2016

Can Corpus Linguistics Help Make Originalism Scientific

Lawrence Solan
Brooklyn Law School, larry.solan@brooklaw.edu

Recommended Citation
126 Yale L.J. F. 57 (2016-2017)

THE YALE LAW JOURNAL FORUM
MAY 27, 2016

CAN CORPUS LINGUISTICS HELP MAKE ORIGINALISM SCIENTIFIC?

Lawrence M. Solan

I. A NEW CORPUS OF FOUNDING ERA TEXTS

James Phillips, Daniel Ortner, and Thomas Lee begin their engaging essay, *Corpus Linguistics & Original Public Meaning: A New Tool To Make Originalism More Empirical*,[1] by pronouncing originalism "the predominant interpretive methodology for constitutional meaning in American history."[2] They then describe and attempt to justify a new tool to improve originalist methodology: a large corpus of Founding-era documents, representative of a host of genres available to educated people of that period. As their title suggests, the brand of originalism they set out to improve is the version at times dubbed "the new originalism"[3]—an iteration that seeks to construe the Constitution in accordance with the understanding of the state constitutional convention members who read its words and heard its supporters at the time.

This brief Essay expresses support for the project, but also focuses on its limitations in advancing originalist argumentation. While better empirical tools for determining original public meaning are valuable, they only get us so far, as a) there may be multiple original public meanings or no clear meaning that emerges from the corpora; b) we are lacking a coherent theory to justify when one original public meaning rather than another should be relied upon; and c) for abstract concepts such as "abridging the freedom of speech," which we are likely to encounter in the constitutional context, it is unclear whether the original meaning ought to be interpreted thickly to include specific examples of the concept or thinly to define only the concept itself.

1. 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. F. 21 (2016).
2. *Id.* at 21.
3. *See, e.g.*, *Symposium: The New Originalism in Constitutional Law*, 82 FORDHAM L. REV. 371, 371 (2013).

The new corpus, COFEA (Corpus of Founding Era American English), will feature at least 100 million words of text written between 1760 and 1799, taken from a variety of sources.[4] The project will be the third publicly available research corpus of general American English created by linguistic scholars from Brigham Young University, supplementing COCA (Corpus of Contemporary American English) (beginning 1990)[5] and COHA (Corpus of Historical American English) (covering 1820 through 1989).[6] The goal of this project is to provide legal theorists with a research tool better able to reveal "original pubic meaning" than either the Founding-era dictionaries relied upon by legal scholars and judges today[7] or the even less reliable practice of extrapolating from a small sample of instances of a word or phrase's usage.

The argument that the new corpus will improve originalist methodology is straightforward: if scholars want to investigate how the public likely understood the Constitution's words, then scholars would benefit from examining the data contained in a large corpus of English from that era rather than only examining the snapshot that a lexicographer took—a method for which Justice Scalia's originalism received substantial criticism.[8] Furthermore, COFEA will likely come with its own software that permits not only searches of individual words, but also searches of words that co-occur in proximity to one another.[9] This tool makes it possible to take into account syntactic and semantic structures larger than single words.

4. Phillips, Ortner & Lee, *supra* note 1, at 31.

5. THE CORPUS OF CONTEMPORARY AMERICAN ENGLISH, BYU, http://corpus.byu.edu/coca/ [https://perma.cc/6JKW-QLPH].

6 THE CORPUS OF HISTORICAL AMERICAN ENGLISH, BYU, http://corpus.byu.edu/coha/ [https://perma.cc/5HPR-FTNU].

7. *See* Gregory E. Maggs, *A Concise Guide To Using Dictionaries from the Founding Era To Determine the Original Meaning of the Constitution,* 82 GEO. WASH. L. REV. 358 (2014).

8. *See* Phillip A. Rubin, Note, *War of the Words: How Courts Can Use Dictionaries Consistent with Textualist Principles*, 60 DUKE L.J. 167, 200-06 (2010). As Phillip Rubin observes, "Justice Scalia chose definitions (that 'arms' means any kind of weapon, and that 'keep arms' means to have such weapons) and invoked the dictionary to say that those meanings were correct *because* the dictionary contained them. But the extent of what a dictionary can be used to say about the matter is that the words *could* have the meanings Justice Scalia attributed to them—not that they *must* have those meanings in a given context." *Id.* at 202.

9. These are the tools available as part of COCA and COHA, COFEA's sister corpora. *See* THE CORPUS OF CONTEMPORARY AMERICAN ENGLISH, *supra* note 5; THE CORPUS OF HISTORICAL AMERICAN ENGLISH, *supra* note 6. Moreover, apart from the BYU corpora, the field of corpus linguistics has developed a host of tools designed specifically to make such tasks possible with either corpora that have already been developed by linguists and other scholars, or with corpora developed by scholars for particular research projects. *See, e.g.*, TONY MCENERY, RICHARD XIAO & YUKIO TONO, CORPUS-BASED LANGUAGE STUDIES: AN ADVANCED RESOURCE BOOK (2006); GRAEME KENNEDY, AN INTRODUCTION TO CORPUS LINGUISTICS (1998).

Moreover, as the Essay's authors argue, judges infer meaning from corpora even now. In his majority opinion in *Muscarello v. United States*,[10] Justice Breyer surveyed literature, the Bible, and contemporary newspapers to determine the ordinary meaning of the word "carry."[11] Judge Posner, in his Seventh Circuit opinion in *United States v. Costello*,[12] used Google searches to demonstrate that the verb "to harbor" had a meaning narrower than "to house." Finally, one of the authors, Utah Supreme Court Associate Chief Justice Lee, used COCA as a tool to determine the ordinary meaning of "discharge" in his concurring opinion in *State v. Rasabout*.[13]

If judges are already using various corpora to determine a word's ordinary meaning in the context of statutory interpretation, the authors argue, then scholars should develop a corpus with accompanying tools so that the task can be accomplished at a higher level of precision and professionalism. The authors are correct on this point. Whatever one's commitment to new originalism, its proponents have every reason to develop its methods to enhance the empirical basis of claims that one interpretation of the Constitution better effectuates its original public meaning than another. Moreover, like dictionaries, the corpus is neutral in the sense that those whose writing contributes to it had no agenda with respect to the constitutional debates that occur now, some 250 years after the texts were written. For these reasons, COFEA is a promising tool.

II. THE CORPUS AS SOURCE MATERIAL FOR A FOREIGN LANGUAGE DICTIONARY

Yet a tool is only a tool, and the authors acknowledge some of COFEA's limitations. First, the authors acknowledge that a general corpus is not very helpful when defining legal terms of art.[14] For these terms, legal sources are superior. Second, even after using the corpus, originalists must still exercise judgment to determine how the various occurrences of words or phrases should inform their meaning in the Constitution.[15] Third, and most importantly, the authors recognize that originalism is under-theorized in the

10. 524 U.S. 125 (1998).

11. *Id.* at 129-30. The fact of this effort is more impressive than its execution. For criticism from the perspective of a corpus linguist/practicing lawyer, see Stephen C. Mouritsen, *The Dictionary Is Not a Fortress: Definitional Fallacies and a Corpus-Based Approach to Plain Meaning*, 2010 BYU L. REV. 1915.

12. 666 F.3d 1040, 1044 (7th Cir. 2012).

13. 356 P.3d 1258, 1281-82 (Utah 2015) (Lee, J., concurring).

14. Phillips et al., *supra* note 1, at 29.

15. *Id.* at 30

sense that it typically chooses the most typical meaning as the target but does not adequately defend that choice.[16]

The second and third caveats are closely related and merit further discussion. The words and phrases used during the Founding do not necessarily have the same meaning that they have today, just as the meanings of the words in Shakespeare's plays are not always the same as the meanings of those same words today. Otherwise, there would be no reason to resort to earlier texts. We could instead use COCA, or *Webster's Third New International Dictionary*. Of course, we may discover in our research that eighteenth-century English and twenty-first-century English have a lot of vocabulary in common. Scholars must first assume, however, that the meanings of words may have changed over time.

Lawrence Solum, a leading theorist of new originalism, makes this point in his description of the originalist method:

> If we want to know what a text means and the text was not written very recently, we need to be aware of the possibility that it uses language somewhat differently than we do now. Moreover, meaning is in part a function of context—and context is time-bound. So if we want to know what a text means, we need to investigate the context in which the text was produced.[17]

A nuanced way to approach the problem is to become lexicographers of the moment, constructing definitions from a large corpus of this foreign language, using the tools of corpus linguistics to determine which terms are typically used together, which senses of a word predominate, and so on. Professor Lawrence Lessig has argued for a translator's perspective to be taken generally in constitutional interpretation.[18] In fact, the claim that the Constitution was written and discussed in a foreign language is not as remote from the truth as it may at first appear. Versions were circulated in both German and Dutch, and comparison of those versions to the English version can be instructive when it comes to understanding what the drafters intended.[19] However, as Jack Balkin

16. *Id.* The authors add a fourth caveat—that corpora currently available to researchers do not adequately represent Founding-era documents. However, COFEA is intended to solve that problem, so I do not discuss it further.

17. Lawrence B. Solum, *The Fixation Thesis: The Role of Historical Fact in Original Meaning*, 91 NOTRE DAME L. REV. 1, 20 (2015).

18. See Lawrence Lessig, *Fidelity in Translation*, 71 TEX. L. REV. 1165, 1189 (1993) ("[T]ranslation is a practice that neutralizes the effect of *changed language* on a text's meaning, where language is just one part of context, and changed language is just one kind of change in context.").

19. *See* Christina Mulligan et al., *Founding Era Translations of the Constitution*, 31 CONST. COMMENT. 1 (2016).

points out,[20] whether one is deciding what to make of the Dutch and German versions or what to make of differences in English between the Founding era and today, interpretive decisions must be made, and these decisions are by no means theoretically neutral. Let us focus on some of them.

First, lexicography is not cut and dried. Lexicographers must make many kinds of judgments. How many tokens of a word or senses of a word must be present before one can responsibly infer a definition?[21] What if there are too many tokens so that some kind of sampling procedure is needed to make sense of the data without losing one's sense of neutrality?[22] What happens if a word or a sense of a word appears disproportionally in one sort of document but not in others?[23] To what extent are examples expansions of the same sense, or entirely different senses?[24] Lexicographers must make all of these judgments and many more.

Second, the decision to assign a word its ordinary meaning rather than a more expansive meaning is a substantive decision with significant interpretive consequences. Assume that the corpus reveals that the phrase "bear arms" was more often than not used in military contexts, but was not restricted to military contexts. What then? Should the interpreter prefer the phrase's narrower, ordinary meaning and limit the Second Amendment's protections to the military context? Should the interpreter prefer the phrase's broader meaning and extend Second Amendment protections to the home? The corpus does not help resolve this interpretive dilemma. Of course, it is better to know these facts than to infer them from less robust data. However, once one commits to original public meaning as a principle of construction, one discovers that there are many original public meanings of an expression, and the corpus does not provide much help in selecting among them. Recent debates over the meaning of "commerce" in the Constitution illustrate the problem.[25]

---

20. Jack M. Balkin, *The Construction of Original Public Meaning*, 31 CONST. COMMENT. 73 (2016).

21. *See, e.g.*, SIDNEY I. LANDAU, DICTIONARIES: THE ART AND CRAFT OF LEXICOGRAPHY 296 (2d ed. 2001).

22. *Id.*

23. Lexicographers speak of "representativeness." *See* BO SVENSEN, A HANDBOOK OF LEXICOGRAPHY: THE THEORY AND PRACTICE OF DICTIONARY-MAKING 64 (2009). As linguist Kevin Tang has pointed out to me in personal communication, the word "asparagus" may appear 100 times in a corpus if the corpus contains ten cookbooks, but that is not the same as a word that appears the same number of times but that is more evenly distributed among genres. Corpus linguists have developed computational tools to adjust for these differences in dispersion among words equally represented in a corpus. *See, e.g.*, Stefan Th. Gries, *Dispersions and Adjusted Frequencies in Corpora*, 13 INT'L J. CORPUS LINGUISTICS 403 (2008).

24. *See* LANDAU, *supra* note 21, at 337-38; Christian M. Meyer & Iryna Gurevych, *Wiktionary: A New Rival for Expert-Built Lexicons? Exploring the Possibilities of Collaborative Lexicography*, *in* ELECTRONIC LEXICOGRAPHY 259, 283-89 (Sylviane Granger & Magali Paquot eds., 2012).

25. See Randy Barnett, *The Original Meaning of the Commerce Clause,* 68 U. CHI. L. REV. 101 (2001) (arguing for a narrow reading based on an analysis of Founding-era documents);

We see analogous problems in the application of the "ordinary meaning rule" in statutory interpretation. Consider *Chisom v. Roemer*,[26] a 1991 Supreme Court case interpreting the Voting Rights Act. In Louisiana, state supreme court justices are elected.[27] Section 2 of the Voting Rights Act applies to the election of "representatives."[28] The plaintiffs argued that Louisiana's at-large election structure for electing justices violated the Act.[29] The defendants argued that the Act did not apply to the election of judges, since judges are not representatives.[30] The case made its way to the U.S. Supreme Court, where a majority of Justices held that the Act did apply, relying both on the law's purpose and on its stated goal of overriding a Supreme Court precedent that had construed the Voting Rights Act narrowly.[31] In dissent, Justice Scalia sharply criticized the majority for not adhering to the ordinary meaning rule.[32]

As a linguistic matter, Justice Scalia was right. Judges do not come within the ordinary meaning of "representative." But the majority was also right in its arguments. There was no reason to believe Congress intended to leave a safe harbor for racism-infected elections of judges. The case boiled down to whether ordinary meaning is a good first approximation of what a legislature intended to communicate, or whether it is a rule of interpretation to be followed as a matter of stare decisis. If the former, then ordinary meaning is defeasible if more specific historical and contextual information suggests that the legislature intended a broader interpretation. If the latter, it is not.

With or without a corpus, originalism presents the same problem as did the Voting Rights Act. There will always be lexicographic decisions to be made about how narrowly or broadly to define a term. These decisions are not neutral. The lexicographer will take into account the dictionary's purpose, audience, and financial resources that may limit the length of permissible definitions. The constitutional interpreter will take into account prior commitments to what counts as a legitimate argument, and will also have to decide the circumstances, if any, under which the ordinary sense of a term can be overridden, leading to a more expansive understanding, if individual

Jack Balkin, *Commerce*, 109 MICH. L. REV. 1, 15-29 (2010) (arguing for a broad reading based on a combination of structure, usage, and dictionary definitions); Robert G. Natelson, *Commerce in the Commerce Clause: A Response to Jack Balkin*, 109 MICH. L. REV. FIRST IMPRESSIONS 55 (2010) (arguing that Balkin focuses excessively on evidence that supports a broad interpretation, missing the fact that a narrow reading follows from ordinary usage).

26. 501 U.S. 380 (1991).

27. La. Const., Art. 5, § 22(A).

28. 52 U.S.C. § 10301(b) (1982).

29. *Chisom*, 501 U.S. at 385.

30. *Id.* at 398-99.

31. *Id.* at 403-04.

32. *Id.* at 405 (Scalia, J., dissenting).

inquiry into the social history of the time suggests that such a move is more likely to be faithful to the intent of the Framers and how they were understood.[33] Nothing in the corpus, or in the methods of corpus-driven lexicography, demands one result or another.

Third, a thorny semantic problem appears to be impervious to the introduction of corpus linguistics into constitutional analysis: To what extent does the meaning of a word include an understanding of the members of the category that the word denotes? Put differently, does the meaning of an abstract concept include concrete instantiations? Dictionaries differ in their commitment to considering examples as part of a definition.[34] In the constitutional realm, the issue arises, for example, in deciding whether the Eighth Amendment's prohibition against "cruel and unusual punishments"[35] should include at least a partial list of acceptable and unacceptable punishments at the time of the Founding, or whether it should be understood abstractly, meaning something like "punishment harsher than acceptable norms would permit."

Lawrence Solum takes the position that interpretation should in this sense be thin, noting that "the facts to which the text can be applied change over time."[36] Jack Balkin, whose "living originalism" espouses thin interpretation more generally in order to be at once faithful to the Founders and responsive to change, agrees.[37] Others who do not subscribe to originalism share the view that the meanings of constitutional terms should not include Founding-era understandings of what came within the concept and what did not.[38] Yet Scalia's arguments supporting the constitutionality of the death penalty today is replete with reference to its ubiquity in the eighteenth and nineteenth centuries.[39] The problem is a linguistic one: words or phrases describing a

33. *See* Ian Bartrum, *Two Dogmas of Originalism*, 7 WASH. U. JURISPRUDENCE REV. 157 (2015) (arguing that judges should not prioritize fixing the text's semantic meaning in a historical moment nor allow a text's fixed semantic meaning to constrain the construction of legal rules); Jonathan Gienapp, *Historicism and Holism: Failures of Originalist Translation*, 84 FORDHAM L. REV. 935 (2015) (arguing that originalism needs a better grounding in the historical method to properly ascertain true original meaning and avoid atomistic translation).

34. *See* SVENSEN, *supra* note 23, at 281-88.

35. U.S. CONST. amend. VIII.

36. Solum, *supra* note 17, at 21.

37. JACK M. BALKIN, LIVING ORIGINALISM 6-7, 100-101 (2011).

38. *See, e.g.*, Robert W. Bennett, *Originalism and the Living American Constitution*, *in* ROBERT W. BENNETT & LAWRENCE B. SOLUM, CONSTITUTIONAL ORIGINALISM: A DEBATE 78 (2011).

39. *See* Antonin Scalia, *Common-Law Courts in a Civil-Law System: The Role of United States Federal Courts in Interpreting the Constitution and Laws*, *in* A MATTER OF INTERPRETATION: FEDERAL COURTS AND THE LAW 46 (Amy Gutmann ed., 1997); Antonin Scalia, *Response*, *in* A MATTER OF INTERPRETATION: FEDERAL COURTS AND THE LAW 132, 145-46; *see also* Baze v.

category may be understood either abstractly or as a function of the category's members. The solution, however, is not linguistic at all. Rather, it requires a decision as to how responsive constitutional interpretation should be to changes in political and social norms over time.

CONCLUSION

The Founding-era corpus project is a good one. It will reduce the reliance on eighteenth- and nineteenth-century dictionaries and the temptation to select among them strategically. Moreover, by making some of the tools of corpus linguistic analysis available to the community of constitutional scholars, the new corpus will encourage analysts to look not only at the single word, but also at the linguistic context in which the word occurs. All of this should bring the practice of originalist analysis closer to its goal of discovering original public meaning.

But perhaps not *much* closer in many instances. Like the lexicographer, the originalist, having found either too few or too many instances of a word in the corpus, will have to decide what constitutes original public meaning. And like the lexicographer, the originalist will have other choices to make about how narrowly or broadly, thinly or thickly, to construe a relevant word. These choices are not strictly linguistic. They depend upon the commitments of the corpus's user, and these commitments depend upon the user's stance with respect to the language being analyzed.

Still, at the end of the day, it is hard to imagine that this wealth of new information will fail to add value to constitutional discourse. At the very least, the corpus will likely provide sufficiently rich new information to generate healthy, open debate about what constitutes good constitutional analysis.

Lawrence M. Solan is the Sidley Austin – Robert D. McLean '70 Visiting Professor of Law at Yale Law School, the Don Forchelli Professor of Law and Director of the Center for the Study of Law, Language and Cognition at Brooklyn Law School. *He wishes to express my gratitude to Christine Kwon for her valuable contributions as my research assistant. He also thanks Jack Balkin, Edward Finegan, Tammy Gales, and Christina Mulligan for their helpful discussion and suggestions.*

Preferred Citation: Lawrence M. Solan, *Can Corpus Linguistics Help Make Originalism Scientific?*, 126 YALE L.J. F. 57 (2016), yalelawjournal.com/forum/can-corpus-linguistics-help-make-originalism-scientific.

Rees, 553 U.S. 35, 94 (2008) (Thomas, J., concurring) (citing STUART BANNER, THE DEATH PENALTY: AN AMERICAN HISTORY 23 (2002)).