418.

89. Lee & Mouritsen, *Judging Ordinary Meaning*, *supra* note 6, at 823.



Like in Justice Lee and Professor Mouritsen's analysis, the collates did not fall neatly within the "carries a firearm" semantic and syntactic structure. The collocates overwhelmingly refer to the more "metaphorical uses" of carries (e.g., felonies carrying certain penalties) that Justice Lee and Professor Mouritsen reference in their analysis.

2. Concordance Lines

I analyzed 85 uses of "carry." Most are used in the metaphorical sense ("fourth-degree crimes carry a prison sentence of up to 18 months and a fine of up to $10,000") or in the "carry out" (put into effect) sense ("the section authorizes $150,000,000 to be appropriated to the Secretary for fiscal years 2022–2026 to carry out future development"). Other "carry" uses include "carry forward" (temporally); "carry on" (endure); "carry over" (in a technical tax sense); and carries in terms of water forces (to direct the course of). "Carry" uses also refer to the inverse example that Justice Lee and Professor Mouritsen highlight (an inanimate object carries a human subject rather than vice versa). Of the 85 uses read, only six met the desired "carries a firearm" structure. I include each instance below.

1. *Months after voters overwhelmingly backed a referendum, the state now allows citizens over 21 years old to carry up to six*

> *ounces of marijuana legally. Police officers will not be allowed to arrest minors in possession or consuming cannabis, nor approach them if they smell it around the person.*[90]

This "carry" usage indicates that "carry" means to possess and have on one's person. The "smell it around the person" and "possession" language support this characterization. This example of "carry" usage does not seem to encompass having a gun locked in a car compartment.

2. *After all, the Constitution begins with "We The People of the United States." I have been carrying the same copy – now very well-worn – in my book bags and briefcases since 1976, when I started high school.*[91]

This example is trickier. One might argue that because the author transports the pocket Constitution in a carrier, which likely has not always been on his person since 1976, "carry" encompasses something less than having on one's person. But one could counter and say that carrying on one's person is part of a book bag or briefcase's nature. A stationary glovebox is quite different. And it seems that the author's point is that he always has the Constitution with him even if he uses some hyperbole to get there.

3. ***Carry all documents required for admission to the U.S.*** *Upon entry to the U.S., some entrants may need to show additional evidence of work authorization or government approval in addition to a passport and valid visa stamp.*[92]

---

90. Herbert Ira Ellis, *New Jersey Cannabis Bill Ends Arrests for Marijuana Possession*, LAWYERS.COM (Mar. 11, 2021), https://blogs.lawyers.com/attorney/criminal-law/new-jersey-cannabis-bill-ends-arrests-for-marijuana-possession-67079/ [https://perma.cc/RVY5-HX8L].

91. Louis Vlahos, *One Step Closer to "Building Back"—Where Do Federal Transfer Taxes Stand?*, JD SUPRA (Nov. 23, 2021), https://www.jdsupra.com/legalnews/one-step-closer-to-building-back-where-3466529/ [https://perma.cc/5Q6M-ZL9N].

92. Survi Parvatiyar et al., *Plan Ahead for Holiday Travel: 2021 Checklist for Foreign Nationals and Employers*, JD SUPRA (Nov. 18, 2021), https://www.jdsupra.com/legalnews/plan-ahead-for-holiday-travel-2021-1740855/ [https://perma.cc/XW4E-BN3H].

"Carry" here seems to indicate that the object ("documents") must be on one's person since it will need to be readily accessible to give to Homeland Security.

4. *It is important to note that the thief need not use or threaten another person with the weapon in order to be convicted of robbery—simply carrying the weapon or having it within one's control at the time of the offense is sufficient. When a weapon is used or brandished, or if harm is inflicted on someone during the course of a robbery, the offense can be elevated to "aggravated" status and more serious penalties apply*.[93]

The use of "or" (between "simply carrying the weapon" and "within one's control") is important. The author seems to use "or" to rephrase "carrying" as "within one's control" and equates the two. Or he is saying "within one's control" is a lesser degree of "carrying." Or the two could be distinct and unrelated. I do not think the third option is at play here. Rather "carrying" seems to be no less than "within one's control." Due also to the temporal "at the time of the offense" phrase, this "carry" usage does not seem to capture leaving a weapon in a locked glove box.

5. [A member of the public submitted a question for a lawyer to answer here. Because it is a harder interpretive question, I provide the full context as Nourse recommends.]

   Q: I was joyriding with some of a few old friends and state troopers had flashed their lights, but I didn't see them until they were behind me completely. Once I had announced to everyone in the car that they were behind us, the friend in the passenger passed back a black hat which unbeknown to me, had a gun in there. So eventually I see an exit ramp and try to come off, but I was going too fast and lost control of the car and crashed. Now I'm and [sic] being charged with fleeing or eluding an officer, firearm not to be

93. Daniel Williams, *Understanding Theft Crimes in Ohio*, LAWYERS.COM (Feb. 5, 2021), https://blogs.lawyers.com/attorney/criminal-law/understanding-theft-crimes-in-ohio-2-66798/ [https://perma.cc/J8S4-HUT5].

> *carried w/o license (no crim-violence) and reckless endangering another person.*
>
> *A: Lawyer up, immediately! Gun cases (Uniform Firearms Act) are taken seriously by the DA and not often bargained down. There is always reasonable doubt for an attorney to work with when there is a gun or drugs found in a car with multiple occupants. You need a good attorney to develop that "black hat" defense. It will be easier for everyone in the car if someone admits to possessing the gun.*[94]

This example is the hardest yet. One snag is that the incident occurred in a car. The key phrase is "passed back a black hat." This phrase indicates that someone had possession of the gun and physical control of it. The joyrider doesn't seem to be the one who had possession of the gun but the lawyer still advises him to "lawyer up, immediately" because of a likely violation of a statute.[95] The lawyer highlights that the carrying issue turns on a "black hat defense." It seems like the attorney views passing a hat with a gun in it as carrying the gun. To me, this "carry" use is distinguishable from a drug deal happening outside a car in which a gun is locked in a glovebox. But if not, this would only be one usage of "carry" that potentially aligns with the *Muscarello* majority's interpretation.

The lawyer corpus's concordance analysis aligns with Justice Lee and Professor Mouritsen's conclusion that "carry on one's person is overwhelmingly the most common use, while carry in a car is a possible but far less common use."[96] However, there is one vital distinction: Justice Lee and Professor Mouritsen's sample is 104 while mine is only six. This inhibits my ability to reach a well-

94. William R. Pelger, *How Do I Go About Handling A Fleeing and Eluding Charge?*, LAWYERS.COM (Nov. 03, 2019), https://blogs.lawyers.com/attorney/criminal-law/how-do-i-go-about-handling-a-fleeing-and-eluding-charge-58796/ [https://perma.cc/7GPP-4ECS].

95. *Id.*

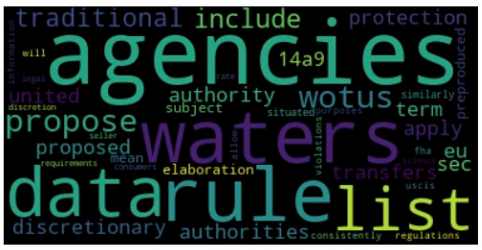
96. Lee & Mouritsen, *Judging Ordinary Meaning*, *supra* note 6, at 848.

formed opinion about how to interpret "carry." Increasing the size of the lawyer corpus will help alleviate this problem.

Considering the amount of non-desired "carry" senses that I read, Justice Ginsburg's argument about not isolating "carry" but including the full "carries a firearm" phrase seems more impactful after conducting the CL analysis. Justice Lee and Professor Mouritsen's method of factoring in the syntactic structure and semantic relationships is a happy medium between the majority and dissent, though. It allows more examples of "carry" to be analyzed while still controlling for features of the "carries a firearm" phrase.

C. Taniguchi v. Kan Pacific Saipan, Ltd.

Strangely, no instances of the word "interpreter" appeared in the entire corpus. This was not a bug in the code since we were able to assess iterations of "interpret" and create a Word Cloud from that (see below). Because there was no data, there was no CL analysis for *Taniguchi*. A future note or article would replace *Taniguchi* and analyze a third case—perhaps *U.S. v. Costello*[97] since Justice Lee and Professor Mourtisen assess that case in *Judging Ordinary Meaning*.[98] Ideally, a third case would have unique insights contrasting from Justice Lee and Professor Mourtisen's CL NOW corpus outcomes.



97. 350 U.S. 359 (1956).
98. *Id.* at 805.

CONCLUSION

The CL "ordinary lawyer" outcomes did not differ significantly from Justice Lee and Professor Mouritsen's "ordinary person" CL analysis. Although the sample sizes differed between the two corpora and the "ordinary lawyer" corpus enveloped more technical uses of words, readers would expect both of those differences. A key difference is the legal theory that underpins the two corpora.

The lawyer corpus operationalizes Professors McGinnis and Rappaport's original methods approach to interpretation into the world of CL. Although a seminal work, Justice Lee and Professor Mouritsen's *Judging Ordinary Meaning* ordinary person approach lacks a strong originalism tether beyond prioritizing public meaning—it does not account for original intent, original meaning, or original methods at the time of enactment. The ordinary lawyer corpus uses modern word usage to find meaning while alleviating fair notice concerns. It also accounts for the interpretive habit of modern and historical readers and writers of legal texts by deferring to lawyerly interpretations. But even more, Professors Baude and Sachs's law of interpretation, Justice Lee and others' register analysis, modern lawyer-client practices, and American history all support the original method approach of finding meaning through lawyerly understanding. Because of this, CL analyses should employ an ordinary lawyer corpus more often.

The key difference between the two corpora is context. Of all the concordance lines that I analyzed, each one used the searched term in a statutory, constitutional, or legal context. This differs significantly from the NOW Corpus. Because the NOW Corpus pools together words from magazines and news articles, many uses of the searched term do not relate to legal issues. Because the issues and contexts that the lawyer articles discuss are like issues that will eventually face courts, relying on lawyers' word usages and understandings seems beneficial to courts, compared to relying on the "ordinary person" understanding. Since words derive meaning

from their context, relying on a more legal semantic, syntactic, and pragmatic context helps courts find ordinary meaning.

As the "ordinary lawyer" corpus is in its early stages, I provide some insights to give the next relay leg a head start. First, transcribe JD Supra's lawyer podcasts. Doing so would increase the number of words that machine learning can scrape and store in a corpus. Next, use more Lawyers.com blogs that extend beyond matters of criminal law. These Lawyers.com blogs focus more on a broader public and use less technical senses of words. Finally, look for more lawyerly websites that meet the *Federalist Papers* criteria. The goal is to expand the size of the corpus without sacrificing parallels to *The Federalist*.