



ESSAY

The Faulty Frequency Hypothesis: Difficulties in Operationalizing Ordinary Meaning Through Corpus Linguistics

Ethan J. Herenstein*

Introduction

Promising to inject empirical rigor into the enterprise of statutory interpretation, corpus linguistics has, over the past couple years, taken the legal academy by storm.¹ A product of linguistics departments, corpus linguistics is an empirical approach to the study of language through the use of

* J.D. Candidate, Stanford Law School, 2019. M.S. Student, Stanford University Symbolic Systems Program, 2019. Many thanks to Bernie Meyler and David Sklansky for their guidance and encouragement; and to the editors of the *Stanford Law Review Online*, particularly Joe DeMott, Dan Brenner, and David Steinbach, for their helpful edits. All errors and omissions are my own.

1. For a sampling of the corpus-linguistics-cum-statutory-interpretation literature, see Thomas R. Lee & Stephen C. Mouritsen, *Judging Ordinary Meaning*, 127 *YALE L.J.* (forthcoming 2018), <https://perma.cc/WJ3C-WEUJ>; Stephen C. Mouritsen, *Hard Cases and Hard Data: Assessing Corpus Linguistics as an Empirical Path to Plain Meaning*, 13 *COLUM. SCI. & TECH. L. REV.* 156 (2011); and Lawrence M. Solan & Tammy A. Gales, *Finding Ordinary Meaning in Law: The Judge, the Dictionary or the Corpus?* (Brooklyn Law Sch. Legal Studies Research Papers, Research Paper No. 474, 2016), <https://perma.cc/H6FV-P4TW>. Note, too, that there is a related strand of originalist scholarship that seeks to leverage the tools of corpus linguistics to make the pursuit of the original public meaning of the Constitution more empirically rigorous. For theoretical justifications of originalism's use of corpus linguistics, see James C. Phillips et al., *Corpus Linguistics & Original Public Meaning: A New Tool to Make Originalism More Empirical*, 126 *YALE L.J. F.* 21 (2016); Lawrence M. Solan, *Can Corpus Linguistics Help Make Originalism Scientific?*, 126 *YALE L.J. F.* 57 (2016); and Lee J. Strang, *How Big Data Can Increase Originalism's Methodological Rigor: Using Corpus Linguistics to Reveal Original Language Conventions*, 50 *U.C. DAVIS L. REV.* 1181 (2016). For particular applications of corpus linguistics to originalist inquiries, see Jennifer L. Mascott, *Who are "Officers of the United States"?*, 70 *STAN. L. REV.* (forthcoming 2018) (using corpus methods to establish the original public meaning of the Appointments Clause), <https://perma.cc/7ZDF-L3D8>; and James C. Phillips & Sara White, *The Meaning of the Three Emoluments Clauses in the U.S Constitution: A Corpus Linguistic Analysis of American English, 1760-1799*, 59 *S. TEX. L. REV.* (forthcoming 2018) (using corpus methods to establish the original public meaning of the Emoluments Clause), <https://perma.cc/HTA4-FK6T>.

large, electronic, and searchable databases of text called corpora.² All sorts of naturally occurring language—novels, essays, poems, news articles—are stored in these corpora, enabling scholars to gather data regarding the actual frequency, context, and collocation of particular words or phrases.³ When faced with the question of the ordinary meaning of an ambiguous word or phrase in a statute, some scholars⁴ and judges⁵ now believe they can search these corpora to uncover how the contested phrase is ordinarily used—and therefore, crucially, commonly understood by ordinary speakers of the English language. These scholars and judges hope that relying on “patterns that emerge from corpus linguistic data”⁶—rather than a hodge-podge and arbitrary collection of dictionaries, newspaper articles, and, well, intuition—will ensure that the search for ordinary meaning will be a principled one.⁷

A forthcoming law review article titled *Judging Ordinary Meaning*, authored by Associate Chief Justice Thomas Lee of the Utah Supreme Court and his former law clerk Stephen Mouritsen, is the latest and most thorough contribution to the law-and-corpus-linguistics debate.⁸ In it, Justice Lee and Mouritsen offer a full-throated argument for the incorporation of corpus linguistics into statutory interpretation. Their justification for relying on corpus linguistics rests, in large part, on an intuitive linguistic hypothesis: “[I]f the search for ordinary meaning entails an analysis of the relative frequency of competing senses of a given term, then corpus linguistics seems the most promising tool.”⁹ In other words, Justice Lee and Mouritsen’s application of corpus linguistics depends on the premise that, where an ambiguous term retains two plausible meanings, the ordinary meaning of the term (and the one that ought to control) is the more frequently used meaning of the term. Call this the Frequency Hypothesis.

2. See Douglas Biber, *Corpus-Based and Corpus-Driven Analyses of Language Variation and Use*, in THE OXFORD HANDBOOK OF LINGUISTIC ANALYSIS 193, 193-94 (Bernd Heine & Heiko Narrog eds., 2d ed. 2015); Mouritsen, *supra* note 1, at 159.

3. See Biber, *supra* note 2, at 194-95.

4. See *supra* note 1.

5. Associate Chief Justice Thomas Lee of the Utah Supreme Court is in many ways the leader of this nascent law and corpus linguistics movement. Aside from his scholarship on the subject, Justice Lee has also employed corpus linguistics in judicial decisions. See, e.g., *State v. Rasabout*, 356 P.3d 1258, 1271 (Utah 2015) (Lee, J., concurring in part and concurring in the judgment); *In re Adoption of Baby E.Z.*, 266 P.3d 702, 724 n.21 (Utah 2011) (Lee, J., concurring in part and concurring in the judgment). I will return to Justice Lee’s advocacy for, and use of, corpus linguistics in Part II, below.

6. See Lee & Mouritsen, *supra* note 1 (manuscript at 26).

7. *Id.* (manuscript at 33-34).

8. See generally *id.*

9. *Id.* (manuscript at 34).

There are, however, good reasons to doubt the Frequency Hypothesis. Carissa Byrne Hessick has articulated one such reason in a forthcoming essay¹⁰ and a series of blog posts.¹¹ In short, Hessick argues that, at least with respect to the interpretation of criminal statutes, “looking to the frequency with which a term is used a certain way . . . creates problems of notice and accountability.”¹² There are notice concerns because criminal defendants, who are unlikely “to perform their own corpus searches and analyses,” will not be informed about the content of the laws they are accused of having broken.¹³ And there are accountability concerns because legislators might be ill-informed about the scope of the very laws they are tasked with passing.¹⁴

This Essay identifies and explores another, more fundamental reason to doubt the Frequency Hypothesis: A word might be used more frequently in one sense than another for reasons that have little to do with the ordinary meaning of that word. Specifically, a word’s frequency will not necessarily reflect the “sense of a word [or] phrase that is most likely implicated in a given linguistic context,”¹⁵ but will instead, at least partly, reflect the prevalence or newsworthiness of the underlying phenomenon that the term denotes. Whereas Hessick’s critique of the Frequency Hypothesis might be limited to the criminal context, my critique pertains to the entire enterprise. Accordingly, I am less optimistic than are Justice Lee and Mouritsen about corpus linguistics’s place in statutory interpretation.

I. The Frequency Hypothesis In Action

Despite the fact that Justice Lee and Mouritsen have not expressly embraced the Frequency Hypothesis,¹⁶ their support for corpus linguistics’s role in statutory interpretation effectively commits them to it. Moreover, Justice Lee has actually relied on the Frequency Hypothesis in his

10. See Carissa Byrne Hessick, *Corpus Linguistics and the Criminal Law*, 2018 BYU L. REV. (forthcoming 2018), <https://perma.cc/9GGA-7ZP3>.

11. See Carissa Byrne Hessick, *Corpus Linguistics and Criminal Law*, PRAWFSBLAWG (Sept. 6, 2017, 9:45 AM), <https://perma.cc/5MN2-Z2GV>; Carissa Byrne Hessick, *More on Corpus Linguistics and the Criminal Law*, PRAWFSBLAWG (Sept. 11, 2017, 1:01 PM), <https://perma.cc/3SEH-HURC>; Carissa Byrne Hessick, *Corpus Linguistics Re-Redux*, PRAWFSBLAWG (Sept. 25, 2017, 9:56 AM), <https://perma.cc/5LS2-WM8Z>.

12. See Hessick, *supra* note 10 (manuscript at 5).

13. *Id.* (manuscript at 6).

14. *Id.*

15. See Lee & Mouritsen, *supra* note 1 (manuscript at 5).

16. Instead, they have merely suggested that corpus linguistics’s usefulness in statutory interpretation depends, *pro tanto*, on the strength of the Frequency Hypothesis. See *id.* (manuscript at 34) (“[I]f the search for ordinary meaning entails an analysis of the relative frequency of competing senses of a given term, then corpus linguistics seems the most promising tool.”).

jurisprudence. *State v. Rasabout*,¹⁷ which afforded Justice Lee his most substantial opportunity to employ corpus linguistics, clearly illustrates this reliance.

At issue in *Rasabout* was a statutory prohibition against the unlawful “discharge of a firearm.”¹⁸ Andy Rasabout fired twelve shots at a house in a gang-related drive-by shooting, and the jury convicted him of twelve felony counts of unlawful discharge of a firearm.¹⁹ The only issue on appeal was the allowable unit of prosecution for the crime of unlawful discharge of a firearm: (1) each discrete shot expelled from the gun, or (2) the continuous intent that motivates one or more shots.²⁰ If (1) were correct, Rasabout could be convicted of twelve counts of the crime; if (2) were correct, Rasabout could be convicted of only a single count. The court’s task, thus, was to determine “the meaning of the term ‘discharge’ in the context of a ‘dangerous weapon or firearm.’”²¹ The majority, relying on traditional canons of construction, held that *discharge* refers to each discrete shot expelled from a gun.²² Accordingly, the court upheld Rasabout’s conviction of twelve counts of the unlawful discharge of a firearm.²³

In a concurring opinion, Justice Lee turned to corpus linguistics to help establish the ordinary meaning of *discharge*—“the meaning [*discharge*] would have in the mind of a ‘reasonable person familiar with the usage and context of the language in question.’”²⁴ Justice Lee searched the Corpus of Contemporary American English²⁵ to find examples of how people actually used the term *discharge*.²⁶ His search returned “eighty-six instances of the verb *discharge* within five words of the nouns *firearm*, *firearms*, *gun*, and *weapon*.”²⁷ Twelve of the results clearly “linked *discharge* to a single bullet.”²⁸ Sixteen of

17. 356 P.3d 1258 (Utah 2015).

18. *Id.* at 1260.

19. *Id.*

20. *Id.* at 1262.

21. *Id.* at 1263.

22. *Id.* at 1262.

23. *Id.*

24. *Id.* at 1272 (Lee, J., concurring in part and concurring in the judgment) (quoting *Olsen v. Eagle Mountain City*, 248 P.3d 465, 469 (Utah 2011)).

25. CORPUS OF CONTEMPORARY AM. ENGLISH, <https://perma.cc/6X9P-ET6D> (archived Nov. 28, 2017) (containing “more than 520 million words of text . . . equally divided among spoken, fiction, popular magazines, newspapers, and academic texts”).

26. *Rasabout*, 356 P.3d at 1281-82 (Lee, J., concurring in part and concurring in the judgment).

27. *Id.*

28. *Id.* at 1282.

the results strongly suggested that *discharge* referred to a single bullet.²⁹ Fifteen other results “seemed to imply a single shot.”³⁰ Thirty-six of the results were inconclusive.³¹ Only one “seemed consistent with the firing of multiple shots.”³² Through this corpus analysis, Justice Lee “confirmed that the single shot sense of [*discharge*] is overwhelmingly the ordinary sense of the term in this context.”³³ Justice Lee’s opinion rests on the Frequency Hypothesis: Because *discharge* is more frequently invoked in reference to a single shot, that is its ordinary meaning—and thus, its meaning in the criminal statute under which Rasabout was charged.

What’s more, Justice Lee is not the only judge who has relied on the Frequency Hypothesis. Justice Breyer³⁴ and Judge Posner³⁵ have both relied on it when conducting their own versions of corpus-style analysis.³⁶ In short, the Frequency Hypothesis is a common assumption—and one upon which Justice Lee’s use of corpus linguistics in large part depends. Because corpus linguistics’s primary contribution to statutory interpretation is to help empirically adjudicate between competing senses of an ambiguous term by uncovering the frequency of those senses, to the extent that the Frequency Hypothesis is flawed, so is the incorporation of corpus linguistics into statutory interpretation.

II. A Flaw in the Frequency Hypothesis

Intuitive as the Frequency Hypothesis may seem, the frequency of a particular term’s usage does not necessarily reveal that term’s ordinary meaning. To see why, let’s return to *Rasabout*. There, Justice Lee assumed that because *discharge* was more frequently used to refer to a single gunshot than to

29. *Id.* (deeming accidental discharges consistent with the single shot sense of *discharge*, “as it seems highly unlikely if not impossible that an accidental trigger-pull could result in a release of all of the bullets in a gun’s magazine”).

30. *Id.*

31. *Id.*

32. *Id.*

33. *Id.*

34. See *Muscarello v. United States*, 524 U.S. 125, 129 (1998) (treating the frequency with which a particular usage of the term *carry* appears in computerized newspaper databases as evidence of the term’s ordinary meaning).

35. See *United States v. Costello*, 666 F.3d 1040, 1044 (7th Cir. 2012) (treating the frequency with which various phrases containing the term *harboring* appear in Google search results as evidence of the term’s ordinary meaning).

36. Professor Randy Barnett, too, relies on the Frequency Hypothesis in his analysis of the original public meaning of the Commerce Clause. See Randy E. Barnett, *New Evidence of the Original Meaning of the Commerce Clause*, 55 ARK. L. REV. 847, 856-862 (2003). For more on Barnett’s corpus analysis, see *infra* Part II.B.

multiple gunshots, an ordinary speaker of the language would more likely have understood the term *discharge* as referring to a single gunshot.³⁷

But a word might be invoked more frequently in one sense than another for reasons that have little to do with the common understanding of that word. More specifically, the frequency with which a word carries a particular meaning will, at least partly, reflect the prevalence or newsworthiness of the underlying phenomenon that it denotes.

A. The Prevalence of the Underlying Phenomenon

Those who turn to corpus linguistics to help uncover the ordinary meaning of a disputed term rely on the Frequency Hypothesis—that the more frequently a particular usage appears in the corpus, the more ordinary that usage is. However, the frequency of a particular usage will also reflect the prevalence of the underlying phenomenon that the term denotes.

Suppose, for example, that it is more common for a person shooting a gun to fire a single bullet than an entire chamber. There would then be fewer opportunities for newspapers, magazines, journals, and talk shows (all of which are included in the corpus that Justice Lee used in *Rasabout*)³⁸ to invoke the term *discharge* in reference to multiple gunshots. But the relative frequency of the different senses of *discharge* would have less to do with the ordinary meaning of the term and more to do with the way guns are commonly used. So, the mere fact that Justice Lee, through corpus linguistics, discovered more instances of *discharge* relating to a single gunshot than relating to multiple gunshots does not tell us which of the two usages is the ordinary meaning of the term—that is, it does not necessarily reveal “the sense of a word [or] phrase that is most likely implicated in a given linguistic context.”³⁹ Frequency data, therefore, could obfuscate the term’s ordinary meaning by overweighting those usages that denote more prevalent phenomena—like the firing of a single gunshot over the firing of multiple gunshots. Thus, frequency data may not be as weighty as supporters of corpus linguistics hope.

37. *Rasabout*, 356 P.3d at 1281-82 (Lee, J., concurring in part and concurring in the judgment).

38. See CORPUS OF CONTEMPORARY AM. ENGLISH, *supra* note 25 (to locate, select “large and balanced” hyperlink in panel on right side of main page) (noting that the corpus “is evenly divided between the five genres of spoken, fiction, popular magazines, newspapers, and journals”).

39. Lee & Mouritsen, *supra* note 1 (manuscript at 5).

B. The Newsworthiness of the Underlying Phenomenon

A usage's frequency may also reflect the newsworthiness of the underlying phenomenon. Consider, in this light, Randy Barnett's corpus⁴⁰ survey of the *Pennsylvania Gazette's* archives to establish the original public meaning of the term *commerce*,⁴¹ as it appears in the Commerce Clause.⁴² Barnett's article emerged as a response to the Supreme Court's decision in *United States v. Lopez*,⁴³ in which the Court sought to determine the original public meaning of the term *commerce*.⁴⁴ There, Justice Thomas, in a concurring opinion, lamented that the Court's expansive reading of the Commerce Clause had "drifted far from [its] original understanding," which he took to consist only of "selling, buying, and bartering, as well as transporting for these purposes."⁴⁵ Following the decision, some academics questioned Justice Thomas's narrow reading of the term, claiming that, in fact, the original meaning of *commerce* included not only trade and transportation but also productive activity intended for trade.⁴⁶

In support of Justice Thomas's narrow interpretation of the Commerce Clause, Barnett, employing a form of corpus analysis,⁴⁷ examined every use of

40. While Barnett did not describe his methodology as corpus-based, Justice Lee and his co-authors have described Barnett's approach as "equivalent" to corpus linguistics. Phillips et al., *supra* note 1, at 27; *see also id.* at 24 (describing Barnett's methodology as "similar to the use of a specialized, unstructured corpus").

41. *See* Barnett, *supra* note 36, at 856-862. Barnett's corpus analysis centered on capturing the original public meaning of a constitutional term, rather than the ordinary meaning of a statutory term, *see id.* at 856, but for the purposes of this Essay nothing turns on that difference. Like Justice Lee's use of corpus linguistics in the statutory context, Barnett's use of corpus linguistics in the constitutional context relies on the Frequency Hypothesis. Therefore, my criticism of the Frequency Hypothesis applies equally to both. Moreover, Justice Lee has supported the application of corpus linguistics to constitutional interpretation. *See generally* Phillips et al., *supra* note 1, at 24 (arguing that corpus linguistics can help reveal the original public meaning of the Constitution).

42. U.S. CONST. art. I, § 8, cl. 3 (granting Congress the power to "regulate Commerce with foreign Nations, and among the several States, and with the Indian Tribes").

43. 514 U.S. 549 (1995).

44. *See id.* at 552-59; *see also* Barnett, *supra* note 36, at 848-49.

45. *Lopez*, 514 U.S. at 584-85. (Thomas, J., concurring).

46. *See, e.g.,* Grant S. Nelson & Robert J. Pushaw, Jr., *Rethinking the Commerce Clause: Applying First Principles to Uphold Federal Commercial Regulations but Preserve State Control over Social Issues*, 85 IOWA L. REV. 1, 101 (1999) (arguing that the original public meaning of "commerce" included "all market-based activities such as production, banking, and insurance").

47. I describe it as a "form" of corpus analysis because Barnett only searched the archives of a single newspaper. Such a narrow corpus does not satisfy the representativeness requirement of corpus construction. *See* SVENJA ADOLPHS & RONALD CARTER, SPOKEN CORPUS LINGUISTICS: FROM MONOMODAL TO MULTIMODAL 6 (2013) ("The corpus should be as representative as possible of the target language."); *see also supra* note 40.

the term *commerce* from 1728-1800 in the *Pennsylvania Gazette*, a widely circulated eighteenth-century newspaper.⁴⁸ Using a computer to exhaustively search the *Pennsylvania Gazette's* electronic archives, Barnett found that the term *commerce* appeared 1594 times, and in nearly every single instance the term was employed in Justice Thomas's narrow sense, rather than the broader sense championed by the Court.⁴⁹ Barnett took this "overwhelming consistency" of usage as powerful evidence of the term's original public meaning.⁵⁰

It is clear that Barnett was relying on the Frequency Hypothesis: He explicitly assumed that "[w]ere the term 'commerce' to have had a readily understood broad meaning, one would expect it to have made its appearance in this typical newspaper."⁵¹ It is important, however, to consider again what exactly a term's frequency reveals. For Barnett, the more frequently we encounter a particular usage of *commerce*, the more ordinary that usage is. But this assumption is not self-evident, nor, at any rate, does Barnett attempt to justify it. We might imagine that the narrow usage of *commerce*—pertaining only to trade and transportation—more frequently appeared in the newspaper because newspaper writers had greater reason to invoke the term in that sense.⁵² Why might this be the case? Perhaps the exchange of a commodity is more often newsworthy than its production. The frequency of the narrow meaning of *commerce* thus might reveal more about what the *Pennsylvania Gazette* took to be newsworthy than it does about the original public meaning of *commerce*. To the extent commercial transactions are more newsworthy than commercial production, Barnett's reliance on the Frequency Hypothesis is flawed; and, more importantly, so is his corpus-based conclusion that Justice Thomas's narrow meaning of *commerce* is the original public meaning. Once again, therefore, evidence derived from corpus analysis may not be as weighty as supporters of corpus linguistics hope.

48. Barnett, *supra* note 36, at 856-57.

49. *Id.* at 857-58 (finding that the usage of "commerce" almost exclusively referred to trading activity); *id.* at 861-62 (finding that just three of the 1594 appearances of *commerce* "suggested a possible broader meaning, though the content of whatever broader meaning they might convey is completely obscure").

50. *Id.* at 858, 862 ("[T]his survey clearly establishes that . . . the normal, conventional, and commonplace public meaning of commerce from 1728-1800 was 'trade and exchange,' as well as transportation for this purpose.").

51. *Id.* at 857.

52. The newsworthiness argument is not limited to newspapers. Magazines and news shows—both of which are included in the Corpus of Contemporary American English, *see supra* note 25—are both susceptible to the same newsworthiness argument.

III. Fixing the Frequency Hypothesis?

A. Broadening the Inquiry Beyond the Corpus

Those set on salvaging the Frequency Hypothesis could partly neutralize this argument by incorporating the prevalence and newsworthiness of the underlying phenomena into their corpus analysis. For example, instead of merely aggregating instances of *discharge* and comparing the frequency of the two competing usages, scholars could try to account for how often actual gun users discharge a single bullet and how often they discharge an entire magazine. Similarly, scholars could attempt to quantify or otherwise measure the newsworthiness of particular phenomena. Incorporating the prevalence and newsworthiness of the underlying phenomenon into the corpus analysis would obviate the worry that those factors distort the results of the corpus analysis. In essence, scholars could discount the frequency of a term's usage by the prevalence and newsworthiness of the underlying phenomenon to which the term refers.

This approach raises two problems of its own, however. First, it is far from clear how scholars would go about implementing this solution. How, in other words, would they ascertain trends in gun usage? The social sciences would be of assistance, surely, but studying gun usage could be a lot messier than studying word usage.⁵³ For example, what sort of gun users—police officers? gang members? hunters?—would be considered? And measuring the newsworthiness of underlying phenomena would lead the analysis into even murkier methodological territory. Scholars would struggle to settle on principled means to determine how to weight the respective coefficients of this discount formula.

This leads to the second problem. Expanding the interpretive inquiry beyond the confines of the corpus threatens to undermine the very reason why scholars and judges, like Justice Lee and Mouritsen, turned to corpus linguistics in the first place: to inject empirical rigor into the interpretive process.⁵⁴ Faced with two competing usages of a particular term, Justice Lee and Mouritsen look to corpus linguistics for an answer. By promising a principled and systematic method to ascertain the ordinary meaning of words, corpus linguistics is supposed to blunt methodological criticisms of statutory interpretation. But salvaging the Frequency Hypothesis—by expanding the inquiry into gun usage and newsworthiness and whatever other underlying

53. See generally JOHN LAW, AFTER METHOD: MESS IN SOCIAL SCIENCE RESEARCH (2004) (observing that the social sciences are messy).

54. See *State v. Rasabout*, 356 P.3d 1258, 1277 (Utah 2015) (Lee, J., concurring in part and concurring in the judgment) (advocating for “an empirical check on [judges’] (imperfect) linguistic intuition”); Lee & Mouritsen, *supra* note 1 (manuscript at 54) (claiming that corpus linguistics provides “an empirical ground” for conclusions about ordinary meaning).

phenomena might affect the frequency of a term's usage—now threatens to bolster such criticisms. The Frequency Hypothesis now demands the very thing supporters of corpus linguistic hope to avoid: nonlinguistic, messy, real-world facts. And, again, scholars would struggle to settle on principled means to determine how heavily to discount the prevalence and newsworthiness of the underlying phenomenon in determining the ordinary meaning of the term. Intuition would, once again, infect statutory interpretation. Salvaging the Frequency Hypothesis in this manner might thus prove a pyrrhic victory; it would leave a corpus linguistics that bolsters, rather than blunts, methodological criticisms of statutory interpretation.

B. Broadening the Inquiry Within the Corpus

There is a second approach—a methodological adjustment—for which defenders of the Frequency Hypothesis might advocate. Instead of relying on the search results for only the disputed term, scholars could conduct a more robust corpus analysis for synonymous terms in order to determine whether the frequency data for the disputed term reflects the prevalence and newsworthiness of the underlying phenomenon or the fact that the disputed term is peculiarly used in a particular sense. For example, in his efforts to determine the ordinary meaning of *discharge*, Justice Lee could have also come up with a list of synonyms—say, *fire* and *shoot*—and gathered data regarding the usage of these terms as well. This additional data would help Justice Lee determine whether *discharge* is peculiarly associated with firing a single shot or whether synonymous verbs are also more frequently invoked in reference to a single shot than multiple shots (which would suggest that multiple shots may be less prevalent or less newsworthy than single shots). Similarly, in Barnett's efforts to determine the original public meaning of *commerce*, he could have conjured up a list of activities included in the broader sense of *commerce*—like *production* and *manufacturing*—and gathered data regarding the usage of these terms as well. This additional data would help Barnett determine whether the newspaper did report these activities but simply did not describe them as *commerce*, or whether these activities—under any name—just do not appear as frequently in the corpus (suggesting that such activities are either less prevalent or less newsworthy than trade and transportation). This methodological adjustment would reduce the worry that the frequency of usage's appearance in a corpus surreptitiously reflects the pervasiveness or newsworthiness of the underlying phenomenon rather than the ordinariness of that usage.

But the improvements purchased by this methodological adjustment do not come without costs. First, the process of conjuring up a list of synonyms might reintroduce the very subjectivity that corpus linguistics is intended to avoid. This is because it is possible, even likely, that scholars will not agree on which additional terms are relevant for this more robust corpus analysis. While there are corpus tools—such as a collocative search, which reveals

“what words occur near other words”⁵⁵—that would help to empiricize this selection process, these tools will not always eliminate discretion. For example, while a search for collocates of *firearm* would empirically reveal *fire* and *shoot* as substitute search terms for *discharge*, it is less obvious how to systemically identify the universe of activities included in the broader sense of *commerce*. To the extent that different people have different ideas about what activities are potentially forms of commerce, the corpus analysis might reflect the linguistic intuitions of those engaged in the corpus analysis—which, of course, is precisely what corpus linguistics is intended to avoid.⁵⁶

Second, assuming that this broadened corpus analysis is workable, it is possible that it will result in even stronger evidence against the Frequency Hypothesis. If this more robust corpus analysis fails to reveal a peculiar usage of the disputed term, that is strong evidence that the frequency data does not reflect the ordinariness of the usage but some combination of its prevalence and newsworthiness. To reiterate an example from above, if *fire* and *shoot* are as frequently invoked in reference to single shots as is *discharge*, that is especially strong evidence that the frequency with which *discharge* is invoked in reference to single shots is a reflection not of the ordinariness of that usage but instead of some combination of the prevalence and newsworthiness of single shots. Thus, while I encourage defenders of the Frequency Hypothesis to adopt this methodological adjustment, I suspect its results will provide more reason to doubt the Frequency Hypothesis.

Conclusion

There are two important points to take away from this Essay. First, and most importantly, the Frequency Hypothesis is contestable. Justice Lee, Mouritsen, and other corpus-linguists-cum-legal scholars have yet to explain, much less defend, their reliance on the Frequency Hypothesis. Instead, they have implicitly relied on the intuitiveness of the Frequency Hypothesis. And that is precisely why it is important to recognize the Frequency Hypothesis for what it is: a contestable hypothesis about the ordinary meaning of words—and one which scholars ought to contest. Second, this Essay offers a particular argument against the Frequency Hypothesis: The frequency with which a term’s usage appears in a corpus will, at least partly, reflect the prevalence or newsworthiness of the underlying phenomenon that the term denotes.

Before judges and scholars embrace corpus linguistics as a tool for statutory interpretation, there ought to be a more thorough reckoning of the Frequency Hypothesis. I hope this Essay helps to start that conversation.

55. CORPUS OF CONTEMPORARY AM. ENGLISH, *supra* note 25 (to locate, select “Collocates”); see, e.g., Lee and Mouritsen, *supra* note 1, at 40-43 (identifying *electric*, *motor*, and *plug-in* as the words most commonly collocated with—that is, found nearby—*vehicle*).

56. See *supra* text accompanying notes 6-7.