CORPUS LINGUISTICS AND *HELLER*

James C. Phillips[*] *and Josh Blackman*[**]

In District of Columbia v. Heller*, the Supreme Court sharply divided over the meaning of the twenty-seven words in the Second Amendment. Justice Scalia wrote the majority opinion. He concluded that the Second Amendment "protects an individual right to possess a firearm unconnected with service in a militia, and to use that arm for traditionally lawful purposes, such as self-defense within the home." In short, an "individual" right. Justice Stevens, in his dissent, contended that the Second Amendment "is most naturally read to secure to the people a right to use and possess arms in conjunction with service in a well-regulated militia." That is, a "collective" right.*

Justice Scalia and Justice Stevens both made linguistic claims about four elements of the Second Amendment: "right of the people," "keep and bear arms," "keep arms," and "bear arms." Both the majority and the dissent used various textualist approaches to consider these four phrases, but their toolkit in 2008 was limited. They considered only a fairly narrow range of sources to interpret the text. Today, we can do better. In this Article, we will grade the four linguistic claims made in the Heller case using corpus linguistics.

We rely on the Corpus of Founding Era American English ("COFEA"). In 2015, one of us conceptualized and oversaw the initial development of COFEA. We performed five queries with COFEA. First, we queried right of the people*. Second, we queried* keep and bear arms *(and synonyms). Third, we queried the word* right *within six words of* arms*. Fourth, we queried the word* keep*, and variants of* keep*, within six words of* arms*. Fifth, we queried*

* Assistant Professor, Chapman University Dale E. Fowler School of Law.

** Professor, South Texas College of Law Houston. We presented a version of this paper at the 2019 Brigham Young University Corpus Linguistics Conference and the 2019 Originalism Works in Progress Conference. We published our preliminary results in Josh Blackman & James C. Phillips, *Corpus Linguistics and the Second Amendment*, HARV. L. REV. BLOG (Aug. 7, 2018), https://blog.harvardlawreview.org/corpus-linguistics-and-the-second-amendment/.

the word bear*, and variants of* bear*, within six words of* arms*. We used multiple coders who independently coded their results using a type of double-blind methodology.*

Both the majority and the dissenting opinions erred with respect to some of their linguistic claims. Justices Scalia and Stevens should have expressed far more caution when reaching their textualist conclusions based on the narrow subset of founding-era sources they reviewed. Additionally, corpus linguistic theory reveals that there are inconsistencies in both Justice Scalia's and Stevens's descriptions of the Second Amendment's original public meaning.

TABLE OF CONTENTS

INTRODUCTION

Constitutional litigation is inevitably binary: the Supreme Court will select one of two competing readings of the Constitution. *District of Columbia v. Heller*[1] held that the Second Amendment protects an *individual* right to keep and bear arms.[2] The dissent countered that the right was *collective*.[3] In *NLRB v. Noel Canning*,[4] the Court

1. 554 U.S. 570 (2008).
2. *Id.* at 595.
3. *Id.* at 645 (Stevens, J., dissenting).
4. 573 U.S. 513 (2014).

concluded that the Recess Appointments Clause applies to *inter-session* recesses, as well as *intra-session* recesses.[5] The concurrence insisted it only applied to the former.[6] And the Court found in *Zivotofsky ex rel. Zivotofsky v. Kerry*[7] that the Recognition Power belongs *exclusively* to the President.[8] Chief Justice Roberts' dissent maintained that the President and Congress *share* this authority.[9]

Were these cases rightly decided? Or did the dissents have the better arguments? In these cases of first impression, there were no binding precedents.[10] Rather, the Court had to interpret constitutional text from a blank slate. The majority and dissenting opinions considered a range of factors, including text, history, structure, and practice.[11] No single factor was dispositive.

It is difficult to grade the accuracy of such complex decisions in their entirety. But each case heavily relied on textualist arguments. Now, scholars can use corpus linguistics, a method for studying language, to assess the accuracy of those textualist claims.[12] Here, we will analyze the Supreme Court's landmark Second Amendment decision.

This Article proceeds in five parts. Part I revisits *Heller*. The Second Amendment provides: "A well regulated Militia, being necessary to the security of a free State, the right of the people to keep and bear Arms, shall not be infringed."[13] Justice Scalia wrote the majority opinion in *Heller*.[14] He concluded that the Second Amendment "protects an individual right to possess a firearm unconnected with service in a militia, and to use that arm for traditionally lawful purposes, such as self-defense within the home."[15] In short, an *individual* right. Justice Stevens, in his dissent, contended that the Second Amendment "is most naturally read to secure to the people a right to use and possess arms in conjunction with service in a well-regulated militia."[16] That is, a *collective* right.

Justice Scalia and Justice Stevens each made linguistic claims about four elements of the Second Amendment: *right of the people*,

5. *Id.* at 519.
6. *See generally id.* at 575–93 (Scalia, J., concurring).
7. 576 U.S. 1 (2015).
8. *Id.* at 28.
9. *Id.* at 63 (Roberts, C.J., dissenting).
10. *See, e.g.*, District of Columbia v. Heller, 554 U.S. 570, 635 (2008).
11. *See, e.g.*, *id.* at 579–92; *id.* at 644–51 (Stevens, J., dissenting); *id.* at 681–84 (Breyer, J., dissenting).
12. Jesse Bowman, *Corpus Linguistics*, 108 ILL. BAR J. 52, 52 (2020).
13. U.S. CONST. amend. II.
14. 554 U.S. at 572.
15. *Id.* at 577.
16. *Id.* at 651 (Stevens, J., dissenting).

keep and bear arms, *keep arms*, and *bear arms*.[17] Both the majority and the dissent used various textualist approaches to consider these four phrases, but their toolkit in 2008 was limited. They considered only a fairly narrow range of sources to interpret the text. Justice Scalia admitted that his analysis was limited to the "written documents of the founding period that *we have found*."[18] Likewise, Justice Stevens's dissent cited "*dozens* of contemporary texts."[19] Over a decade later, big data allows us to perform much more sophisticated queries.

Part II introduces corpus linguistics and the Corpus of Founding Era American English ("COFEA"). COFEA (which rhymes with Sophia) contains over 100,000 texts of approximately 140 million words ranging in date from the start of the reign of King George III (1760) to the death of George Washington (1799).[20] COFEA allows us to parse the twenty-seven words of the Second Amendment in dynamic ways with information that was simply inaccessible in 2008, when the Court decided *Heller*. We will discuss several tools and practices used in corpus linguistics research: frequency data, concordance lines, coding, collocation, and clusters (or n-grams). Finally, we will suggest four best practices to perform corpus linguistics research. These practices are designed to reduce the risk of coding bias and ensure that others can replicate corpus linguistic queries.[21]

Part III reviews the work of several other scholars who have used corpus linguistics to assess the textualist claims made in *Heller*: Professor Dennis Baron, Professor Alison LaCroix, Josh Jones, and Neal Goldfarb.

Part IV presents our corpus linguistics analysis. We performed five queries. First, we queried *right of the people*. Second, we queried *keep and bear arms* (and synonyms). Third, we queried the word *right* within six words of *arms*. Fourth, we queried the word *keep*, and variants of *keep*, within six words of *arms*. Fifth, we queried the word *bear*, and variants of *bear*, within six words of *arms*. To reduce the risk of coding bias, we used eight coders who independently coded their results using a type of double-blind methodology.

Finally, Part V grades the four linguistic claims made in the *Heller* opinions. Both the majority and the dissenting opinions erred with respect to some of their linguistic claims. Justices Scalia and Stevens should have expressed far more caution when reaching their

17. *See id.* at 579–92 (majority opinion); *id.* at 644–51 (Stevens, J., dissenting).

18. *Id.* at 582 (majority opinion) (emphasis added).

19. *Id.* at 647 (Stevens, J., dissenting) (emphasis added).

20. *See Corpus of Founding Era American English (COFEA)*, BYU L.: L. & CORPUS LINGUISTICS (Oct. 11, 2019), https://lcl.byu.edu/projects/cofea/.

21. Our consolidated dataset is available at https://bit.ly/3y2AcaT.

textualist conclusions based on the narrow subset of founding-era sources they reviewed.

On balance, a corpus linguistics analysis shows that the meaning of the operative clause of the Second Amendment is a much closer call than either the *Heller* majority or the dissent were willing to admit. In fact, we found linguistic evidence that supports both views. However, *Heller* relied on far more than just linguistic evidence, and this Article does not consider other historical methodologies to inform the original public meaning of the Second Amendment. Nor does this Article consider the meaning of the right to keep and bear arms as applied to the states by virtue of the Fourteenth Amendment.[22]

I. *DISTRICT OF COLUMBIA V. HELLER*

In 1976, the District of Columbia criminalized the possession of handguns.[23] Dick Heller, a resident of the District of Columbia, worked as a special policeman.[24] Heller was required to use a firearm for his work at the Thurgood Marshall Judiciary Building, but the District of Columbia denied his application to register a handgun that

22. *Cf.* McDonald v. City of Chicago, 561 U.S. 742, 765 (2010) ("Instead, the Court decisively held that incorporated Bill of Rights protections [through the Due Process Clause] 'are all to be enforced against the States under the Fourteenth Amendment according to the *same standards* that protect those personal rights against federal encroachment.'" (emphasis added) (quoting Malloy v. Hogan, 378 U.S. 1, 10 (1964))). *But see* Josh Blackman, *Originalism at the Right Time?*, 90 TEX. L. REV. *SEE ALSO* 269, 277 (2012) ("Originalism at the right time remains a valid methodology that constrains interpretive inquires to the proper timeframe."); Josh Blackman & Ilya Shapiro, *Keeping Pandora's Box Sealed: Privileges or Immunities, The Constitution in 2020, and Properly Extending the Right to Keep and Bear Arms to the States*, 8 GEO. J.L. & PUB. POL'Y 1, 51 (2010) ("Originalism demands that the interpreter select the proper temporal location in which to seek the text's original public meaning."); Alan Gura, Ilya Shapiro & Josh Blackman, *The Tell-Tale Privileges or Immunities Clause*, 2010 CATO SUP. CT. REV. 163, 196 (arguing that "the correct timeframe for analyzing the Fourteenth Amendment's substantive protections is the Reconstruction era."). Even critics of *Heller* recognize that the Second Amendment had a more individualized meaning during Reconstruction. AKHIL REED AMAR, AMERICA'S CONSTITUTION: A BIOGRAPHY 390 (2005) ("Yet when filtered through the well-chosen language of 'privileges or immunities of citizens,' the Founders' Second Amendment could be refined into a rather different kind of right: a right/privilege to keep a gun at home for self-protection—a right of all citizens, female as well as male, acting individually rather than in a collective militia, wielding weapons in a private space rather than mustering on the public square.").

23. *Heller*, 554 U.S. at 574–75.

24. *Id.* at 575.

he wished to keep at home.[25] Heller then challenged the constitutionality of the District of Columbia handgun ban.[26]

In 2008, the Supreme Court considered whether the District of Columbia's handgun ban violated the Second Amendment, which provides that "[a] well regulated Militia, being necessary to the security of a free State, the right of the people to keep and bear Arms, shall not be infringed."[27] The Court sharply divided, 5-4.[28] Justice Scalia wrote the majority opinion, which was joined by Chief Justice Roberts and Justices Kennedy, Thomas, and Alito.[29] The Court found that the Second Amendment guarantees an individual right to possess a gun in the home for self-defense.[30] Justice Stevens, joined by Justices Souter, Ginsburg, and Breyer, dissented, concluding that the Second Amendment refers to a collective right.[31]

Professor Cass Sunstein observed that *Heller* may be "the most explicitly and self-consciously originalist opinion in the history of the Supreme Court."[32] Both Justice Scalia's majority opinion, as well as Justice Stevens's dissent, relied on a slew of historical arguments.[33] In this Article, we do not engage *all* of the originalist arguments advanced by Justices Scalia and Stevens. Nor does our analysis assess the voluminous body of historical research that the parties advanced in *Heller*.

Rather, we consider four specific linguistic claims that the majority and dissent addressed. First, what is the significance of the phrase *right of the people*? Second, was the phrase *keep and bear arms* a single linguistic unit? Third, what is the meaning of *keep arms*? Fourth, what is the meaning of *bear arms*? We will begin with an analysis of Justice Scalia's textualist methodologies.

A. Justice Scalia's Textualist Methodologies

At the outset of *Heller*, Justice Scalia laid out his interpretive principle: "[T]he Constitution was written to be understood by the voters; its words and phrases were used in their normal and ordinary as distinguished from technical meaning."[34] He explained that the "[n]ormal meaning" of words and phrases would "include an idiomatic

25. *Id.*
26. *Id.* at 575–76.
27. U.S. CONST. amend. II.
28. *Heller*, 554 U.S. at 572.
29. *Id.*
30. *Id.* at 635.
31. *Id.* at 636, 645–46 (Stevens, J., dissenting).
32. Cass R. Sunstein, *Second Amendment Minimalism:* Heller *as* Griswold, 122 HARV. L. REV. 246, 246 (2008).
33. *See, e.g.*, *Heller*, 554 U.S. at 580–82, 640–42 (showing two of the many examples of Justices Scalia and Stevens using historical arguments).
34. *Id.* at 576 (quoting United States v. Sprague, 282 U.S. 716, 731 (1931)).

meaning," or a meaning that is not necessarily deducible from the phrase's individual words but is still commonly understood.[35] But, Justice Scalia continued, the "normal meaning" of words and phrases would "exclude[] secret or technical meanings that would not have been known to ordinary citizens in the founding generation."[36]

Justice Scalia's interpretive principle made two significant errors. First, he conflated two different *publics*. He referred, interchangeably, to "the voters" and "ordinary citizens."[37] These *speech communities* were distinct in the late eighteenth century. For example, "ordinary citizens" would include women, but women could only vote in New Jersey.[38] Also, different states had different restrictions on suffrage—not all men could vote either. Moreover, the category of "citizens" excluded slaves, and freedmen in certain states.[39] And what did it mean to be an "ordinary citizen"? Not all "citizens" would have been deemed "ordinary" when interpreting the Constitution. Some scholars would argue that the views of George Washington, James Madison, and Alexander Hamilton are entitled to more weight than the views of the average voter.[40] The two groups Justice Scalia identified overlap like a Venn diagram, with some people included in both categories, but other people excluded from both categories. Justice Scalia's imprecise standard could potentially lead to different linguistic conclusions, and thus, different original public meanings.

We identify a second significant error with Justice Scalia's interpretive principle: certain terms would not have been readily understood by common folk. The Constitution often uses words or phrases that have technical legal meanings that may have been obscure to "ordinary citizens in the founding generation."[41] Professors John McGinnis and Michael Rappaport identified at least thirteen unambiguously legal terms in the Constitution, such as "duty of tonnage," "attainder of treason," or "writ of habeas corpus."[42]

35. *Id.* at 576–77.

36. *Id.*

37. *Id.*

38. *See* THOMAS G. WEST, VINDICATING THE FOUNDERS: RACE, SEX, CLASS, AND JUSTICE IN THE ORIGINS OF AMERICA 75–77 (1997) (noting that women could vote initially in New Jersey from 1776–1807); *see also* Jennifer Schuessler, *On the Trail of America's First Women to Vote*, N.Y. TIMES (Aug. 7, 2020), https://www.nytimes.com/2020/02/24/arts/first-women-voters-new-jersey.html.

39. *See* WEST, *supra* note 38, at 27–28.

40. *See, e.g.*, John O. McGinnis & Michael B. Rappaport, *The Constitution and the Language of the Law*, 59 WM. & MARY L. REV. 1321, 1360–61 (2018) (stating that the Constitution was drafted in the language of the law by learned elites who understood such language better than the average voter).

41. *Heller*, 554 U.S. at 577; *see also* McGinnis & Rappaport, *supra* note 40, at 1370.

42. *See* McGinnis & Rappaport, *supra* note 40, at 1370–73.

Those trained in the law may have grasped such legalese, but it is unlikely that ordinary citizens in the late eighteenth century would have understood these terms. Professors McGinnis and Rappaport also identified forty-four terms that could have both a technical, as well as an ordinary meaning, such as "corruption of blood."[43]

Modern originalist scholars have grasped this nuance. Generally, they seek the Constitution's original meaning, also known as its "objective social meaning" or "semantic meaning."[44] This process searches for the "meaning" that words and phrases in the Constitution "would have had at the time they were adopted as law, within the [legal] and linguistic community that adopted" that text.[45] If we are seeking the "ordinary" meaning of a text, we would look to how "ordinary citizens in the founding generation" would have understood that text.[46] But for a technical, legal term, it would be a mistake to look to "ordinary citizens." That audience is the wrong *public*. Instead, we would consider the views of people of the day who were learned in the law.[47] Thus, the original public meaning "can typically be discovered by empirical investigation" of an ordinary public or of a technical public—whichever is the appropriate speech community.[48]

If the Second Amendment uses phrases that would have been understood only by a technical audience, then Justice Scalia's focus on the "ordinary citizen" was misplaced.

B. Justice Scalia's Four Linguistic Claims

The Second Amendment has two portions. The first part is referred to as the prologue, the preamble, or the prefatory clause.[49] It provides, "[a] well regulated Militia, being necessary to the security of a free State . . . ."[50] The second part is known as the operative clause.[51] It provides, "the right of the people to keep and bear Arms shall not be infringed."[52]

The structure of the Second Amendment is complicated. Indeed, it may be the most syntactically complex provision of the Bill of

43. *Id.* at 1371, 1373.

44. Randy E. Barnett, *Interpretation and Construction*, 34 HARV. J.L. & PUB. POL'Y 65, 66 (2011).

45. Vasan Kesavan & Michael Stokes Paulsen, *The Interpretive Force of the Constitution's Secret Drafting History*, 91 GEO. L.J. 1113, 1131 (2003).

46. *See Heller*, 554 U.S. at 576–77.

47. *See* McGinnis & Rappaport, *supra* note 40, at 1411.

48. Barnett, *supra* note 44, at 66.

49. *See Heller*, 554 U.S. at 577, 610, 624 (using all three words interchangeably to describe the first part of the Second Amendment).

50. U.S. CONST. amend. II, cl. 1.

51. *Heller*, 554 U.S. at 577.

52. U.S. CONST. amend. II, cl. 2.

Rights. On its face, the prefatory clause appears to be in tension with the operative clause. The former discusses service in the militia.[53] The latter discusses the right of the people to keep and bear arms.[54] Moreover, the grammar of the Second Amendment is complicated. Does it protect a single right to *keep and bear arms*? Or does it protect separate rights to *keep arms* and *bear arms*? And how does the phrase *right of the people* affect these two rights? Finally, what is the relationship between a *well regulated militia* and the *right of the people*? The answers to these questions are not evident from the text.

Justice Scalia made four linguistic claims about the operative clause. First, he wrote that where the phrase *the people* is coupled with a *right*, the Constitution speaks to an individual right.[55] Where the phrase *the people* is not coupled with a right, the Constitution speaks to some sort of collective action.[56] Second, he dissected the phrase *keep and bear Arms* into two components: *keep Arms* and *bear Arms*.[57] Third, he concluded that the phrase "'keep Arms' was simply a common way of referring to possessing arms, for militiamen *and everyone else*."[58] In other words, this right could be collective *or* individual in nature. Fourth, Justice Scalia acknowledged that *bear Arms* had an "idiomatic meaning" that reflected militia service.[59] But, he countered, "'bear Arms' . . . . [*U*]*nequivocally* bore that idiomatic meaning only when followed by the preposition 'against,' which was in turn followed by the target of the hostilities."[60] We will consider each claim in turn.

1. "Right of the People"

Justice Scalia's analysis began with the phrase *right of the people*.[61] He relied on an interpretive methodology that Professor Akhil Reed Amar dubbed "intratextualism."[62] With this approach, courts should "read a contested word or phrase that appears in the Constitution in light of another passage in the Constitution featuring the same (or a very similar) word or phrase."[63] The phrase *right of the people* appears elsewhere in the Bill of Rights. For example, that phrase is used "in the First Amendment's Assembly–and–Petition Clause and in the Fourth Amendment's Search–and–Seizure

53. *See* U.S. CONST. amend. II, cl. 1 (referencing "a well regulated Militia").
54. *See id.* cl. 2 (referencing "the right of the people to keep and bear Arms").
55. *Heller*, 554 U.S. at 579, 581.
56. *Id.* at 579–80.
57. *Id.* at 582–84.
58. *Id.* at 583.
59. *Id.* at 586.
60. *Id.*
61. *Id.* at 579.
62. Akhil Reed Amar, *Intratextualism*, 112 HARV. L. REV. 747, 748–49 (1999).
63. *Id.* at 748.

Clause."[64] Justice Scalia observed that the Ninth Amendment, which refers to "rights . . . retained by the people," also "uses very similar terminology."[65] Justice Scalia found a common meaning among these Amendments: "[a]ll three of these instances unambiguously refer to individual rights, not 'collective' rights, or rights that may be exercised only through participation in some corporate body."[66]

Next, Justice Scalia expanded upon his intratextualist analysis. He observed that other provisions of the Constitution refer to "'the people' in a context other than 'rights.'"[67] For example, the preamble refers to "We the People."[68] Section Two of Article I provides that "the people" can "choose members of the House."[69] And the Tenth Amendment recognizes that "those powers not given the Federal Government remain with 'the States' or 'the people.'"[70] But these references to "the people," Justice Scalia contends, differ from the reference to "the people" in the Second Amendment because these three provisions "deal with the exercise or reservation of powers, not rights."[71] Therefore, the people can only act "collectively."[72]

Justice Scalia concluded, "[n]owhere else in the Constitution does a 'right' attributed to 'the people' refer to anything other than an individual right."[73] In short, where the phrase *the people* is coupled with a *right*, the Constitution speaks to an individual right. Where the phrase *the people* is not coupled with a right, the Constitution speaks to some sort of collective action. Indeed, "in all six other provisions of the Constitution that mention 'the people,' the term unambiguously refers to all members of the political community, not an unspecified subset."[74] Justice Scalia relied on this intratextualist analysis to form the basis of his opinion: he "start[ed] . . . with a strong presumption that the Second Amendment right is exercised individually and belongs to all Americans."[75]

Intratextualism may potentially be a valuable tool in the interpretive toolkit. It makes sense that a word or phrase used one way in a legal document, like the Constitution, will be used in the same sense in other places in the document. However, there are three important limitations in this context. First, assume that the phrase

64. *Heller*, 554 U.S. at 579.
65. *Id.* (quoting U.S. CONST. amend. IX).
66. *Id.*
67. *Id.*
68. *Id.* (quoting U.S. CONST. pmbl.).
69. *Id.* (quoting U.S. CONST. art. I, § 2, cl. 1).
70. *Id.* (quoting U.S. CONST. amend. X).
71. *Id.* at 579–80.
72. *Id.*
73. *Id.* at 580.
74. *Id.*
75. *Id.* at 581.

right of the people in the Second Amendment has the same meaning as the phrase *right of the people* in the Fourth Amendment. That relationship still does not tell us what *right of the people* means. The text is silent about whether a right is *individual* or *collective*. In this way, intratextualism runs into the danger of interpretive circularity.

Second, Justice Scalia presumed that the phrase *right of the people* is a phrase that would have been "known to ordinary citizens in the founding generation."[76] But he did not establish that the phrase *right of the people* lacked a specialized, technical meaning.[77] He did not perform any research to prove this point. Justice Scalia simply made an assumption. This approach could lead to a double error. Assume that the phrase *right of the people* had *both* a legal and an ordinary meaning. Justice Scalia may have concluded that the phrase *right of the people* in the Fourth Amendment has an ordinary meaning.[78] But what if he was wrong? What if *right of the people* in the Fourth Amendment had a specialized, legal meaning? That is potential error #1. Next, Justice Scalia relied on that assumption to perform an intratextualist analysis: he concluded that the phrase *right of the people* in the Second Amendment shares the same ordinary meaning as *right of the people* in the Fourth Amendment.[79] That is potential error #2. In short, intratextualism based on flawed assumptions can compound constitutional errors.

Intratextualism has a third limitation: there may not be enough data points to establish an intratextualist link. The First, Second, and Fourth Amendments use the phrase *right of the people*.[80] The Ninth Amendment uses similar, but slightly different language.[81] Is this small sample size sufficient to establish how an entire population would have understood the phrase in that context? Even if the sample size is sufficient, Justice Scalia's intratextualist analysis of *right of the people* still appears to rely on several unfounded assumptions.

2. "Keep and Bear Arms"

Justice Scalia continued his analysis with the phrase *keep and bear arms*. Justice Scalia dissected the phrase into two components: *keep arms* and *bear arms*.[82] However, it is not self-evident that this phrase can be dissected. The phrase *keep and bear arms* could have been a single linguistic unit, as Justice Stevens suggested.[83] Perhaps it was a specialized, technical phrase that would not have been

76. *Id.* at 576–77.
77. *See id.* at 579–81.
78. *Id.* at 579.
79. *Id.* at 576–79.
80. U.S. CONST. amends. I, II, IV.
81. U.S. CONST. amend. IX.
82. 554 U.S. at 582–84.
83. *Id.* at 651 (Stevens, J., dissenting).

understood by "ordinary citizens." Justice Scalia rejected this proposition without any independent analysis.[84] He simply assumed that *keep and bear arms* refers to separate rights.

3. "Keep Arms"

Third, Justice Scalia considered the phrase *keep arms*. He defined the word *keep*, standing by itself.[85] Then he returned to the full phrase, *keep arms*.[86] We will consider each component separately.

a. "Keep"

According to Samuel Johnson's 1773 edition of *A Dictionary of the English Language*, the word *keep* was defined as: "'[t]o retain; not to lose,' and '[t]o have in custody.'"[87] Both the majority and the dissent relied on Johnson's dictionary, as well as other founding-era dictionaries.[88] Looking to these dictionaries is a common practice in originalist analysis,[89] but there are several significant limitations of using such founding-era dictionaries.[90]

First, dictionaries tend to define individual words, rather than phrases. This practice was common at the founding and endures to the present day. For this reason, Justice Scalia could not locate an entry in any dictionaries for *keep arms*. Instead, he had to focus on *keep*.[91] This practice violates the linguistic principle of non-compositionality, which provides that "the communicative content of a phrase isn't always the sum of its parts."[92]

84. *Id.* at 591 (majority opinion).

85. *Id.* at 582.

86. *Id.*

87. *Id.* (quoting 1 SAMUEL JOHNSON, A DICTIONARY OF THE ENGLISH LANGUAGE 1095 (4th ed. 1773)).

88. *See, e.g.*, *id.* at 581–82, 685 (relying on other founding-era dictionaries).

89. 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, 372–73 (2014).

90. These criticisms have been raised elsewhere. *See* Thomas R. Lee & James C. Phillips, *Data-Driven Originalism*, 167 U. PA. L. REV. 261, 284–89 (2019); James Cleith Phillips & Sara White, *The Meaning of the Three Emoluments Clauses in the U.S. Constitution: A Corpus Linguistic Analysis of American English from 1760–1799*, 59 S. TEX. L. REV. 181, 188–91 (2017). See generally Maggs, *supra* note 89, for additional analysis of the benefits and pitfalls of relying on dictionaries for constitutional interpretation.

91. *Heller*, 554 U.S. at 592.

92. Lee & Phillips, *supra* note 90, at 283. Judge Easterbrook colorfully described this principle:

> [T]he [appellee] produced . . . nothing but a dictionary. It did not offer any evidence about how [people] use or understand the phrase *as a unit*. It offered only lexicographers' definitions of the individual words. That won't cut the mustard, because dictionaries reveal a range of historical

Second, dictionaries from the founding era usually had one or two authors.[93] As a means to save time and effort, these authors would often *lump* definitions together: "[l]umpers are definers who tend to write broad definitions that can cover several more minor variations on that meaning."[94] By contrast, *splitters* would "write discrete definitions for each of those minor variations."[95] Johnson, a lumper, included definitions that were often quite broad.

Third, founding-era dictionaries often plagiarized from each other. Noah Webster's 1828 dictionary copied from Johnson's 1773 dictionary, and Johnson in turn copied from earlier dictionaries.[96] As a result, if an earlier dictionary lumped, and excluded a narrow sense of a word, then that lumping decision would be repeated in later dictionaries.

Finally, "language usage and meaning shifts over time."[97] Linguistics refers to this phenomenon as *linguistic drift*.[98] Such drift can occur quickly in a language, sometimes even in the span of a single decade.[99] However, dictionaries that plagiarize older usage may not discern such drift. Thus, for example, Webster's 1828 dictionary may have relied on definitions from a century earlier—and from a different country.[100] Meaning can drift at different paces in different locations, even if those locations once shared an original meaning. For example, language usage in the thirteen colonies from the 1770s till the 1780s may have evolved at a different pace than

> meanings rather than how people use a particular phrase in contemporary culture. (Similarly, looking up the words "cut" and "mustard" would not reveal the meaning of the phrase we just used.)

TE-TA-MA Truth Found.—Fam. of URI, Inc. v. World Church of the Creator, 297 F.3d 662, 666 (7th Cir. 2002) (emphasis added).

93. *See, e.g.*, THOMAS DYCHE & WILLIAM PARDON, A NEW GENERAL ENGLISH DICTIONARY (18th ed. 1781).

94. KORY STAMPER, WORD BY WORD: THE SECRET LIFE OF DICTIONARIES 119 (2017); *see also* Elizabeth Walter, *Using Corpora to Write Dictionaries*, *in* THE ROUTLEDGE HANDBOOK OF CORPUS LINGUISTICS 428, 434 (Anne O'Keeffe & Michael McCarthy eds., 2010) (discussing "lumpers" and "splitters").

95. STAMPER, *supra* note 94, at 119.

96. *See, e.g.*, ALLEN REDDICK, THE MAKING OF JOHNSON'S DICTIONARY 1746–1773, at 11 (1996); Maggs, *supra* note 89, at 383 ("Samuel Johnson apparently relied on Bailey's definitions when he prepared his dictionary[.]").

97. *See* Lee & Phillips, *supra* note 90, at 265.

98. *Id.*

99. Stefan Th. Gries, *What is Corpus Linguistics?*, 3 LANGUAGE & LINGUISTICS COMPASS 1225, 1233 (2009).

100. *Cf.* Lawrence B. Solum, *Triangulating Public Meaning: Corpus Linguistics, Immersion, and the Constitutional Record*, 2017 BYU L. REV. 1621, 1642 (observing that Samuel Johnson's influential dictionary, which Noah Webster relied on to write his own 1828 dictionary, was first published thirty-two years before the drafting of the Constitution).

usage on the other side of the pond.[101] That usage may have evolved at a far greater pace over the course of a century.[102] Moreover, founding-era dictionary writers often provided examples of language taken from Shakespeare or even the King James Bible—both published centuries before the Constitution was written.[103]

Judges should acknowledge the deficiencies of founding-era dictionaries when performing originalist analyses.

b. "Keep Arms"

After defining the word *keep*, Justice Scalia turned to the broader phrase *keep arms*. He observed that "[t]he phrase 'keep arms' was not prevalent in the written documents of the founding period that we have found."[104] Justice Scalia identified one source that used the phrase *keep arms*,[105] and two sources in which the word *keep* appeared close to the word *arms*.[106] These "few examples[] all . . . favor viewing the right to 'keep Arms' as an individual right unconnected with militia service."[107] In a footnote, Justice Scalia cited ten other sources, at least nine of which appear to be legal documents, that use the phrase *keep arms*.[108] From this evidence, he concluded that "'[k]eep arms' was simply a common way of referring to possessing arms, for militiamen *and everyone else*."[109]

Justice Scalia observed that "[n]o party has apprised us of an idiomatic meaning of 'keep Arms.'"[110] That is, the phrase *keep arms* did not have a specialized meaning and would have been understood by "ordinary citizens." Justice Scalia concluded that "the most natural reading of 'keep Arms' in the Second Amendment" is to be discerned from the dictionary definition of *keep*: "to 'have weapons.'"[111]

Justice Scalia reached these textualist conclusions about the phrase *keep arms* based on a fairly small set of evidence: Samuel Johnson's dictionary and thirteen examples of the phrase, twelve of

101. JACK LYNCH, THE LEXICOGRAPHER'S DILEMMA 131, 141 (2009).

102. *See id.* at 139 (discussing the dramatic differences between modern American English and modern British English).

103. *See, e.g.*, Maggs, *supra* note 89, at 385–86.

104. District of Columbia v. Heller, 554 U.S. 570, 582 (2008).

105. *Id.* at 582 (quoting 4 WILLIAM BLACKSTONE, COMMENTARIES *55) ("Catholics convicted of not attending service in the Church of England . . . were not permitted to 'keep arms in their houses.'").

106. *Id.* (citing 1 W. & M., ch. 15, § 4, in 3 Eng. Stat. at Large 422 (1689) ("[N]o Papist . . . shall or may have or keep in his House . . . any Arms . . . ."); 1 W. HAWKINS, TREATISE ON THE PLEAS OF THE CROWN 26 (1771) (similar)).

107. *Id.*

108. *Id.* at 583 n.7.

109. *Id.* at 583.

110. *Id.* at 582.

111. *Id.*

which were legal sources.[112] Additionally, five of Justice Scalia's sources were from well after 1791, some as late as the 1840s.[113] It is possible that the meaning of this phrase drifted over time.[114] Justice Scalia, therefore, might have relied on a meaning of the phrase that did not yet exist when the Second Amendment was ratified.

Finally, Justice Scalia relied on several sources to analyze the phrase *keep arms* that were English rather than American.[115] It is not self-evident that words had the same meaning in both nations.

In 2008, this limited data set was the best the Supreme Court could muster. But it is problematic to generalize from these few sources, which have inherent limitations, about how almost four million Americans in 1791 would have understood the phrase *keep arms*.[116]

4. "Bear Arms"

Justice Scalia concluded his textualist analysis with the phrase *bear arms*. He began this analysis with a dictionary definition of *bear*.[117] Next, he reviewed how *bear arms* was used in founding-era sources.[118] He then claimed that "bear arms . . . . [*U*]*nequivocally* bore that idiomatic [military] meaning only when followed by the preposition 'against,' which was in turn followed by the target of the hostilities."[119] Finally, Justice Scalia countered that *bear arms*, without *against*, could in certain contexts refer to an individual right.[120]

a. "Bear"

Once again, Justice Scalia focused on a single word, *bear*. And once again, to define that word, Justice Scalia cited Johnson's dictionary, as well as Sheridan's 1796 dictionary.[121] Both sources define *bear* as "carry."[122] In the modern era, Justice Scalia concluded,

112. *Id.* at 582–83, 583 n.7.
113. *Id.* at 583 n.7.
114. Solum, *supra* note 100, at 1639–41 (discussing "the phenomenon of linguistic drift").
115. *Heller*, 554 U.S. at 582, 583 n.7.
116. The population of the United States in 1790 was about four million. *Pop Culture: 1790*, U.S. CENSUS BUREAU, https://www.census.gov/history/www/through_the_decades/fast_facts/1790_fast_facts.html (last visited Sept. 10, 2021).
117. *Heller*, 554 U.S. at 584.
118. *Id.* at 584–86.
119. *Id.* at 586.
120. *Id.* at 588–89, 589 n.11.
121. *Id.* at 584.
122. *Id.*

the word *bear* has the same meaning.[123] Our earlier criticism of founding-era dictionaries applies equally here.[124] Specifically, these dictionaries suffer from the problems of non-compositionality, the phenomenon of lumping, and the possibility of plagiarism. It is also possible that Sheridan plagiarized from Johnson.[125]

b. "Bear Arms" in "Founding-Era Sources"

Next, Justice Scalia sought to interpret the "natural meaning" of *bear arms* through a "review of founding-era sources."[126] Justice Scalia primarily relied on the text of eleven state constitutional provisions from the late eighteenth and early nineteenth centuries.[127] He also looked to antebellum state court decisions interpreting state constitutions.[128] Finally, he reviewed the *Collected Works of James Wilson*.[129] All of these sources, he concluded, "demonstrate . . . that 'bear arms' was not limited to carrying of arms in a militia."[130]

The contemporaneous state constitutional provisions may be the most "analogous linguistic context" to determine the meaning of the Second Amendment.[131] But the other legal sources Justice Scalia cited seem one step removed from the sort of evidence that an "ordinary citizen" would have understood. Legal texts are just as likely to provide evidence of legal meaning as they are to provide Justice Scalia's intended target: "natural meaning."[132] Justice Scalia did not address why the "ordinary citizens" at the time of the founding would have shared the understandings of the specialized sources he consulted. Several of Justice Scalia's later-in-time sources pose another risk: The understanding of the words in the Second Amendment may have shifted after 1791.[133] Unfortunately, Justice Scalia did not acknowledge this risk of *linguistic drift*.

c. "Bear Arms Against" referred to service in the militia

Earlier in his opinion, Justice Scalia concluded that the phrase *keep arms* does not have an idiomatic meaning separate from its natural meaning.[134] However, "at the time of the founding," he wrote, the phrase "'bear Arms" did have "an idiomatic meaning that was

123. *Id.*
124. *See supra* Part I.B.3.a.
125. *See supra* notes 90–100 and accompanying text.
126. *Heller*, 554 U.S. at 584.
127. *Id.* at 584–85, 585 n.8.
128. *Id.* at 585–86, 585 n.9.
129. *Id.* at 585.
130. *Id.* at 585–86.
131. *Id.*
132. *Id.* at 586.
133. Lee & Phillips, *supra* note 90, at 265.
134. *Heller*, 554 U.S. at 582.

significantly different from its natural meaning."[135] That specialized meaning was "'to serve as a soldier, do military service, fight' or 'to wage war.'"[136] Yet, Justice Scalia stressed that the phrase "bear arms," by itself, did not generally embrace this idiomatic meaning.[137] Rather, "bear arms . . . . [*U*]*nequivocally* bore that idiomatic meaning only when followed by the preposition 'against,' which was in turn followed by the target of the hostilities."[138] For example, Justice Scalia cited the Declaration of Independence, which states that the King "has constrained our fellow Citizens taken Captive on the high Seas to *bear Arms against* their Country . . . ."[139]

Justice Scalia found support for his position in an amicus brief filed by three professors of linguistics and English, Dennis E. Baron, Richard W. Bailey, and Jeffrey P. Kaplan.[140] (We will return *infra* to Professor Baron's recent criticism of *Heller*.)[141] Justice Scalia wrote, "[e]very example given by petitioners' *amici* for the idiomatic meaning of 'bear arms' from the founding period either includes the preposition 'against' or is not clearly idiomatic."[142] Let us unpack that claim. The linguistics professors treated the terms *bear arms*, *bearing arms*, and *bear arms against* as synonymous.[143] The linguistics professors identified several sources in which the phrase *bear arms against* takes the idiomatic military meaning.[144] But the linguistics professors also identified several sources in which *bear arms*, by itself, also seems to take the idiomatic military meaning.[145] It is unclear how Justice Scalia thought this latter category was "not clearly idiomatic." That language is an ambiguous hedge. We do not know how much certainty Justice Scalia demanded for the language to be "clearly idiomatic."

135. *Id.* at 586.

136. *Id.* (quoting Brief for Professors of Linguistics and English Dennis E. Baron et al. as Amici Curiae at 18, *Heller*, 554 U.S. 570 (No. 07-290)).

137. *Id.*

138. *Id.*

139. *Id.* (emphasis added) (quoting THE DECLARATION OF INDEPENDENCE para. 28 (U.S. 1776)).

140. *Id.* (citing Brief for Professors, *supra* note 136, at 18–23).

141. *See infra* Part III.A.

142. *Heller*, 554 U.S. at 586 (citing Brief for Professors, *supra* note 136, at 18–23).

143. Brief for Professors, *supra* note 136, at 20–21.

144. *Id.* at 19, 21 (identifying, for example, "He bure armes, and made weir against the King"; "An ample . . . pardon to all who had born arms against him."; "[C]ompelled to *bear arms against* the friends of their country").

145. *Id.* at 22–23 ("In many of the provinces not a man of the nobility able to *bear arms* remains"; "[T]he power of subjecting all men able to *bear arms* to martial law at any moment, should remain vested in congress"; "That no person, conscientiously scrupulous of *bearing arms* in any case, shall be compelled personally to serve as a soldier.").

This distinction between *bear arms* and *bear arms against* was central to Justice Scalia's analysis. He used this distinction to respond to the historical usages collected by the linguistics professors.[146] We are not certain how Justice Scalia drew this distinction, but we will discuss *infra* that this distinction simply does not hold up.[147]

d. "Bear Arms" was not limited to a militia sense

Later in his opinion, Justice Scalia seems to back away from the argument premised on *bear arms against*. He wrote that *bear arms* need not be limited to its military meaning even though "it was often used in that context."[148] Why? He found it "especially unremarkable that the phrase was often used in a military context in the federal legal sources (such as records of congressional debate) that have been the focus of petitioners' inquiry."[149] Further, "[t]hose sources would have had little occasion to use it *except* in discussions about the standing army and the militia."[150] Likewise, Justice Scalia dismissed the linguistics professors' brief that "supposedly show[ed] that the phrase 'bear arms' was most frequently used in the military context."[151] He explained that "the fact that the phrase was commonly used in a particular context does not show that it is limited to that context, and, in any event, we have given many sources where the phrase was used in nonmilitary contexts."[152] Here, Justice Scalia raised an important theoretical question for originalists: how should courts weigh different types of linguistic evidence? We will address this question in Part II.B.

Finally, Justice Scalia observed that "the phrases used primarily in those military discussions include not only 'bear arms' but also 'carry arms,' 'possess arms,' and 'have arms'—though no one thinks that those *other* phrases also had special military meanings."[153] Justice Scalia added that "legal sources frequently used 'bear arms' in nonmilitary contexts," and "if one looks beyond legal sources, 'bear arms' was frequently used in nonmilitary contexts."[154] However, he did not explain why it makes sense to give equal weight to legal and nonlegal sources. Perhaps legal sources are the best type of evidence for interpreting the Constitution. But if the Court is trying to determine how "ordinary citizens" would understand the operative

146. *Heller*, 554 U.S. at 586.
147. *See infra* Part V.D.
148. *Heller*, 554 U.S. at 587.
149. *Id.*
150. *Id.*
151. *Id.* at 588.
152. *Id.*
153. *Id.* at 587.
154. *Id.* at 587–88.

phrases, the nonlegal sources arguably could be more important than the legal ones.

C. Justice Stevens's Four Linguistic Claims

Justice Stevens wrote the principal dissent in *Heller*.[155] (Justice Breyer also wrote a dissent, but he did not provide any textualist analysis.)[156] Justice Stevens analyzed the same four aspects from the operative clause as Justice Scalia did: *right of the people*, *keep and bear arms*, *bear arms*, and *keep arms*.

Justice Stevens's methodology is more opaque than that of the majority opinion because he failed to state many of his assumptions. For example, Justice Stevens did not explain what public, or speech community, he was trying to generalize about. By contrast, Justice Scalia was quite transparent about his approach. As a result, we are left to speculate about why Justice Stevens draws certain conclusions. In this Part, we attempt to draw all inferences in the light most favorable to Justice Stevens's position.

1. "Right of the People"

First, Justice Stevens began with the phrase *the right of the people*. Like Justice Scalia, Justice Stevens also relied on intratextualism with respect to the First and Fourth Amendments. Justice Stevens wrote that the First and Fourth Amendments "contemplate collective action."[157] He added that *the people* in the Second Amendment "refer[s] back to the object announced in the Amendment's preamble."[158] Justice Stevens concluded that the phrase *right of the people* in the Second Amendment could only refer to a *collective* right.[159] He recognized that the rights in the First, Second, and Fourth Amendments can be exercised by individuals.[160] But, Justice Stevens reasoned, the phrase *right of the people*, and the rights themselves, referred to collective action.[161] The right to peaceable assembly, he explained, is "concern[ed] . . . with action engaged in by members of a group, rather than any single individual."[162] And the right to petition the government "is primarily collective in nature."[163] Justice Stevens reasoned that for this right "to be effective, petitions must involve groups of individuals acting in

155. *Id.* at 636 (Stevens, J., dissenting).
156. *Id.* at 681 (Breyer, J., dissenting).
157. *Id.* at 645 (Stevens, J., dissenting).
158. *Id.* at 645.
159. *Id.* at 645–46.
160. *Id.* at 644–46.
161. *Id.* at 645.
162. *Id.*
163. *Id.*

concert."[164] Thus, the dissent concluded that even rights with individual components can be ultimately collective.

2. "Keep and Bear Arms"

Second, Justice Stevens turned to the phrase *to keep and bear arms*. However, Justice Stevens did not dissect the phrase into two separate rights—as Justice Scalia did—at least, not right away. Justice Stevens seems to reject the notion that the Second Amendment refers to a right to *keep arms* and a separate right to *bear arms*. To the contrary, he wrote, these five words "describe a unitary right."[165] Justice Stevens did not perform any linguistic analysis to support this claim. Rather, his assertion appears to have been an unsupported assumption.

Justice Stevens concluded that the Second Amendment "protects only one right, rather than two."[166] He explained that "the single right that it does describe is both a duty and a right to have arms available and ready for military service, and to use them for military purposes when necessary."[167] He asserted that the "unitary right" serves a very specific purpose: "to possess arms if needed for military purposes and to use them in conjunction with military activities."[168]

Justice Stevens's analysis does not neatly fit the binary frame of an "individual" right or a "collective" right. He seems to contend that *keeping* arms was an "individual" right but that *bearing* arms was a "collective" one. Perhaps an individual could keep a gun at home but only use it when fighting with the militia. Alternatively, perhaps guns could be kept in a government storage facility, where they could be retrieved for militia use. More than a decade later, we are still not entirely sure how Justice Stevens conceived of the *right to keep and bear arms*. In any event, this dual nature of the right undermines Justice Stevens's argument that the Second Amendment "protects only one right, rather than two."[169]

3. "Bear Arms"

Even though Justice Stevens claimed that *keep and bear arms* is a "unitary right," he still linguistically analyzed the two component rights.[170] He might have done so to offer a belt-and-suspenders response to Justice Scalia. Or perhaps he felt compelled to break up the terms because the phrase was rarely used in founding-era

164. *Id.*
165. *Id.* at 646.
166. *Id.* at 651.
167. *Id.*; *see also id.* at 651 n.13 (engaging in intratextualism by contrasting the Second Amendment with the First Amendment).
168. *Id.* at 646.
169. *Id.* at 651.
170. *See id.* at 646–52.

documents. Another possibility is that Justice Stevens was not fully committed to the "unitary right" concept. We are not certain.

Justice Stevens began with the phrase *bear arms*.[171] (In contrast, Justice Scalia began with *keep arms*.[172]) This phrase, Justice Stevens wrote, is "a familiar idiom" having a military meaning.[173] He relied on several sources, including contemporaneous state constitutions, the linguistics professors' brief, and an eighteenth-century dictionary.[174]

Justice Stevens also looked to the Latin etymology of the phrase, "*arma ferre*, which, translated literally, means 'to bear [*ferre*] war equipment [*arma*].'"[175] However, relying on the etymology of the phrase is potentially problematic.[176] The Latin roots are millennia old, and the meaning of words can shift over time. For example, the word *December* derives from the Latin root for "tenth month,"[177] but December is the twelfth month on the modern Gregorian calendar.[178] Justice Stevens did not provide any evidence to suggest that the meaning from the Latin roots has remained static.

4. "Keep Arms"

Justice Stevens concluded his textualist framework with a brief, one-paragraph analysis of *keep arms*. He wrote that "[t]he Amendment's use of the term 'keep' in no way contradicts the military meaning conveyed by the phrase 'bear arms' and the Amendment's preamble."[179] He observed that "a number of state militia laws in effect at the time of the Second Amendment's drafting used the term 'keep' to describe the requirement that militia members store their arms at their homes, ready to be used for service when necessary."[180] Justice Stevens did not perform any further linguistic analysis beyond examining a single Virginia Militia Act from 1785.[181]

171. *Id.* at 646.
172. *See id.* at 582 (majority opinion).
173. *Id.* at 646 (Stevens, J., dissenting).
174. *Id.* at 646–47.
175. *Id.* at 646 (quoting Brief for Professors, *supra* note 136, at 19).
176. *See* Lee & Phillips, *supra* note 90, at 265 & n.115.
177. *Id.* at 288.
178. *See* Steve Hendrix, *On New Year's, Our Calendar's Crazy History, and the Switch That Changed Washington's Birthday*, WASH. POST (Dec. 31, 2017), https://www.washingtonpost.com/news/retropolis/wp/2017/12/31/on-new-years-our-calendars-crazy-history-and-the-switch-that-changed-washingtons-birthday/.
179. *Heller*, 554 U.S. at 650 (Stevens, J., dissenting).
180. *Id.*; *see also id.* at 650 n.12 (quoting such laws from New Jersey, Delaware, and Connecticut).
181. *Id.* at 650–51.

***

Like Justice Scalia, Justice Stevens mingled legal and nonlegal sources in his analysis.[182] He did not attempt to defend, or even seem to recognize, his confused methodology.

II. THE POTENTIAL AND LIMITS OF CORPUS LINGUISTICS

Corpus linguistics is a method to empirically study how language is used.[183] Through this methodology, researchers can search, or *query*, databases of texts that include naturally occurring language usage. These databases, or bodies of words, are known as a corpus (singular) or corpora (plural).[184] This Part will introduce COFEA and identify some constraints on using this corpus. Next, this Part discusses four tools to use with corpus linguistics research: frequency data, concordance lines, collocation, and clusters (or n-grams). Finally, we suggest four best practices to perform corpus linguistics research. These practices are designed to reduce the risk of bias in corpus analysis and ensure that others can replicate corpus linguistic queries.

A. Corpus Linguistics and COFEA

Attorneys are already familiar with databases that can be considered rudimentary type of corpora, such as Westlaw and Google.[185] These databases index newspaper articles, books, cases, and other sources. But these services are not true corpora in the linguistic sense. They are not structured to represent a particular speech community and do not contain linguistic tools for analysis. *General* linguistic corpora focus on a broad speech community, such as an entire nation.[186] In contrast, *special* linguistic corpora focus on a specific speech community, such as a particular dialect or region of a country.[187] While the number of words in a corpus is important, a large corpus that does not match the relevant speech community is not very helpful. Instead, the corpus must represent the group about which one wants to draw inferences.

Today, COFEA is the largest and most reliable corpus available for American English in the late eighteenth century.[188] COFEA

182. *See id.* at 650 (discussing Virginia's state militia law); *id.* at 646–47 (invoking etymologies and nonlegal dictionaries in his analysis).

183. James C. Phillips et al., *Corpus Linguistics and Officers of the United States*, 42 HARV. J.L. & PUB. POL'Y 871, 877 (2019).

184. *Id.*

185. *See id.* at 880 (discussing concordance line (a major corpus linguistics tool) use in modern legal searches).

186. Gries, *supra* note 99, at 1232.

187. *See id.* (discussing differences between general and specific corpora).

188. *Corpus of Founding Era American English*, *supra* note 20.

contains over 100,000 texts with approximately 140 million words from the start of the reign of King George III (1760) to the death of George Washington (1799).[189] COFEA is a *historical* corpus, which means that it contains language only from this time period.[190] COFEA is not a *monitor* corpus, which is updated over time.[191]

COFEA is also the most prominent corpus for American constitutional interpretation.[192] It enables scholars and jurists to investigate how "ordinary citizens" would have understood the Bill of Rights in 1791.[193] The federal courts have begun to rely on COFEA. For example, Justice Thomas cited COFEA in his *Carpenter v. United States*[194] dissent.[195]

COFEA consists of three smaller corpora. First, the Evans Early American Imprints ("Evans") collection consists of "nearly two-thirds of all books, pamphlets, and broadsides known to have been printed in this country" during the time period covered.[196] We believe that the Evans sub-corpus is most relevant to determining what an "ordinary citizen" would have understood in the 1790s.

The second sub-corpus contained in COFEA is the Founders Online database, which is published by the National Archives. This sub-corpus consists of the "correspondence and other writings of seven major shapers of the United States: George Washington, Benjamin Franklin, John Adams (and family), Thomas Jefferson, Alexander Hamilton, John Jay, and James Madison."[197] This database includes correspondence written by these luminaries, as well as letters written by other founders and by "ordinary citizens."[198] We do not think Washington and his peers would be deemed "ordinary citizens." Rather, this sub-corpus may be of more value to determine how the Constitution's Framers understood and used language. Yet,

189. *Id.*

190. *Id.*

191. TONY MCENERY & ANDREW HARDIE, CORPUS LINGUISTICS: METHOD, THEORY AND PRACTICE 11 (2012) (contrasting a "monitor corpus" with a "snapshot corpus" and showing that a monitor corpus is not isolated to only one moment in time).

192. *Corpus of Founding Era American English*, *supra* note 20.

193. *See id.* (noting that the corpus contains a range of publications from the relevant timeframe, which include nonlegal sources like books, pamphlets, and periodicals).

194. 138 S. Ct. 2206 (2018).

195. *Id.* at 2238 n.4 (Thomas, J., dissenting).

196. *Evans Early American Imprints (Evans) TCP*, TEXT CREATION P'SHIP, https://textcreationpartnership.org/tcp-texts/evans-tcp-evans-early-american-imprints/ (last visited July 17, 2021).

197. *Founders Online*, NAT'L ARCHIVES, https://founders.archives.gov/ (last visited July 17, 2021).

198. *See id.* (showcasing other authors and recipients who are frequently searched but not listed as the "seven major shapers").

not all documents in this corpus are of equal weight: letters written by "ordinary citizens" to these founders may not be as probative.

The third sub-corpus in COFEA is the Hein database. This sub-corpus includes legal materials such as cases, statutes, legislative debates and related materials, legal papers, and the like.[199] We think this database is most helpful to determine how those trained in the law would have understood legal terms. These three corpora allow us to study how language was used in three speech communities.

TABLE 1. THE DATABASES THAT CONTRIBUTE TO COFEA[200]

Sub-Corpus	**Word Count**	**# of Files**	**Document Types**	**Authors**
Evans	53.4 million	2,881	Books, pamphlets, broadsides, speeches, sermons, etc.	More "ordinary" folks with some founders
Founders Online	43.9 million	115,281	Letters, diaries, other writings, etc.	Mostly founders with some "ordinary" folks
Hein/Legal	48.6 million	351	Statutes, cases, legislative records, legal treatises, etc.	Legal bodies, elected officials and judges, others

We acknowledge there are several constraints on COFEA. As a threshold matter, not all documents from the late eighteenth century have been preserved in databases.[201] Rather, the documents that were preserved and digitized are skewed toward favored segments of society.[202] This demographic was more likely to (1) get something published or (2) have their writings preserved for posterity. As a result, almost all of the sub-corpus documents, especially the Founders Online corpus, were written by white males.[203] And even among white men, the records of elite figures are far more likely to

199. Phillips et al., *supra* note 183, at 884.

200. We performed queries in an earlier version of COFEA in December 2018. As a result, these numbers may differ somewhat from the totals on the most recent version of COFEA.

201. *Corpus of Founding Era American English*, *supra* note 20.

202. *See id.*

203. *See id.*

have been preserved.[204] For instance, the papers of the Federal Farmer were saved,[205] but the papers of most real farmers were not.

This disparity creates two important difficulties for COFEA. First, if COFEA is limited to what elite, white males believed, then the database cannot fully represent the broader category of "ordinary citizens." Second, COFEA cannot tell us to what degree the language usage and understanding of mostly elite white males differed from other groups of Americans. For example, did non-elites, women, African Americans, American Indians, and other groups have different understandings of speech? What about Americans who spoke English as a second language?[206] Additionally, how would the literacy rate affect this query? According to one study, during the time of the framing, white males in New England had a literacy rate of about ninety percent, while women had a literacy rate of about forty-eight percent.[207] If these varied groups shared their understanding of language, then COFEA's skewing may not be problematic. But if there was a greater disparity, then the skewing does not give us a complete picture of language in the late eighteenth century among "ordinary citizens." Of course, this selection effect is not limited to corpus linguistics. This difficulty is present in all studies of historical documents. In future writings, we will address this challenge for originalism.

We identify another constraint on COFEA. As broad as the database is, it lacks representativeness for some types of documents. These different types of documents are called *genres* or *registers*.[208] People use different types of words in different modes of communication. People tend to speak differently than they write. For example, a newspaper article will differ from an academic journal. Indeed, COFEA does not include a dedicated database of founding-era newspapers.[209] The significance of this omission is uncertain. However, this gap may not be as critical as it appears at first glance. During the founding era, newspapers tended to consist of an

204. *See id.*

205. The Federal Farmer was a prominent anti-Federalist. *See* Paul Finkelman, *Complete Anti-Federalist*, 70 CORNELL L. REV. 182, 194 (1984) (discussing the Federal Farmer's views on the need for the Constitution).

206. *See generally* Christina Mulligan et al., *Founding-Era Translations of the U.S. Constitution*, 31 CONST. COMM. 1 (2016) (analyzing translations of the Constitution made in 1787 for the German- and Dutch-speaking populations of Pennsylvania and New York).

207. *See* KENNETH A. LOCKRIDGE, LITERACY IN COLONIAL NEW ENGLAND: AN ENQUIRY INTO THE SOCIAL CONTEXT OF LITERACY IN THE EARLY MODERN WEST 39 (1974) (extrapolating literacy rate from rate of signatures used).

208. *See* Douglas Biber, *What Can a Corpus Tell Us About Registers and Genres?*, *in* THE ROUTLEDGE HANDBOOK OF CORPUS LINGUISTICS 241, 241–42 (Anne O'Keeffe & Michael McCarthy eds., 2010).

209. *See Corpus of Founding Era American English*, *supra* note 20.

assortment of sermons, speeches, letters, essays, and the like.[210] Newspapers from the late eighteenth century did not focus on current events articles written in a distinctive style. As a result, adding newspapers to COFEA may not add a distinctive genre of speech. Finally, COFEA does not include state ratification debates.[211] These documents could provide some insight into how the ratifiers used and understood language.

Despite those shortcomings, COFEA is the only corpus of American English for this time period. (The Corpus of Historical American English starts its coverage three decades after ratification in 1820.[212]) COFEA also presents marked improvement over existing sources and tools. Further, with only a few exceptions, the documents in COFEA were not created with the intent to influence any specific interpretation of the Constitution. Likewise, the Evans collection, Founders Online, and the Hein database were not built with any particular constitutional interpretive question in mind. Thus, COFEA is, for all intents and purposes, agnostic as to the question researchers seek to answer. Still, even with COFEA, there are best practices researchers should follow.

B. Corpus Linguistics Tools

We rely on several tools to perform corpus linguistics research. The first and most basic tool is known as *frequency data*.[213] This method asks how often a word is used over time or in different types of documents.[214] Here, we are looking for a simple count.

A second tool is called *concordance lines*.[215] These lines resemble the search results from a Westlaw or Google query, with a little more information. We can illustrate this concept with an example. Assume you search a corpus for the phrase *bear arms*. It is helpful to know the words that appear before and after *bear arms* in a given document. A concordance line will display the search term (*bear arms*), as well as the words that appear before and after it. A single

210. *See* Robert McNamara, *History of Newspapers in America*, THOUGHTCO. https://www.thoughtco.com/history-of-newspapers-in-america-4097503 (last updated Feb. 24, 2020) (noting that newspapers were often highly political and that they contained essays and letters attempting to drive political action or express political opinion).

211. *See Corpus of Founding Era American English*, *supra* note 20 (failing to list state ratification debates on the COFEA database homepage).

212. *Corpus of Historical American English*, ENGLISH-CORPORA, https://www.english-corpora.org/coha/ (last visited July 17, 2021).

213. Jane Evison, *What are the Basics of Analysing a Corpus?*, *in* THE ROUTLEDGE HANDBOOK OF CORPUS LINGUISTICS 122, 123–26 (Anne O'Keeffe & Michael McCarthy eds., 2010).

214. *See id.* (discussing how to use frequency data in linguistic analysis).

215. *Id.* at 122, 128–30.

row of output is known as a concordance line. A corpus can generate fifty, one hundred, or more randomized concordance lines.[216] The size depends on the total number of search results and the parameters of the search. Moreover, a concordance line need not be limited to a single sentence. Most corpora can display an entire paragraph of material surrounding a particular concordance line. With such expansive searches, concordance lines can provide the sort of context that dictionaries cannot. Concordance lines can also be used to generate frequency data.[217]

Once the concordance lines are generated, researchers can classify, or code, the search results.[218] Corpus linguistics draws on the best practices and principles of content analysis and survey methodologies.[219] Consider the same example we used earlier. You perform a search for *bear arms*. Then, you classify each concordance line according to a particular sense of the phrase *bear arms*. Perhaps some documents are used in the sense of an individual right to bear arms. Other documents are used in the sense of a collective, militia-related right to bear arms, and in some documents, the reference to *bear arms* does not clearly fit into either category. Coding is a qualitative endeavor that requires reading a large amount of material. For example, the coders who performed the analysis in this Article read about 127,000 words. That total exceeds the words in any of the first three *Harry Potter* novels.[220]

In our view, classifying concordance lines may be the most useful corpus linguistic tool for interpreting the Constitution. This approach has the most potential to decide which sense of a word or phrase is more common. Further, this approach avoids cherry-picking results, as the corpus can generate a large number of concordance lines. This tool presents an obvious difficulty: deciding how to code specific lines.

There are two types of coding: *manifest coding* and *latent coding*.[221] Manifest coding is used where the categorization is apparent on the surface.[222] For example, a researcher counting how

216. *See id.* at 129–30 (discussing the process of using concordance lines).

217. *See id.* at 131 (providing an example of the intersection of frequency data and concordance lines through the analysis of the word "now").

218. *See* EARL BABBIE, THE PRACTICE OF SOCIAL RESEARCH 400 (12th ed. 2010).

219. *See generally* James C. Phillips & Jesse Egbert, *Advancing Law and Corpus Linguistics: Importing Principles and Practices from Survey and Content-Analysis Methodologies to Improve Corpus Design and Analysis*, 2017 BYU L. REV. 1589 (2017) (explaining how law and corpus linguistics can import principles and practices from content-analysis and survey methodologies from other fields, such as media studies, social sciences, and natural sciences).

220. *See How Many Words Are There in the* Harry Potter *Book Series?*, WORDCOUNTER: BLOG (Nov. 23, 2015), https://wordcounter.net/blog/2015/11/23/10922_how-many-words-harry-potter.html.

221. *See* BABBIE, *supra* note 218, at 338.

222. *See id.*

many times a particular term appears in search results is manifest coding. This task is purely objective and involves no subjectivity. However, researchers can also perform latent coding.[223] This task requires subjective judgment about the text's underlying meaning. This Article relies on latent coding. Properly classifying latent data is a qualitative endeavor that requires subjective judgment.[224] We will address this challenge in Part II.C.

We also rely on another tool known as *collocation*.[225] Some word pairs appear together more frequently than other word pairs. We refer to such *word neighbors* as collocates (pronounced KAH-la-kits). These semantic patterns of word association are often intuitive. For example, we would expect the word "dark" to appear fairly often in the same semantic environment as the word "night." But "dark" is less likely to appear near "perfume."

In some cases, these patterns can be surprising. For example, the adjective "impending" is often collocated with bad events, such as "doom," "death," "disaster," "crisis," "attack," and "danger."[226] However, one of the top collocates of "impending" in modern American English is also "marriage"[227]—something we may not normally place in the same category as the prior negative events. Thus, collocation can uncover patterns that our intuition alone would miss. The courts have long recognized the principle of collocation with the *noscitur a sociis* canon of construction: "it is known by its associates."[228] Linguists state the rule in a slightly different fashion: "You shall know a word by the company it keeps!"[229]

Another tool looks at *lexical bundles*, which are a repeated series or groupings of three or more words.[230] Searching for lexical bundles in a corpus is usually done with the N-grams or clusters feature.[231]

223. *Id.*

224. *See id.*

225. *See* TONY MCENERY & ANDREW WILSON, CORPUS LINGUISTICS: AN INTRODUCTION 85–88 (2d ed. 2001).

226. *Corpus of Historical American English*, *supra* note 212 (analyzing the results of searching for collocates of "impending" as an adjective).

227. *Id.*

228. *Noscitur a sociis*, BLACK'S LAW DICTIONARY (11th ed. 2019).

229. John Rupert Firth, *A Synopsis of Linguistic Theory, 1930–1955*, *in* STUDIES IN LINGUISTIC ANALYSIS 1, 11 (1962).

230. DOUGLAS BIBER ET AL., LONGMAN STUDENT GRAMMAR OF SPOKEN AND WRITTEN ENGLISH 444 (2002) ("defin[ing] a lexical bundle as a recurring sequence of three or four words.").

231. *See generally* Omer Ari, *Review of Three Software Programs Designed to Identify Lexical Bundles*, 10 LANGUAGE LEARNING & TECH. 30 (2006) (evaluating different software programs that help find lexical bundles using both N-grams and clusters).

Such units of words "are not complete phrases."[232] For example, two of the most common lexical bundles in conversational English are "Do you want me to" and "I don't know what."[233] Now, consider a legal example. COFEA contains forty-nine occurrences of a lexical bundle that includes "agriculture," "commerce," and a third word.[234] In thirty-one of those bundles, the third word was "manufactures."[235] In the other eleven bundles, no word appeared more than four times.[236] These results suggest that the relationship between "agriculture," "commerce," and "manufactures" is stronger than the relationship between "agriculture," "commerce," and other words. Lexical bundles can also suggest that certain words, placed in a specific order, have taken on an idiomatic, or specialized, meaning.[237] For example, where the words "nice" and "fat" appear together, they appear in that order, not as "fat" and "nice."[238] The latter is a *hendiadys* (pronounced hen-DIE-ah-dus), which we will discuss in Part IV.B.4.

C. Four Best Practices for Corpus Linguistics Research

We acknowledge that there is always a risk of bias when coding or classifying data. Here, we do not use the word "bias" in a pejorative fashion. Rather, everyone is affected by motivated reasoning and confirmation bias.[239] These problems are not specific to corpus linguistics but are part of human nature. Scholars in all disciplines must contend with these issues. Fortunately, corpus linguistics researchers can adopt certain measures to reduce the concerns of coding bias.

First, researchers should ask others to independently code the data. This approach helps reduce the risk of coding bias.[240] For our

232. GENA R. BENNETT, USING CORPORA IN THE LANGUAGE LEARNING CLASSROOM: CORPUS LINGUISTICS FOR TEACHERS 9 (2010).

233. BIBER ET AL., *supra* note 230, at 443.

234. Lee & Phillips, *supra* note 90, at 308 tbl.5.

235. *Id.*

236. *Id.*

237. While lexical bundles can be idiomatic, they need not be. *See* Susan M. Conrad & Douglas Biber, *The Frequency and Use of Lexical Bundles in Conversation and Academic Prose* 20 LEXICOGRAPHICA, 56, 69 (2004).

238. Samuel L. Bray, *"Necessary AND Proper" and "Cruel AND Unusual": Hendiadys in the Constitution*, 102 VA. L. REV. 687, 689 (2016) (using the example of a "nice and fat" cow to show that this is a single descriptor, not two individual qualities of the cow being lauded).

239. *See* Ziva Kunda, *The Case for Motivated Reasoning*, 108 PSYCH. BULL. 480, 480 (1990).

240. *See* JEROME KAGAN, THE THREE CULTURES: NATURAL SCIENCES, SOCIAL SCIENCES, AND THE HUMANITIES IN THE 21ST CENTURY 41 (2009) ("Social scientists are more often concerned with the meanings of verbal statements or actions. . . . [S]ocial scientists rely on consensual agreement among trained experts as a way to protect against the biased perspective of a single observer.").

research, we asked law students to code the data. We did not code the data ourselves. And each researcher coded the data independently. When each coder was blind to what other coders did, we eliminated intercoder bias that could distort the results. That is, Coder #1 did not see the results of Coder #2, and Coder #2 did not see the results of Coder #1.

Second, the coders should be given as little information as possible about the project so as not to create any preconceptions. For our Article, we only gave the students the specific categories, or senses, we were researching. We did not provide them with any pre-conception of what they might find or what we expected the results to be. That way, we reduced any risk that our views would shade their work, and vice versa.[241] Medical trials follow this double-blind methodology of experiments: both the person receiving a treatment and the person dispensing a treatment do not know which pill is the drug and which is the placebo.[242] This double layer of independence—between each coder and between coders and researchers—approximates the double-blind technique that is the gold standard of experimental methodology.

Third, researchers should use multiple coders for each query. We used *at least* two coders for each query. Our goal was to obtain intercoder reliability. Here, we aimed for an agreement of at least 70% between the coders. When the percentage of agreement was lower, we included additional coders. There are much more sophisticated statistical measures of intercoder reliability beyond the percentage of agreement.[243] However, for this research, we were satisfied that our coders generally exceeded the 70% threshold. We did not deem it necessary to impose more stringent standards.

Fourth, researchers should disclose their research. To that end, we have published all of our data sets and have tried to be as transparent as possible. This way, other researchers can scrutinize and try to recreate our work.

We think these four steps can reduce the risk of coding bias and thus represent best practices for corpus linguistics research. The steps are also fairly straightforward. They can be employed in academic institutions, law firms, and, we hope, judicial chambers.[244]

241. *See* Phillips & Egbert, *supra* note 219, at 1614.

242. *See* BABBIE, *supra* note 218, at 234–35.

243. *See* Phillips & Egbert, *supra* note 219, at 1616.

244. *See* Josh Blackman, *Originalism and Stare Decisis in the Lower Courts*, 13 NYU J.L. & LIBERTY 44, 60–62 (2019) (discussing how courts can perform corpus linguistics research).

III. OTHER CORPUS LINGUISTIC ANALYSES OF *HELLER*

Recently, several scholars have used COFEA to assess the linguistic claims made in *Heller*. First, Professor Dennis Baron criticized Justice Scalia's interpretation of "bear arms."[245] (Professor Baron was the main author for the linguistics professors' amicus brief that Justices Scalia and Stevens both cited in *Heller*.)[246] Second, Professors Alison LaCroix and Jason Merchant developed a project titled "Historical Semantics and Legal Interpretation."[247] In a blog post, Professor LaCroix wrote that the Second Amendment's "prefatory clause's reference to a 'well regulated Militia' [is] more meaningful than the Court suggested in *Heller*."[248] Third, Josh Jones criticized both Justice Scalia's majority opinion and Justice Stevens's dissent.[249] Fourth, Neal Goldfarb has written numerous briefs and blog posts that are critical of *Heller*.[250]

We agree on several points with these four scholars. But we also have some points of disagreement. At the outset, we offer a general criticism. These scholars did not follow the best practices we identified above, such as employing double-blind coding. Instead, as far as we can tell, the scholars performed all the coding themselves. Therefore, they did not take the necessary steps to reduce coding bias. We will assume for present purposes that their coding is accurate. Neal Goldfarb, to his credit, did publicize most of his data sets.[251]

245. *See* Dennis Baron, *Corpus Evidence Illuminates the Meaning of* Bear Arms, 46 HASTINGS CONST. L.Q. 509, 510 (2019).

246. District of Columbia v. Heller, 554 U.S. 570, 577, 586, 588–89 (2008); *id.* at 646–47, 647 n.9 (Stevens, J., dissenting).

247. *See* Alison L. LaCroix, *Historical Semantics and the Meaning of the Second Amendment*, PANORAMA: EXPANSIVE VIEWS FROM *J. EARLY REPUBLIC* (Aug. 3, 2018), http://thepanorama.shear.org/2018/08/03/historical-semantics-and-the-meaning-of-the-second-amendment/.

248. *Id.* (quoting U.S. CONST. amend. I).

249. *See* Josh Jones, Note, *The "Weaponization" of Corpus Linguistics: Testing* Heller*'s Linguistic Claims*, 34 BYU J. PUB. L. 135, 135 (2020) ("This Note introduces new corpus linguistics research that suggests both Justice Antonin Scalia's majority opinion and Justice Stevens's dissenting opinion relied on inaccurate historical linguistic claims.").

250. *See infra* Part III.D.

251. *See* Neal Goldfarb, *A (Mostly Corpus-Based) Linguistic Reexamination of* D.C v. Heller *and the Second Amendment* 2, https://ssrn.com/abstract=3481474 (last updated Feb. 27, 2021) [hereinafter Goldfarb SSRN Collection]; *see also* Neal Goldfarb, *Comments on Two Responses to My (Mostly Corpus-Based) Analysis of the Second Amendment. Part 1: Gun-Rights Advocates' Amicus Brief*, LAWNLINGUISTICS (June 25, 2020), https://lawnlinguistics.com/2020/06/25/comments-on-two-responses-to-my-mostly-corpus-based-analysis-of-the-second-amendment-part-1-gun-rights-advocates-amicus-brief/ [hereinafter Goldfarb Comments]; Neal Goldfarb, *Corpora and the Second Amendment*,

This transparency allowed people to more carefully scrutinize his work. Our disagreement with these scholars, therefore, is primarily methodological. We also disagree with the queries they chose and how much weight they placed on their results.

In this Part, we will analyze each of these scholars' corpus linguistics analyses.

A. Professor Dennis Baron's Corpus Linguistics Analysis of Heller

Professor Baron focused on part of the Second Amendment's operative clause: *bear arms*. He found 310 instances of the phrase *bear arms* in COFEA.[252] He also searched the Corpus of Early Modern English ("COEME"). COEME (which is pronounced koh-EEM) includes documents in English from 1475–1800; most of the documents are British English, though some are American English.[253] Professor Baron located 1,578 instances of *bear arms* in COEME.[254] He "was able to examine about 1,300 of these instances in context."[255] Professor Baron examined 1,000 instances from COEME and, based on our estimates, probably all 310 from COFEA.[256] But we cannot be sure of the exact counts because Professor Baron did not provide the precise contours of his research.

Professor Baron then eliminated approximately 400 of these 1,300 instances that he found to be duplicates, leaving "roughly 900 separate occurrences of *bear arms* before and during the founding era."[257] He wrote that these instances overwhelmingly "refer to war, soldiering, or other forms of armed action by a group rather than an individual."[258] Professor Baron found that only seven instances of the phrase *bear arms* "were either ambiguous or carried no military connotation."[259] He provided some examples of the military sense of *bear arms* that he characterized as "typical."[260] He also reproduced all seven instances that he deemed to be ambiguous or have a non-military sense.[261]

LAWNLINGUISTICS (Aug. 8, 2018), https://lawnlinguistics.com/corpora-and-the-second-amendment/ [hereinafter Goldfarb Corpora].

252. Baron, *supra* note 245, at 510.

253. *BYU-Corpus of Early Modern English: About the Corpus*, BYU L. & CORPUS LINGUISTICS, https://lawcorpus.byu.edu/byucoeme/concordances;showCorpusDescription=true/search (last visited July 17, 2021).

254. Baron, *supra* note 245, at 510.

255. *Id.*

256. *Id.* at 511 n.5 ("Since COEME only returns a maximum of 1,000 hits for a collocation search, I was not able to examine 578 of the 1,578 citations with *bear arms*.").

257. *Id.*

258. *Id.*

259. *Id.* at 510–11.

260. *Id.* at 511.

261. *Id.* at 512–13.

Next, Professor Baron searched for the phrase *keep arms*. He found twenty-eight instances in COEME and ten instances in COFEA.[262] He eliminated thirteen instances that were duplicative or irrelevant.[263] Professor Baron was left with "twenty-five of the remaining occurrences [that] refer to weapons for use in the military or the militia, and one [that is] ambiguous."[264]

Professor Baron's methodology has several shortcomings. First, as a threshold matter, he did not comply with any of the best practices discussed above.[265] Professor Baron did not publish his data sets. We attempted to duplicate his searches. We found the same number of instances of *keep arms* in COFEA (10) but fewer instances of *bear arms* (297 as compared to his 310). This disparity may be due to the fact that we used version 4.0 of COFEA to check his work. This newer version, which was released on February 2, 2019, may have cleaned up some of the duplicates.

With COEME, we found the same number of search results for *keep arms* (28) but fewer results for *bear arms* (1,452 compared to Professor Baron's 1,578). Ultimately, we were able to largely verify Professor Baron's raw counts. However, we cannot be certain how he coded individual entries. We do not know how he decided that a result was "either ambiguous or carried no military connotation."[266] Moreover, as best as we can tell, he coded all of his own research. He did not appear to use double-blind coding, that is, multiple coders who worked independently of each other and were largely detached from the research's purpose.[267] In short, Professor Baron did not take any steps to mitigate coding bias. It is therefore very difficult to verify his categorization standards.

Second, Professor Baron only focused on the phrases *bear arms* and *keep arms*.[268] He did not research other elements of the Second Amendment, such as *right of the people*. He also did not consider that the phrase *arms* may have different senses when it appears in the context of *rights*. His research, by itself, only speaks to a portion of the Second Amendment in an incomplete way.

Third, the queries that Professor Baron did perform were underinclusive. He only searched for the exact phrases: *bear arms*

262. *Id.* at 513.
263. *Id.*
264. *Id.*
265. *See supra* Part II.C.
266. Baron, *supra* note 245, at 510–11.
267. Professor Baron's article does not mention coders or his coding process. *See generally id.* Professor Baron also states that he personally coded the results. *See id.* at 511 n.5 ("*I* was not able to examine 578 of the 1,578 citations . . . . *I* hope to be able [to] retrieve and analyze the missing citations in the near future[.]" (emphasis added)).
268. *Id.* at 509–18.

and *keep arms*.[269] He did not search for various forms of the operative verbs and differing grammatical constructions. For example, he did not search for phrases like *bearing arms, kept arms*, the *arms were kept*, and so on. Professor Baron did not explain why he limited his analysis to the exact phrases in the Second Amendment.

Professor Baron's underinclusive searches limited the size of his data set. We also found 310 instances of *bear arms* in COFEA, but, when we used broader searches of various forms of that phrase, we found 718 examples.[270] Likewise, Professor Baron only found 10 instances of *keep arms* in COFEA.[271] But our broader search parameters, which included various forms of the phrase, returned 237 instances.[272] Our broader search parameters generated irrelevant results, but Professor Baron's underinclusive searches missed many relevant results. This disparity may explain why his findings on *keep arms* differed significantly from ours.

Fourth, Professor Baron mixed the results from COFEA and COEME.[273] Yet these two corpora arguably represent different speech communities: COFEA includes American English from 1760–1799,[274] and COEME includes mostly British English (with some American English) from 1475–1800.[275] Professor Baron did not contend that *bear arms* and *keep arms* were British terms of art that Americans imported into their lexicon.[276] Without this necessary background, it is not self-evident that the British data is relevant. Indeed, there is a cost to mixing these two data sets: there may be differences between how this language developed on either side of the pond. However, there does not seem to be much of a difference between the results of the two databases when researching the precise phrases of *keep arms* and *bear arms*.

Ultimately, Professor Baron focused on a fairly narrow set of inquiries. Without more, his article cannot prove or disprove *Heller*'s linguistic analysis.

269. *Id.* at 510, 513.

270. *See infra* Part IV.E.

271. Baron, *supra* note 245, at 513.

272. *See infra* Part IV.D.

273. *See* Baron, *supra* note 245, at 510.

274. *Corpus of Founding Era American English: About the Corpus*, BYU L. & CORPUS LINGUISTICS, https://lawcorpus.byu.edu/cofea/concordances/search (last visited July 17, 2021).

275. *BYU-Corpus of Early Modern English: About the Corpus*, *supra* note 253.

276. *See generally* Baron, *supra* note 245.

B. Professors Alison LaCroix and Jason Merchant's Corpus Linguistics Analysis of Heller

In 2018, Professor Alison LaCroix wrote a blog post about her linguistic analysis of phrases from the Second Amendment.[277] Professor LaCroix and her colleague, Professor Jason Merchant, did not use COFEA. Instead, they relied on the Google Books Corpus.[278] They "search[ed] a range of published materials dating from the period between 1760 and 1795 for the phrase 'bear arms.'"[279] Their query generated 181 results.[280] The scholars then categorized the results according to three senses in which *bear arms* was used: "collective, individual, or undeterminable."[281] They also classified "the type of subject that accompanied the phrase (plural, singular, or undeterminable)."[282] They also coded for "heraldic uses."[283] The authors did not define the phrase "heraldic uses." We suspect they were referring to heraldry, the design of coats of arms and armorial bearings.

The scholars found that 64.09% of the 181 instances of *bear arms* used "a collective sense with a plural subject."[284] For example, "Slaves were not permitted to bear arms."[285] The scholars found that 18.23% of the instances used "an individual sense with a singular subject."[286] For example, "When I was strong, and able to bear arms."[287] The remaining results were smaller: 3.31% of the instances had a collective sense with a singular subject, 5.52% of the instances were undeterminable but plural, 4.42% being "undeterminable altogether," and 4.42% were heraldic.[288] They further observed that "[t]he results for newspaper are even more dramatic," but they do not report anything related to such research.[289]

We are hesitant to place too much weight on a blog post. Professors LaCroix and Merchant's complete research may be more rigorous. Still, we have some serious concerns with their research as presented. First, the authors do not appear to follow any of the best practices for corpus linguistics research. They wrote that they

277. *See* LaCroix, *supra* note 247.
278. *Id.*
279. *Id.*
280. *Id.*
281. *Id.*
282. *Id.*
283. *Id.*
284. *Id.*
285. *Id.*
286. *Id.*
287. *Id.*
288. *Id.*
289. *Id.*

"classified by hand" the results, but we are not told anything more.[290] We assume they coded the results on their own. Still, we do not know whether Professors LaCroix and Merchant independently coded the data, they coded together, or just one of them coded. There is thus a huge risk of coding bias, as the scholars took no apparent steps to mitigate those concerns.[291] Moreover, so far, Professors LaCroix and Merchant have not published *any* of their data sets.

Second, the scholars only considered Google Books.[292] This database is limited to just one register—books.[293] They did not review COFEA,[294] which includes multiple genres of texts, such as books, pamphlets, broadsides, speeches, sermons, and letters.[295] Frequency often varies across registers.[296] The scholars may not gain a complete understanding of the language used at the time by focusing on a single register that includes only books.

Third, we have doubts about Professors LaCroix and Merchant's coding categories. Consider an example that would fall outside of Google Books but that is in COFEA. During the Virginia Ratifying Debate, Alexander White published an essay. He referred to "the rights of bearing arms for defence, or for killing game."[297] This entry can be read as referring to people hunting together. Therefore, under Professors LaCroix and Merchant's categorization, this entry would seem to be a collective use of a plural subject. But that categorization is not very helpful to understanding whether the Second Amendment protects a right to bear arms in the militia or a right to have a firearm for personal use. We are not sure what their findings actually reveal.

Fourth, the Second Amendment cannot be reduced to the phrase *bear arms*. Professors LaCroix and Merchant, however, appear to make just that reduction. They did not analyze any other words or phrases from the amendment. The blog post concludes, "our research demonstrates that the language of the Second Amendment points toward a more collective interpretation of the right of gun ownership

290. *Id.*

291. *See supra* Part II.C.

292. LaCroix, *supra* note 247.

293. *See About Google Books*, GOOGLE, https://books.google.com/intl/en/googlebooks/about/index.html (last visited July 17, 2021).

294. *See generally* LaCroix, *supra* note 247 (referencing only the Google Books corpus and failing to mention COFEA).

295. *See supra* Table 1.

296. *See* Phillips & Egbert, *supra* note 219, at 1601 (examining the term *discharge* in the Corpus of Contemporary American English and noting that "no two registers are similar in their frequency" of containing the term and that the register with the highest frequency contained the term about eight times more than the register with the lowest frequency).

297. *Ratification of the Constitution by the States: Virginia*, *reprinted in* 8 THE DOCUMENTARY HISTORY OF THE RATIFICATION OF THE CONSTITUTION 404 (John P. Kaminski et al. eds., 1988).

than Justice Scalia's opinion acknowledges."[298] Respectfully, Professors LaCroix and Merchant's research does little to address the full scope of *Heller*'s linguistic analysis.

C. Josh Jones's Corpus Linguistics Analysis of Heller

Josh Jones published a student note on corpus linguistics and the Second Amendment.[299] First, he tested Justice Stevens's claim that *keep and bear arms* was a unitary right.[300] He searched COFEA for "any variant of *keep* (e.g., *kept*, *keeping*) appear[ing] within six words of any variant of *bear* (e.g., *bore*, *bearing*)."[301] These queries resulted in 105 hits.[302] He deemed eighty of them irrelevant, and nineteen results were duplicates of the Second Amendment.[303] Jones was then left with only six results.[304] Only one of those results was consistent with Justice Stevens's unitary right.[305] This research comports with our own.[306] We commend Jones for testing the *Heller* dissent. The other scholars focused solely on Justice Scalia's majority opinion.

Second, Jones searched COFEA "for all the instances in which any variant of *bear* was found within six words of *arm* or *arms*."[307] He found 727 results.[308] He then selected a random sample of 329 concordance lines.[309] Jones eliminated 105 as irrelevant or duplicative, leaving 224 lines to analyze.[310] He coded these lines for either a "'literal' carrying sense" or the "'figurative' specialized sense."[311] By *literal*, Jones meant *literally* carrying a gun, and by *specialized*, he meant the *idiomatic version* in which one fights in the militia.[312]

Jones found that the specialized sense occurred 65.6% of the time, the literal sense occurred 21% of the time, and the remaining 13.4% of the results were too ambiguous to classify.[313] Jones concluded that "this preliminary corpus data raises doubts about whether one can dismiss the *Heller* Court's adoption of the literal reading of *bear arms*

298. LaCroix, *supra* note 247.
299. Jones, *supra* note 249, at 135.
300. *Id.* at 160.
301. *Id.* at 159.
302. *Id.*
303. *Id.*
304. *Id.* at 160.
305. *Id.*
306. *See* discussion *infra* Part V.B.
307. Jones, *supra* note 249, at 161.
308. *Id.*
309. *Id.*
310. *Id.*
311. *Id.*
312. *Id.* at 136.
313. *Id.* at 161.

(even if it ultimately only makes up 21% of this Note's sample)."[314] In other words, Jones questioned whether *Heller*'s linguistics claims can be dismissed even though only 21% of the results matched the figurative (or idiomatic) meaning of "bear arms."

Third, Jones considered whether the phrase *bear arms*, or a grammatical variant, took on a specialized sense.[315] He considered 147 instances.[316] Only thirty-six results (24.5%) used the preposition *against*.[317] Jones concluded that Justice Scalia was wrong to claim that *bear arms* only took on its military idiomatic meaning when followed by that preposition.[318] We agree with Jones's conclusion.[319]

Our criticism of Jones's work is limited. He was a law student who prepared a note. He did not have adequate resources to perform double-blind coding. And he did not investigate other elements of the Second Amendment, such as *right of the people* and *keep arms*. He acknowledged this limitation.[320] Yet, Jones's work is still more comprehensive than that of Professor Baron and Professors LaCroix and Merchant. And in many aspects, Jones's research resembles our own.

D. Neal Goldfarb's Corpus Linguistics Analysis of Heller

Neal Goldfarb has written extensively on corpus linguistics and the Second Amendment.[321] Goldfarb has also adapted his scholarship into advocacy.[322] In a Supreme Court amicus brief, for instance, Goldfarb argued that *Heller*'s interpretation of the prefatory clause has become "untenable."[323] The quantity of his output is truly impressive. It is impossible to respond to all of Goldfarb's work in a single article. Indeed, we find large areas of agreement between his work and ours. Here, we will focus on arguments presented in Goldfarb's series of blog posts, which he collected into a single posting

314. *Id.* at 164.
315. *See id.* at 161.
316. *Id.*
317. *Id.* at 165.
318. *Id.*
319. *See infra* Part V.D.
320. Jones, *supra* note 249, at 164–65 ("[I]deally additional corpus linguistics research and other tools of constitutional interpretation will be used to shed further light on the original understanding of the *right to keep and bear arms*.").
321. *See generally* Neal Goldfarb, Archive of Blog Posts, LAWNLINGUISTICS: CORPORA AND THE SECOND AMENDMENT, https://lawnlinguistics.com/corpora-and-the-second-amendment/ (last visited July 17, 2021).
322. *See generally* Neal Goldfarb, Archive of Briefs, LAWNLINGUISTICS: BRIEFS, https://lawnlinguistics.com/briefs/ (last visited July 17, 2021).
323. Brief of Neal Goldfarb as Amicus Curiae in Support of Respondents at 25–26, N. Y. State Rifle & Pistol Ass'n v. City of New York, 140 S. Ct. 1525 (2020) (No. 18-280).

on SSRN.[324] This document, we think, provides the most complete statement of his work.[325]

1. "Keep"

Goldfarb began his analysis with the word *keep*. He did so to showcase how a corpus works rather than to determine the meaning of the Second Amendment.[326] He observed, "I'm not sure at this point that I have anything worthwhile to say about [*keep arms*]."[327]

2. "Bear"

Next, Goldfarb turned to the other critical verb in the operative clause, *bear*.[328] He searched for this word in both COFEA and COEME. And, to ensure that he was looking at instances from the same time period in both databases, he limited his COEME search to only retrieve results from 1760–1799.[329] Goldfarb sought to determine whether *bear* was synonymous with *carry*.[330]

It was common enough for an individual to *carry* a weapon for personal use.[331] If *bear* and *carry* had the same or similar meaning, then we can draw an inference: a person could also *bear* a weapon for personal use. Goldfarb's focus on *carry* approached the contentious debate over *bear* from a different angle. We think this framework was clever and original.

To test this hypothesis, Goldfarb "tr[ied] to learn which nouns acted most frequently as [the] direct object" of *carry* and *bear*.[332] Goldfarb performed a query "that would return a list of the nouns that occurred within two words to the right of" *carry* and *bear*.[333] In other words, he searched for whether the nouns that are likely to appear near *bear* are also likely to appear near *carry*. Goldfarb found that different types of nouns are likely to appear near *carry* and *bear*.[334]

324. *See* Goldfarb SSRN Collection, *supra* note 251, at 1.

325. *See id.*; *see also* Goldfarb Comments, *supra* note 251; Goldfarb Corpora, *supra* note 251.

326. *See* Goldfarb SSRN Collection, *supra* note 251, at 17–18.

327. *Id.* at 18.

328. *Id.*

329. *Id.*

330. *Id.* at 18–19.

331. *See id.* at 3 ("[*C*]*arry* was often used to denote the physical carrying of tangible objects."); *see also id.* at 18 ("With respect to *carry*, many of the nouns in the search results . . . denoted tangible objects. These included *arms*, *gun(s)* . . . .").

332. *Id.* at 18.

333. *Id.*

334. *See id.* at 18–19 (describing how Justice Scalia's statement in *Heller* that "[a]t the time of the founding, as now, to 'bear' meant to 'carry'" was an oversimplification of patterns of use for these verbs (quoting District of Columbia v. Heller, 554 U.S. 570, 584 (2008))).

For *carry*, "the most frequent" adjacent nouns were "tangible objects."[335] However, for *bear*, Goldfarb found that nouns representing tangible objects "were much less in evidence."[336] His research did identify some tangible objects that were adjacent to *bear*, for example, *arms*, *torches*, *flags*, *prize*, *cross*, and *sword*.[337] And, as far as we can tell, Goldfarb observed that *arms* was the most frequent object found in his searches for both verbs.[338]

Goldfarb made several other observations. First, *carry* was used in contexts in which "various categories of human beings . . . were described as being carried from one place to another."[339] But *bear* was not used in this context. You could *carry* a person from place to place, but you would not *bear* a person from place to place. Second, the results for *bear* included some nouns, such as *fruit*, *wheat*, *wood*, and *corn*, that denoted a "sense of 'bringing forth.'"[340] Third, Goldfarb also found that *carry* was often associated with nouns such as *home*, *trade*, and *places*. These "nouns did not act as the direct object of carry, but . . . nevertheless had to do with the carrying of tangible objects . . . ."[341] Goldfarb did not make a similar observation with respect to *bear*.[342] Fourth, Goldfarb found uses for *carry* "in which the literal, physical sense of the word was metaphorically extended to include the carrying of information . . . ."[343] For example, one could *carry messages* or *carry news*. But he did not find analogous uses of *bear messages* or *bear news*.[344] Fifth, Goldfarb found some "significant overlap between the results" for *carry* and *bear* in one context: when the document referenced *carrying/bearing burdens* or *weight*.[345]

Goldfarb concluded that, during the founding era, "*bear* was in general not synonymous with *carry*."[346] He acknowledged that the word *bear* was "sometimes used to denote the kind of physical

335. *Id.* at 18.
336. *Id.* at 19.
337. *Id.*
338. *See id.* at 18–19 ("With respect to *carry*, many of the nouns in the search results . . . included *arms*, *guns*, *provisions* . . . . For *bear* . . . the most frequent one by far was *arms* . . . . ").
339. *Id.* at 19.
340. *Id.*
341. *Id.* at 20.
342. *See id.* at 19–20 ("The biggest qualitative difference [between *carry* and *bear*] was seen in the data for burden(s), with the literal use being 100 times as frequent for *carry burden(s)* as for *bear burden(s).* The literal-to-metaphoric ratio for *carry* was roughly 10:1, while for *bear* it was the inverse, 1:10.").
343. *Id.* at 20.
344. *Id.*
345. *Id.*
346. *Id.* at 23.

carrying" of a weapon that Justice Scalia referred to in *Heller*.[347] But, Goldfarb explained, "those uses were infrequent and were exceptions to the general pattern of usage."[348] Thus, Goldfarb rejected Justice Scalia's conclusion that "at the time of the founding, as now, to 'bear' meant to 'carry.'"[349] Goldfarb wrote that Justice Scalia's definition did not "accurately reflect how *bear* was ordinarily used."[350]

Ultimately, this analysis of *bear* and *carry* fails to shed much light on the correctness of *Heller*. Goldfarb conceded that this analysis "doesn't resolve the question of how *bear arms* as used in the Second Amendment was likely to have been understood by the American public of 1789."[351] Goldfarb found that "*arms* was one of the most common nouns that acted as the direct object of *bear*."[352] He acknowledged that the phrase *bear arms* in the Second Amendment, "is potentially the biggest exception to the general rule that *bear* didn't mean carry."[353] We agree. Ultimately, Goldfarb's extensive analysis of *carry* and *bear* does not tell us much about the meaning of the Second Amendment.

3. "Bear Arms"

Next, Goldfarb considered *arms*.[354] His analysis, though thorough, is beyond the scope of this Article, as the nature of the Second Amendment right does not turn on what constitutes an "arm." Justice Scalia and Justice Stevens did not really dispute this point. Rather, the word *arms* could be relevant to determining what types of weapons are protected by the Second Amendment, such as a musket, rifle, handgun, AR-15, or bazooka.

By contrast, Goldfarb's analysis of *bear arms* is quite relevant to this Article. Once again, Goldfarb queried COFEA and COEME, with the latter corpus limited to results from 1760–1799.[355] Goldfarb searched for the verb *bear* and its variants within four words of the noun *arms*.[356] Specifically, Goldfarb searched for words that share the same *lemma* of *bear*, such as *bears*, *bearing*, etc. (A lemma is the canonical or dictionary form of the word, sometimes referred to as a

347. *Id.*
348. *Id.*
349. *Id.* (quoting District of Columbia v. Heller, 554 U.S. 570, 584 (2008)).
350. *Id.*
351. *Id.*
352. *Id.*
353. *Id.*
354. *See id.* at 23–27 (describing the Supreme Court's discussion in *Heller* of what *arms* means, tracing its meaning according to the Oxford English Dictionary over the history of English through the end of the eighteenth century, and examining the relevant corpus data).
355. *Id.* at 34.
356. *Id.*

headword.[357]) Goldfarb used more sophisticated searches than Professors Baron and LaCroix employed. The latter scholars simply searched for the word *bear* without searching for its variants.[358]

Goldfarb combined the results from both corpora and removed duplicates, which generated 756 concordance lines.[359] Goldfarb then removed 221 results for one of two reasons. First, he excluded a concordance line "when the topic was the right to bear arms in the United States or in one of the [country's] states and the phrase occurred in the text (or proposed text) of a constitutional provision protecting that right."[360] He excluded these instances of *bear arms* because he deemed that "there was nothing to be learned from considering uses of that very phrase or of closely related variants, in a similar context."[361] Second, he excluded instances of *bear arms* that were linguistically irrelevant.[362] This latter set of exclusions was understandable, and we performed a culling similar to Goldfarb's latter set of exclusions.[363]

After this filtering, Goldfarb was left with 535 remaining results.[364] Based on these concordance lines, Goldfarb found eleven lines where the phrase *bear arms*, or a grammatical variant, "was unambiguously used to convey what would generally be thought of as its literal sense: 'carry weapons.'"[365] In another fifteen lines, "*bear arms* was not unambiguously used in its literal sense, but . . . [Goldfarb] couldn't point to a specific factor ruling out the possibility that they would have been understood to express the 'natural meaning' that was declared by *Heller*."[366] In other words, between 2.1%–4.9% of the data supported Justice Scalia's reading. As for the other 95% or so, Goldfarb stated that he thought "that all of those uses would most likely have been understood as conveying the idiomatic sense relating to the military: 'serve in the militia,' 'fight in a war,' and so on."[367] These findings are similar to our own.[368]

357. *See* MCENERY & HARDIE, *supra* note 191, at 245.
358. *See* Baron, *supra* note 245, at 510–11; LaCroix, *supra* note 247.
359. Goldfarb SSRN Collection, *supra* note 251, at 34–35.
360. *Id.* at 35.
361. *Id.*
362. *Id.*
363. *See infra* Part IV.E.
364. Goldfarb SSRN Collection, *supra* note 251, at 35.
365. *Id.*
366. *Id.* at 36.
367. *Id.* at 39.
368. *See infra* Part IV.E.

4. "Keep and Bear Arms"

Fourth, Goldfarb searched COFEA and COEME for the phrase *keep and bear arms*.[369] However, these queries did not return any useful concordance lines.[370] We reached a similar result.[371] This term was seldom used outside the context of the Second Amendment. However, this research undermines Justice Stevens's claim that *keep and bear arms* was a single linguistic unit. Goldfarb heavily criticized Justice Scalia's majority opinion, but he did not address this shortcoming of Justice Stevens's dissent.[372]

5. "The Right of the People"

Fifth, Goldfarb performed queries to better understand the phrase *the right of the people*.[373] He searched in COFEA and COEME "for instances of *militia* appearing within six words on either side of *the people*."[374] It is a bit unclear how many total concordance lines Goldfarb found, but he reported nine uses.[375] He conceded this sample size cannot "establish that in contexts having to do with the militia [that] *the people* was ordinarily equated with *the militia*."[376] We agree with this conclusion. But Goldfarb drew an inference from these results: the phrase *the people* in the Second Amendment "can reasonably be interpreted as referring to the subset of Americans who were eligible to serve in the militia."[377]

6. An Analysis of Goldfarb's Methodological Choices

We applaud Goldfarb's rigorous research, but we still have concerns with some of his methodological choices. First, Goldfarb did not employ double-blind coding.[378] It seems he performed all of the coding himself. But, to his credit, Goldfarb published his data sets. That transparency allows the public to carefully review his work.

Second, Goldfarb chose to combine the data from COEME and COFEA.[379] He relied on an unstated assumption: that British English should be given the same weight as American English to

369. Goldfarb SSRN Collection, *supra* note 251, at 60.
370. *Id.*
371. *See infra* Part IV.B.1.
372. Goldfarb SSRN Collection, *supra* note 251, at 3, 7, 9.
373. *Id.* at 56–59.
374. *Id.* at 57.
375. *Id.* at 59.
376. *Id.*
377. *Id.*
378. Goldfarb states that he personally coded the results. *See id.* at 40 ("*I* downloaded [the data] to a spreadsheet, to which *I* added a column that could be used for coding each line. In the image below, you can see the column that *I* used for coding . . . ." (emphasis added)).
379. *Id.* at 18.

interpret language used in the Constitution of the United States. Granted, there may be little difference between the two corpora, but there also could be divergence that would affect the research. By lumping the results from both corpora together, it is very difficult to measure that divergence.

Third, Goldfarb reported the number of results he found after deleting duplicates,[380] but he did not state how many total results he found in each corpus or overall. These omissions make it difficult to recreate all of his research.

Fourth, Goldfarb deleted what arguably was the most relevant data from his search of *bear arms*: discussions of that term in the context of constitutional rights.[381] We acknowledge that including portions of that data may be unhelpful. It is circular to analyze the text of the Second Amendment to discern what the Second Amendment means. But we still think this exclusion was overbroad. Analyzing other legal texts that use *bear arms* in relation to rights seems crucial to understanding the Second Amendment. We cannot be certain how excluding this material affected Goldfarb's results. We also do not know how much weight to give these materials. Goldfarb could have published his analysis with and without those instances of discussions of *bear arms* in a constitutional context so that the reader could compare the outcomes.

Fifth, Goldfarb searched for the lemmas of *keep* and *bear*, but he did not search for synonyms of those words.[382] He established that *bear* was not synonymous with *carry*, but there were other words that were closer to *bear* that could have been queried. For example, we queried *use*.[383] Goldfarb also did not consider what synonyms of *keep* exist. We queried *have*, *own*, and *possess*.[384] Broadening the searches could have generated a larger sample size.

Sixth, Goldfarb did not analyze the whole phrase, *the right of the people*.[385] He did not explain this choice. This element of the operative clause could alter the meaning of the Second Amendment in important ways that Goldfarb simply did not consider.

These shortcomings reduce our confidence in Goldfarb's analysis and findings. In any event, however, his underlying findings for *bear arms* were quantitatively similar to our own. Ultimately, Goldfarb undermined some aspects of Justice Scalia's linguistic analysis. He also undermined a central aspect of Justice Stevens's linguistic analysis. Nevertheless, Goldfarb's analysis still does not cast doubt on all four linguistic claims from *Heller*.

380. *Id.* at 20.
381. *Id.* at 35.
382. *Id.* at 11–12, 34.
383. *See infra* Part IV.B.3.
384. *See infra* Part IV.B.3.
385. *See* Goldfarb SSRN Collection, *supra* note 251, at 6.

IV. OUR CORPUS LINGUISTICS ANALYSIS OF *HELLER*

For many constitutional provisions, it is fairly straightforward to generate corpus linguistics queries. For example, the Interstate Commerce Clause gives Congress the power to "regulate Commerce . . . among the several States."[386] This provision has three important elements: "regulate," "Commerce," and "among."[387] The Establishment Clause provides, "Congress shall make no law respecting an establishment of religion."[388] This provision has two relevant elements: "respecting" and "establishment of religion."[389] ("No law" would seem to mean "no law," but the Supreme Court has held otherwise.[390]) The Takings Clause states that "nor shall private property be taken for public use, without just compensation."[391] This provision has four elements: "private property," "taken," "public use," and "just compensation." These provisions of the Bill of Rights have a straightforward syntax. A single verb acts upon a single subject, with simple modifying words.

The Second Amendment does not fit this simple mold. It provides: "A well regulated Militia, being necessary to the security of a free State, the right of the people to keep and bear Arms, shall not be infringed."[392] There are four verbs: *being necessary*, *keep*, *bear*, and *infringed*. There are four subjects: *well regulated Militia*, *security of a free State*, *right of the people*, and *Arms*. The grammar is very complex. No other provision of the Bill of Rights includes a similar prologue. The text by itself does not tell us the relationship between the prefatory clause and the operative clause. Moreover, the Second Amendment is the only provision in the Constitution that uses the verb *infringed*. By contrast, the First Amendment includes the phrase "Congress shall make no law,"[393] and the Fourth Amendment instructs that the protection against unreasonable searches and

386. U.S. CONST. art. I, § 8, cl. 3.

387. *See* Lee & Phillips, *supra* note 90, at 276, 300 (discussing the limitations of the word "commerce").

388. U.S. CONST. amend. I.

389. *See* Stephanie H. Barclay et al., *Original Meaning and the Establishment Clause: A Corpus Linguistics Analysis*, 61 ARIZ. L. REV. 505, 520–21, 529, 531 (2019).

390. *Cf.* N.Y. Times Co. v. United States, 403 U.S. 713, 720 (1971) (Douglas, J., concurring) ("It should be noted at the outset that the First Amendment provides that 'Congress shall make no law . . . abridging the freedom of speech, or of the press.' That leaves, in my view, no room for governmental restraint on the press.").

391. U.S. CONST. amend. V, cl. 4.

392. U.S. CONST. amend. II.

393. U.S. CONST. amend. I.

seizures "shall not be violated."[394] These phrases may or may not be synonymous.

The Framers gave us a very complicated text. The Second Amendment may be the most syntactically intricate element in the entire Bill of Rights. Fittingly, its subject matter concerns one of the most controversial issues in our society: guns. In comparison with other provisions of the Constitution, the Second Amendment provides greater challenges for a corpus linguistics analysis.

To address the four claims made about the operative clause, we performed five searches. First, we queried *right of the people*. Second, we queried *keep and bear arms* (and phrases using synonyms). Third, we queried the word *right* within six words of *arms*. Fourth, we queried the word *keep*, and variants of *keep*, within six words of *arms*. Fifth, we queried the word *bear*, and variants of *bear*, within six words of *arms*.

We did not conduct any corpus linguistic searches concerning the Second Amendment's prefatory clause. For purposes of our analysis, we assumed that the prefatory clause refers to collective militia service. That conclusion is not self-evident, because at the time of the framing, the "militia" in the prefatory clause could be viewed as the same group as the "people" in the operative clause.[395] However, that element was not relevant to our framework. We also did not try to use corpus linguistics to determine the relationship between the prefatory and operative clauses. Eighteenth-century grammar rules were far from standardized, and we did not see how COFEA would readily reveal insights into that topic.[396] We also did not perform any collocate analyses. There is much duplication within COFEA; the same documents will often repeat multiple times. These duplicates can skew the collocate results: A word may seem like a frequent collocate of another word in founding-era American English. In reality, that word may be a collocate in a single document that is found numerous times in COFEA.

Our analysis began with *right of the people*.

A. "Right of the People"

We queried COFEA for *right of the people*. That search generated 194 results. We then removed duplicates, irrelevant results, and

394. U.S. CONST. amend. IV.

395. District of Columbia v. Heller, 554 U.S. 570, 580 (2008) ("[T]he 'militia' in colonial America consisted of a subset of 'the people'—those who were male, able bodied, and within a certain age range.").

396. *See* Brief of Second Amendment Foundation as Amicus Curiae Supporting Respondent at 7–12, District of Columbia v. Heller, 554 U.S. 570 (2008) (No. 07-290) (analyzing the grammatical structure of the Second Amendment).

quotes of the Second Amendment. We were left with ninety-seven results. (Our cull removed exactly 50% of the raw total).

We asked our coders to categorize each result based on five possible senses. First, we asked if the sense of *right of the people* was "collective." Here are two examples of results that our coders placed in this first category:

> "[T]he canons of the primitive church fully recognized this *right of the people*, to choose their own bishop; and declared, that without such election, they should not be considered as bishops."[397]

> "[T]hat whenever any Form of Government becomes destructive of these Ends, it is the *Right of the People* to alter or to abolish it, and to institute new Government . . . ."[398]

Second, we asked if the sense of *right of the people* was "individual." Here is an example that our coders placed in the second category:

> "As to the *right of the people* to think, let him who denies it, deny, at the same time, their right to breathe."[399]

Third, we asked if the sense of *right of the people* was both "collective" and "individual." Here is an example of a result that our coders placed in the third category: [400]

> "[T]hat the *right of the people* to assemble peaceably for the purpose of consulting about public matters, and petitioning or remonstrating to the federal legislature ought not to be prevented . . . ."[401]

Fourth, we asked if the sense of *right of the people* was "ambiguous." Here is an example of an instance that our coders placed in the fourth category:

397. *The American Whig, No. XXII*, PARKER'S N.Y. GAZETTE, Aug. 8, 1768, *reprinted in* A COLLECTION OF TRACTS FROM THE LATE NEWS PAPERS 367 (New York, John Holt 1768) (emphasis added).

398. THE DECLARATION OF INDEPENDENCE para 2. (U.S. 1776) (emphasis added).

399. VICESIMUS KNOX, THE SPIRIT OF DESPOTISM 80 (London 1795) (emphasis added).

400. Two of our four coders placed this result in the "both" category. The other two coders divided: one placed this result in the "collective" category and the other placed it in the "individual" category.

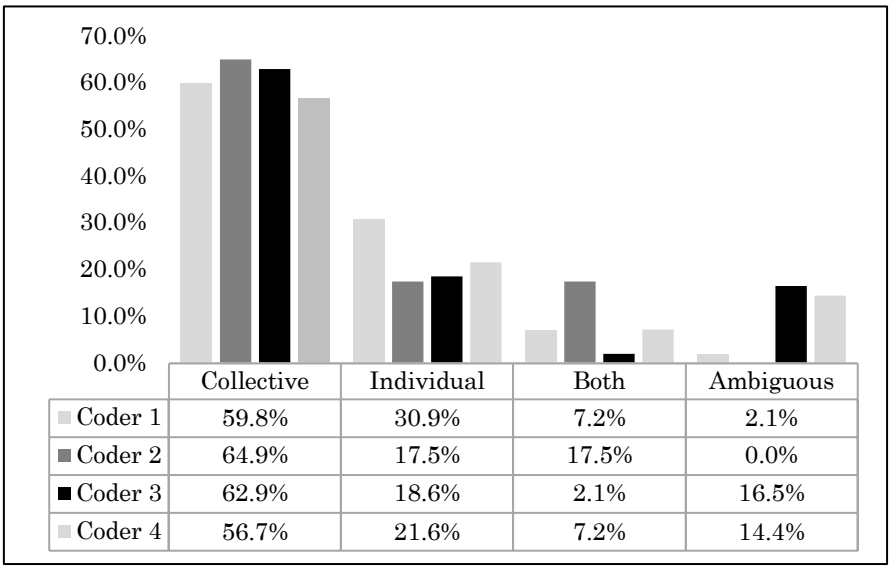
401. NATHANIEL BREADING ET AL., OBSERVATIONS ON THE PROPOSED CONSTITUTION FOR THE UNITED STATES OF AMERICA, CLEARLY SHEWING IT TO BE A COMPLETE SYSTEM OF ARISTOCRACY AND TYRANNY, AND DESTRUCTIVE OF THE RIGHTS AND LIBERTIES OF THE PEOPLE 70 (New York 1788) (emphasis added).

> "That which is the *right of the people*, therefore, is the duty of Government."[402]

Fifth, we asked if there was some other sense for *right of the people*. We posed this final question to ensure we were not unfairly pigeonholing the researchers. However, our coders did not find any other senses beyond the four categories we identified.

Figure 1 reports the results from the four coders for this query. As shown, there is a fair amount of consistency in the percentage of each sense the coders found.

FIGURE 1. SENSE DISTRIBUTION OF *RIGHT OF THE PEOPLE* (N = 97)



	Collective	Individual	Both	Ambiguous
Coder 1	59.8%	30.9%	7.2%	2.1%
Coder 2	64.9%	17.5%	17.5%	0.0%
Coder 3	62.9%	18.6%	2.1%	16.5%
Coder 4	56.7%	21.6%	7.2%	14.4%

The majority sense is the "collective" sense, ranging from 56.7%–64.9% of the results. If we add in the "both" category, then the collective sense ranges from 63.9% (Coder 4)–82.4% (Coder 2) of the results. The "individual" sense occurs between 17.5%–30.9% of the time. If we add in the "both" category, then the "individual" sense occurs between 20.7%–38.1% of the time. No matter how you slice the data, *right of the people* refers to a collective right in a majority of instances, and *right of the people* refers to an individual right in only a minority of instances.

We draw three conclusions from the data. First, in most of the instances, the phrase *right of the people* refers to the right of the people to govern themselves. The Declaration of Independence uses *right of the people* in this sense: "[T]hat whenever any Form of

402. 3 ANNALS OF CONG. 258–59 (1791) (Joseph Gales ed., 1849) (emphasis added).

Government becomes destructive of these Ends, it is the *Right of the People* to alter or to abolish it . . . ."[403] Indeed, we found that the two most common collocates of *right of the people* in COFEA were *alter* and *ends*. The frequent references to the right contained within the Declaration of Independence, even without directly quoting the document, may have skewed the results towards the "collective" category.

Second, our confidence of these results is not high. Indeed, our coders struggled with this query more than any other. As a result, we assigned the task to four independent coders. Most of our intercoder agreements of the individual search results were less than 70%, even though the coders overall came to similar conclusions as to the general percentage of each sense.

TABLE 2. INTERCODER AGREEMENT OF *RIGHT OF THE PEOPLE*

Coder Pairing	**Percent Agreement**
Coder 1 – Coder 2	71.1%
Coder 1 – Coder 3	62.9%
Coder 1—Coder 4	60.8%
Coder 2 – Coder 3	61.9%
Coder 2—Coder 4	62.8%
Coder 3—Coder 4	67.0%

In short, the coders did not reach a strong consensus. Consider the results for the "collective" category: Coder 4 selected 56.7%, Coder 1 selected 59.8%, Coder 3 selected 62.9%, and Coder 2 selected 64.9%. Two coders were below 60% and two coders were just above 60%. We are not certain which percentage is sufficient to declare a majority sense. Is an average of 50.1% enough? An average of 60%? What about 70%?

Generally, when a particular sense or meaning of a word is used more frequently in a given context, then that sense is more likely to be the appropriate one in that context. Yet frequency is not always an indication of the appropriate linguistic meaning.[404] That frequency may also reflect other facts concerning the state of the world.

We confronted other difficult questions. How should we factor the intercoder agreement into that calculation? If the coders have a higher rate of agreement, should a lower threshold be used to determine the majority sense? In contrast, if the coders have a lower

403. THE DECLARATION OF INDEPENDENCE para 2. (U.S. 1776) (emphasis added).

404. *See* MCENERY & HARDIE, *supra* note 191, at 48 (discussing the difficulty in labeling terms as "frequent").

rate of agreement, should a higher threshold be used to determine the majority sense? There are no easy answers to these questions.

The social sciences have selected .05 as the threshold for achieving statistical significance.[405] There is no real reason to select this specific standard other than having a standard that is useful but still relatively hard to obtain. Researchers in the fields of law and corpus linguistics may also need to select an arbitrary standard. Yet we recognize that each linguistic inquiry seems sufficiently dependent on a host of variables, such as how many senses are being categorized. Ultimately, a one-size-fits-all standard may prove unhelpful.

We draw a third conclusion from the data: the "both" category further complicates the selection of a majority sense. All of our coders found that some instances referred to the exercise of "both" an individual right and a collective right. Coder 3 selected 2.1% of the instances for the "both" category, and Coder 2 selected 17.5% for the "both" category. There is a huge sweep between these two poles. Consider an example from the Kentucky Constitution of 1792: "[T]he right of the people to petition for the redress of grievances, to bear arms, and to emigrate from the state."[406] This provision, adopted one year after the Second Amendment was ratified, could reasonably be read to refer to both a collective right and an individual right.

Ultimately, in light of the lower intercoder reliability, we cannot put as much weight on this first query as we can on our other four queries.

B. "Keep and Bear Arms"

Our analysis of *keep and bear arms* began with the text *keep and bear*. First, *keep and bear* was not a single linguistic unit, which is known as a *binomial*. This phrase was rarely used in the founding era. Second, lemmas of the verbs *keep* and *bear* were likewise rarely used. Third, synonyms of the verbs *keep* and *bear* were also seldom used in that phrasal form during the founding era. Fourth, the phrase *keep and bear* was not a specific figure of speech, known as a hendiadys. We thus conclude that this phrase refers to two separate components: *keeping* arms and *bearing arms*.

405. *See* Beatrice Grabowski, *"P < 0.05" Might Not Mean What You Think: American Statistical Association Clarifies P Values*, 108 J. NAT'L CANCER INST. 4, 4–5 (2016) (discussing the arbitrary P value choice and using .05 as the ubiquitous decision-making P value).

406. 1 JEDIDIAH MORSE, THE AMERICAN UNIVERSAL GEOGRAPHY, OR A VIEW OF THE PRESENT STATE OF ALL THE EMPIRES, KINGDOMS, STATES, AND REPUBLICS IN THE KNOWN WORLD AND OF THE UNITED STATES OF AMERICA IN PARTICULAR 636 (4th ed., Boston, Isaiah Thomas & Ebenezer T. Andrews 1802) (describing rights set forth by the Kentucky Constitution of 1792).

1. "Keep and Bear" was not a binomial

We first considered whether the phrase *keep and bear*, stripped of the word *arms*, was a term of art, either in ordinary language or in the language of the law. Specifically, we researched whether the phrase *keep and bear* was a binomial. A binomial is "a coordinated pair of linguistic units of the same word class which show *some* semantic relation," and are often, though not exclusively, noun pairs.[407] In the law, we often refer to binomials as *legal doublets*.[408] In *Heller*, Justice Scalia identified two such binomials that he called "terms of art": "hue and cry" and "cease and desist."[409] A third would be "aid and abet."[410]

In the law, we also deal with *multinomials*. Multinomials are not limited to pairs of two items. Rather, "[m]ultinomials are similarly chained by semantic and syntactic links, but consist of longer sequences of related words."[411] Multinomials are common in ordinary usage, such as "lock, stock, and barrel." The law, which tends to speak in triplicate, is littered with multinomials.[412] For example, "give, devise, and bequeath"[413] or "right, title, and interest" are multinomials.[414] According to one study, binomials and multinomials are found four to five times more frequently in legal writing than in nonlegal writing.[415] (We do not have comparable surveys for writings from the founding era.)

We searched COEME for *keep and bear*. That query only generated ten hits: nine quoted the Constitution, and one had nothing to do with arms. Thus, we can reasonably conclude that *keep and bear* was not a binomial in older British English.

2. Lemmas of the verbs "Keep and Bear" were not binomials

Next, we searched COFEA for the phrase *keep and bear*. We also performed queries for variants of that phrase. That is, we looked for

407. Joanna Kopaczyk & Hans Sauer, *Defining and Exploring Binomials*, *in* BINOMIALS IN THE HISTORY OF ENGLISH: FIXED AND FLEXIBLE 1, 3 (Joanna Kopaczyk & Hans Sauer eds., 2017).

408. *See* BRYAN A. GARNER, THE REDBOOK: A MANUAL ON LEGAL STYLE 224–25 (3d ed. 2013) (listing examples of legal doublets).

409. District of Columbia v. Heller, 554 U.S. 570, 591 (2008).

410. *Aid and Abet*, BLACK'S LAW DICTIONARY (11th ed. 2019).

411. Anu Lehto, *Binomials and Multinomials in Early Modern English Parliamentary Acts*, *in* BINOMIALS IN THE HISTORY OF ENGLISH: FIXED AND FLEXIBLE 261, 262 (Joanna Kopaczyk & Hans Sauer eds., 2017).

412. Marita Gustafsson, *The Syntactic Features of Binomial Expressions in Legal English*, 4 TEXT & TALK 123, 123 (1984).

413. *Give, Devise, and Bequeath*, BLACK'S LAW DICTIONARY (11th ed. 2019).

414. *Interest*, BLACK'S LAW DICTIONARY (11th ed. 2019) (referencing "right, title, and interest").

415. Gustafsson, *supra* note 412, at 123.

lemmas of the verbs, including, for example, *keeping and bearing* and *keeps and bears*. These queries yielded only twenty results. Fifteen of these results were quotes of the Second Amendment. Of the remaining five, one was a duplicate, and two were references to drafts of the Bill of Rights in Congress.

Our research suggests that the phrase *keep and bear* was used only *twice* in the entire corpus prior to 1789—the year the First Congress proposed what would become the Second Amendment.[416] (In fact, the right to keep and bear arms was originally the fourth proposed amendment.)[417] And in both cases, *keep and bear* was immediately followed by *arms*. First, the Massachusetts Declaration of Rights in 1780 provided that "the people have a right to *keep and bear arms* for the common defence."[418] Second, Virginia's ratifying convention proposed an amendment to the Constitution in June 1788: "That the people have a right to *keep and bear arms*."[419] We were also able to locate three ratifying statements outside of COFEA that used the phrase *keep and bear arms*. The North Carolina ratifying convention copied Virginia's declaration verbatim in November 1789.[420] New York's statement in July 1788 provided, "That the people have a right to *keep and bear arms* . . . ."[421] Rhode Island copied New York's statement verbatim in May 1789.[422]

We conclude that *keep and bear* was not a binomial or a term of art. There were very few usages of *keep and bear* throughout the corpora prior to 1789. We cannot be sure why this novel phrase was ultimately used in the Second Amendment; it lacked any strong linguistic pedigree.

This conclusion is bolstered by language used in contemporary state constitutions. These documents only used the phrase *bear arms*, not *keep and bear arms*. For example, in 1776, Pennsylvania's Constitution and Declaration of Rights stated: "That the people have

416. *The Bill of Rights: A Transcription*, NAT'L ARCHIVES, https://www.archives.gov/founding-docs/bill-of-rights-transcript (last visited July 17, 2021).

417. *Id.*

418. MASS. CONST., Declaration of Rights, art. XVII (1780) (emphasis added).

419. *Convention of Virginia* (1788), *reprinted in* 3 DEBATES IN THE SEVERAL STATE CONVENTIONS ON THE ADOPTION OF THE FEDERAL CONSTITUTION 657–59 (Washington, Jonathan Elliot ed., 1st ed. 1836) (emphasis added).

420. *Convention of North Carolina* (1789), *reprinted in* 1 DEBATES IN THE SEVERAL STATE CONVENTIONS ON THE ADOPTION OF THE FEDERAL CONSTITUTION 242–44 (Washington, Jonathan Elliot ed., 1st ed. 1836).

421. *Ratification—New York* (1788), *reprinted in* 1 DEBATES IN THE SEVERAL STATE CONVENTIONS ON THE ADOPTION OF THE FEDERAL CONSTITUTION 327–28 (Washington, Jonathan Elliot ed., 1st ed. 1836) (emphasis added).

422. *Ratification—Rhode Island* (1789), *reprinted in* 1 DEBATES IN THE SEVERAL STATE CONVENTIONS ON THE ADOPTION OF THE FEDERAL CONSTITUTION 371 (Washington, Jonathan Elliot ed., 1st ed., 1836).

a right to *bear arms* for the defence of themselves and the state . . . ."[423] Here, the drafters only used *bear* and not *keep*. Vermont and Kentucky used identical language for their constitutions in 1786 and 1792, respectively.[424] In 1790, Pennsylvania's Constitution adopted similar language: "The right of the citizens to *bear arms*, in defence of themselves and the state, shall not be questioned."[425] And, the Pennsylvania ratifying minority's statement declared: "That the people have a right to *bear arms* for the defense of themselves and their own state, or the United States, or for the purpose of killing game . . . ."[426] These usages, some drafted while the Bill of Rights was being ratified,[427] lend further support to the conclusion that *keep and bear arms* was not a term of art.

The variations do raise a different question. Some state constitutions protected a right to *bear arms*. Other states protected a right to *keep and bear arms*. Did these states think that they were protecting a different right—or at least a right with a different scope? Or were these phrases different verbal formulations of the same right? If the former, we would expect to see contemporaneous discussions of the differences between the states. If the latter, we would not expect people during the founding era to draw attention to these differences in wordings. We have not investigated which option is more supported by the historical record.

3. Synonyms of "Keep and Bear" were not binomials

Next, we turned to another possibility: The words *keep* or *bear* could be switched out with synonyms and yield the same meaning. To test this hypothesis, we searched COFEA for a series of synonyms, including the lemma of each verb. First, we considered three possible synonyms for *keep*: *have*, *own*, and *possess*. Second, we considered one synonym for *bear*: *use*. Third, we considered one synonym for *arms*: *guns*.

These queries yielded only six results; one was irrelevant and two were duplicates of a third. Thus, the entire corpus yielded only three unique, relevant results. First, a publication in 1765 referred to the

423. PA. CONST. of 1776, Declaration of Rights, art. XIII (emphasis added).

424. VT. CONST. of 1786, Declaration of Rights, art. 18; KY. CONST. of 1792, art. XII, § 23.

425. PA. CONST. of 1790, art. IX, § XXI (emphasis added).

426. *Ratification of the Constitution by the States: Pennsylvania*, *in* 2 THE DOCUMENTARY HISTORY OF THE RATIFICATION OF THE CONSTITUTION 597–98 (Merrill Jensen ed., 1976) (emphasis added).

427. *The Bill of Rights: How Did It Happen?*, NAT'L ARCHIVES, https://www.archives.gov/founding-docs/bill-of-rights/how-did-it-happen (last visited July 17, 2021) (explaining that George Washington sent what would later become the Bill of Rights to the states on October 2, 1789, and three-fourths of the states had ratified the Bill of Rights by December 15, 1791).

"natural absolute personal rights of individuals," which included "[t]he right of petitioning for redress of grievances" and "[t]he right of having and using arms for self-defence."[428]

Second, a 1769 book on the laws of colonial Virginia explained:

> [E]very free Negro, Mulatto, or Indian, being a Housekeeper, may be permitted to *keep one Gun, and Powder and Shot*; and all Negroes, Mulattoes, and Indians, bond or free, living at any Frontier Plantation, may be permitted to *keep and use Guns, Powder, Shot, and Weapons, offensive or defensive*, by License from a Justice of Peace of the County wherein such Plantations lie, to be obtained upon the Application of free Negroes, Mulattoes, or Indians, or of the Owners of such as are Slaves.[429]

(This entry appears three times in COFEA.)

Third, an Englishman in 1794 referred to Blackstone's *Commentaries* and "the right of having and using arms for self-preservation and defence."[430]

Prior to 1789, the following construction was very rare: *keep* (or a synonym) and *bear* (or a synonym) *arms* (or a synonym). This construction likely did not have a distinct meaning.

4. "Keep and Bear" was likely not a hendiadys

We considered another possibility: *keep and bear* could be a hendiadys. A hendiadys is "a figure of speech in which two terms, separated by a conjunction, are melded together to form a single complex expression."[431] Examples include "good and dry," "law and order," "try and do better," or "rise and shine."[432] Professor Samuel Bray concluded that, in the Constitution, the phrases "necessary and proper" and "cruel and unusual" are hendiadyses.[433] Professor Bray identified two categories of hendiadyses. In the first category, the first word may modify the second: "nice and hot" would mean "nicely hot."[434] In the second category, "each word ha[s] its due, instead of

428. JAMES OTIS, *A Vindication of the British Colonies*, *in* THE COLLECTED POLITICAL WRITINGS OF JAMES OTIS 189 (Richard Samuelson ed., 2015) (emphasis added).

429. THE ACTS OF ASSEMBLY, NOW IN FORCE, IN THE COLONY OF VIRGINIA 261 (Williamsburg, W. Rind, A. Purdie, & J. Dixon 1769) (emphasis added).

430. THE TRIAL OF JOSEPH GERRALD, DELEGATE FROM THE LONDON CORRESPONDING SOCIETY, TO THE BRITISH CONVENTION 171 (1794) (Samuel Campbell, Pearl-Street 1794) (quoting 1 WILLIAM BLACKSTONE, COMMENTARIES *144).

431. Samuel L. Bray, *"Necessary and Proper" and "Cruel and Unusual": Hendiadys in the Constitution*, 102 VA. L. REV. 687, 695 (2016).

432. *Id.* at 699–700.

433. *Id.* at 688.

434. *Id.* at 704.

letting one serve the other . . . [with] the terms in the hendiadys . . . remain[ing] distinct."[435]

COFEA suggests that *keep and bear* does not fall into the first category of hendiadyses. For example, the Massachusetts Constitution of 1780 stated that "[t]he people have a right *to* keep and *to* bear arms for the common defence."[436] Here, the words *keep* and *bear* are used as distinct verbs in their infinitive forms: *to keep* and *to bear*. Consider another example from the First Congress: The members debated a proposed constitutional right, in which "no person religiously scrupulous shall be compelled to render military service in person."[437] Representative Thomas Scott of Pennsylvania objected to this draft.[438] He worried that if this right became a part of the Constitution, it would, among other ills, "lead to the violation of another article in the [C]onstitution, which secures to the people the right of keeping arms, and in this case recourse must be had to a standing army."[439] The amendment referenced *bear arms*, and Representative Scott's argument referenced *keep arms*. He did not view these terms as synonymous. Representative Scott's argument implies that *keeping* was thought to have an independent meaning from *bearing*. Further, during this time, people often used synonyms of *keep* and *bear* as distinct verbs (e.g., *having* and *using*). The concepts of *keep arms* and *bear arms* appear to be separate. Therefore, the phrase *keep and bear* was likely not the first type of a hendiadys. If *keep and bear* was the second type, it would not matter; each part would have its own meaning.

Thus, we conclude that *keep and bear arms* was not a binomial or a term of art.

C. Rights Related to Arms

The operative clause focuses primarily on the *right to arms*. *Keeping arms* and *bearing arms* are merely ways to exercise that *right*, regardless of whether that right is collective or individual. In order to determine how the concept of a *right* was used in conjunction with *arms*, we queried *right* within six words of *arms*.

Our search of COFEA generated forty-six results. At first glance, this number may seem quite small. And a larger sample size is generally more helpful than a smaller one. But we are not dealing with a sample. The *entire* population of instances within COFEA was forty-six instances. And this population is a significantly higher number of instances than Justices Scalia or Stevens examined.

435. *Id.*
436. MASS. CONST. of 1780, Declaration of Rights, art. XVII (emphasis added).
437. 1 ANNALS OF CONG. 451 (1789) (Joseph Gales ed., 1834).
438. *Id.* at 796.
439. *Id.*

Ultimately, we were not deterred by the small population size. Consider an example: If we were researching trends for the 2020 presidential election, relying on a sample size of forty-six American voters would be malpractice. But if we were researching American veterans of World War I, it would be perfectly acceptable, and commendable, to survey forty-six people. (Of course, the last known American World War I veteran, Frank Buckles, died in 2011.[440]) This small population size from COFEA may reflect the fact that the founding era did not often speak of the concept of rights in the context of arms. Once again, the Framers' choice to use fairly unusual language in the Second Amendment further complicates the interpretation of that provision.

We asked three coders to categorize each result based on five possible senses. First, we asked if the sense of the *right to arms* was "collective/military." Here are three examples of instances that our coders placed in the first category:

> "[I]t was often said, that they had a *Right to oppose with Arms* a Military Force, which was sent to oblige them to submit to unconstitutional Laws . . . ."[441]

> "The people have a *right to keep and to bear arms* for the common defence."[442]

> "That the people have a *right to bear arms*, for the defence of the State . . . ."[443]

Second, we asked if the sense of the *right to arms* was "individual." Here are three examples of instances that our coders placed in the second category:

> "The *right of having and using arms* for self-defence."[444]

> "Christians have a *right to self-defence*. *His disciples carried arms* to defend themselves from enemies, as appears by one of

440. *See* Paul Duggan, *Frank Buckles, Last Known U.S. World War I Veteran, Is Laid to Rest at Arlington*, WASH. POST (Mar. 15, 2011), https://wapo.st/31RAClX.

441. Letter from Governor Bernard to the Earl of Hillsborough (Sept. 16, 1768) (on file at Evans Early American Imprint Collection) (emphasis added).

442. MASS. CONST. of 1780, Declaration of Rights, art. XVII (emphasis added).

443. N.C. CONST. Declaration of Rights, art. § XVII (1776) (emphasis added).

444. JAMES OTIS, A VINDICATION OF THE BRITISH COLONIES, AGAINST THE ASPERSIONS OF THE HALIFAX GENTLEMAN, IN HIS LETTER TO A RHODE-ISLAND FRIEND 8 (Boston, Edes & Gill 1765) (emphasis added).

> them drawing his sword to defend Christ at the time of his surrendry of himself to his persecutors."[445]

> "[A]ll the armed prophets have succeeded, and the unarmed ones have failed. That Mahomet had a *right to take up arms* for his own-defence, may perhaps be allowed . . . ."[446]

Third, we asked if the sense of the *right to arms* was both "collective/military" and "individual." Here are two examples that our coders placed in the third category:

> "That the people have a *right to bear arms* for the defence of themselves and the state . . . ."[447]

> "[T]he people have a *right to bear arms* for the defence of themselves and their own State, or the United States, or for the purpose of killing game."[448]

Fourth, we asked if the sense of the *right to arms* was "ambiguous." Fifth, we asked if there was some other sense, but our coders did not find any other senses beyond the four we identified.

The coders tended to see the overall sense distribution very similarly. We achieved an intercoder agreement above our threshold of 70%. In light of this high level of agreement, we only used three coders.

TABLE 3. INTERCODER AGREEMENT OF *RIGHT* /6 *ARMS*

Coder Pairing	Percent Agreement
Coder 1 – Coder 2	76.1%
Coder 1 – Coder 3	71.7%
Coder 2 – Coder 3	73.9%

Figure 2 reports the results from the three coders we used for this query.

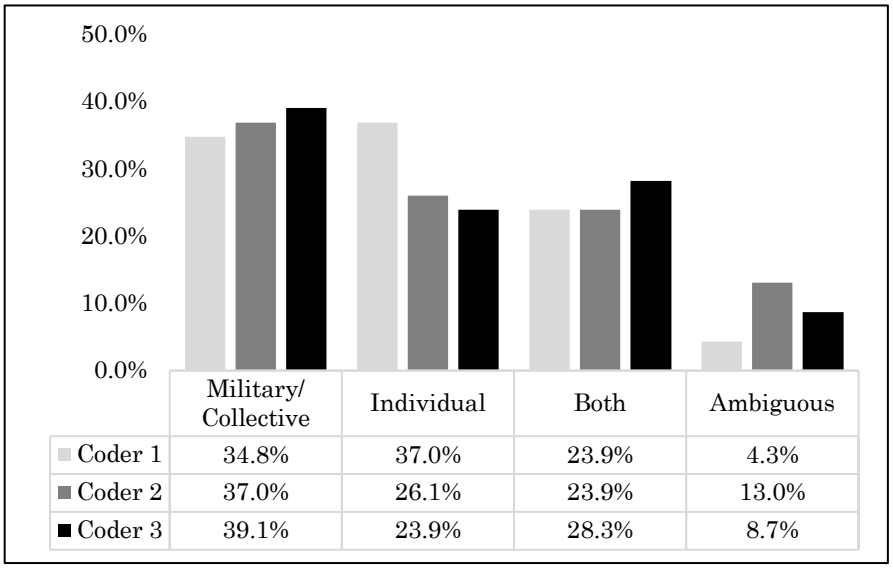
445. DAVID M'CLURE, SERMONS ON THE MORAL LAW; ELUCIDATING THE NATURE, EXTENT AND OBLIGATIONS OF THE VARIOUS SOCIAL AND DIVINE VIRTUES, COMPRISED IN THAT SUMMARY OF UNIVERSAL DUTY – AND ON THE CONNECTION OF THE MORAL LAW AND THE GOSPEL 192 (n.p., 1795) (emphasis added).

446. JAMES WILSON STEVENS, AN HISTORICAL AND GEOGRAPHICAL ACCOUNT OF ALGIERS; COMPREHENDING A NOVEL AND INTERESTING DETAIL OF EVENTS RELATIVE TO THE AMERICAN CAPTIVES 173 (Philadelphia, Hogan & M'Elroy 1797) (emphasis added).

447. PA. CONST. of 1776, Declaration of Rights, art. § XIII (emphasis added).

448. PENNSYLVANIA AND THE FEDERAL CONSTITUTION 1787–1788, at 422 (John Bach McMaster & Frederick Dawson Stone eds., n.p. 1888) (emphasis added).

FIGURE 2. SENSE DISTRIBUTION FOR *RIGHT* /6 *ARMS* (N = 46)



	Military/ Collective	Individual	Both	Ambiguous
Coder 1	34.8%	37.0%	23.9%	4.3%
Coder 2	37.0%	26.1%	23.9%	13.0%
Coder 3	39.1%	23.9%	28.3%	8.7%

There is no majority sense. That is, no sense prevailed more than 50% of the time. On average, the "military" sense (34.8%–39.1%) was used only slightly more frequently than the "individual" sense (23.9%–37.0%). Additionally, 23.9%–28.3% of the results considered both the "military" and "individual" senses.

As a result, we cannot state that the concept of *rights* in the context of *arms* had a dominant public meaning at the time of the framing. And this context—*rights* and *arms*—is arguably the most relevant to the Second Amendment inquiry.

D. "Keep Arms"

Next, we searched COFEA for the lemma of *keep* within six words of *arms*. We will refer to this query as *keep* /6 *arms*. This search allowed us to consider many different grammatical constructions, such as *she kept arms*, *he was keeping arms*, and *the arms were kept by the people*. But this broad search could yield many irrelevant results, such as *she kept her arms above her head while she danced*.

Our query returned 237 results. We disregarded 198 of those lines (83.5% of the results) because they were duplicates, irrelevant results, or quotes of the Second Amendment. After this culling, we were left with only thirty-nine results.

We further refined the data and excluded instances where arms were *kept* by a regular soldier or a professional. Consider a concordance line that discussed the British Army keeping arms in its storehouse. This result does not tell us anything about the Second Amendment, at least in the context of the *Heller* debate. The debate

in *Heller* was over whether the Second Amendment concerned keeping arms in a *non-professional* army setting, either by the community or by individuals for militia or personal use.[449] Therefore, concordance lines about a professional army are not relevant to the collective/individual dichotomy. Even the *Heller* dissenters did not contend that the Second Amendment concerns the right to serve in a professional, standing army. To the contrary, Justice Stevens wrote that the state militias would serve as a check on a federal standing army.[450] To be overinclusive, we included any results that could concern *either* the state militia or a professional army.

For these queries, we did not think a simple dichotomy between collective and individual rights would suffice. Instead, we adopted a more nuanced approach to classifying the lines. We considered separately how the arms were stored and how the arms were utilized. Were arms *kept* in a "collective" or "individual" fashion? And were the arms *used* in a "collective" or "individual" fashion? This classification made the most sense to us given the arguments raised in the various opinions in *Heller*, since knowing only that an individual kept arms without knowing what those arms were kept for would not tell us whether Justice Scalia or Justice Stevens was correct. We asked coders to categorize each result based on five possible senses.

First, we asked if the sense of *keep* /6 *arms* involved "collective" keeping arms for "collective" use—that is, the government stored the arms for the militia to use. Here are two examples of instances that our coders placed in the first category:

> "The powder and *arms* of the country which were *kept* at Boston, were by order of the last court carried to Roxbury and Newtown."[451]

> "Unanimously vote to raise 10,000£ to be laid out by the County Committees in *Arms* and Ammunition, to be kept and disposed of by the Committees, as they shall think proper."[452]

449. *See* District of Columbia v. Heller, 554 U.S. 570, 573 (2008).

450. *Id.* at 661 (Stevens, J., dissenting) ("The history of the adoption of the [Second] Amendment thus describes an overriding concern about the potential threat to state sovereignty that a federal standing army would pose, and a desire to protect the States' militias as the means by which to guard against that danger.").

451. JOHN WINTHROP, THE HISTORY OF NEW ENGLAND FROM 1630–1649, at 251 (James Savage ed., Boston, Phelps & Farnham 1825) (emphasis added).

452. Letter from John Adams to James Warren (Jan. 3, 1775) (Founders Online) (emphasis added), https://founders.archives.gov/documents/Adams/06-02-02-0067.

Second, we asked if the sense of *keep* /6 *arms* involved "individual" keeping arms for "collective" use—that is, individuals stored the arms which they could use in their militia service. We think this is the category closest to Justice Stevens's conception of the Second Amendment. We suspect, however, that Justice Stevens would view the first category as also being consistent with his understanding of the Second Amendment. In the 1770s, arms were sometimes kept in communal storehouses for militia purposes. Paul Revere's famous ride and the Battle of Lexington and Concord involved arms stored in such warehouses.[453]

Here are two examples of instances that our coders placed in the second category:

> "Even those persons who were exempted from appearing at the common military trainings, were obliged to *keep* the same *arms* and ammunition."[454]

> [T]his disciplined militia will be unequal to oppose a sudden and powerful attack, by reason of the dispersed situation of their places of abode, every man, from twenty four to forty years of age, must be subject to military service upon an alarm; and should therefore *keep* himself constantly provided with *arms* and accoutrements; which should be inspected annually by the officers of the company to which he belongs . . . .[455]

Third, we asked if the sense of *keep* /6 *arms* involved "individual" keeping of arms for "individual" use—that is, individuals stored the arms which they could use for their own individual use. We think this category is closest to Justice Scalia's conception of the Second Amendment. Here are two examples of instances that our coders placed in the third category:

> "Fid: John, you must take Mr. Plunket's advice; — you don't know so much about 'em as we do they all *keep* their *arms*, and if you shou'd molest 'em, you'd die first, I'll warrant ye. John: What, do they carry pistols with 'em? Fid: Indeed they do, to my certain knowledge."[456]

---

453. *First Shots of War, 1775*, LIBR. CONG., https://www.loc.gov/classroom-materials/united-states-history-primary-source-timeline/american-revolution-1763-1783/first-shorts-of-war-1775/ (last visited July 17, 2021).

454. 2 JEREMY BELKNAP, THE HISTORY OF NEW HAMPSHIRE: COMPREHENDING THE EVENTS OF SEVENTY FIVE YEARS, FROM 1715 TO 1790, at 389 (Boston, Isaiah Thomas & Ebenezer T. Andrews 1791) (emphasis added).

455. Letter from Timothy Pickering to George Washington (Apr. 22, 1783) (Founders Online) (emphasis added), https://founders.archives.gov/documents/Washington/99-01-02-11138.

456. THE DOUBLE CONSPIRACY, OR TREASON DISCOVERED BUT NOT PUNISHED 84 (n.p., 1783) (emphasis added).

"No slave shall *keep* any *arms* whatever, nor pass, unless with written orders from his master or employer, or in his company, with arms from one place to another. Arms in possession of a slave contrary to this prohibition shall be forfeited to him who will seize them."[457]

Fourth, we asked if the sense of *keep* /6 *arms* involved "collective" keeping arms for "individual" use—that is, the government stored the arms which people could then retrieve for their own individual use. Our coders did not find any instances in this fourth category. We were not surprised. It is difficult to conceive of a situation where a state maintained a repository of weapons that people could then use for private purposes. We are only aware of one historian who advanced such an account. Michael Bellesiles wrote that, in early America, the government maintained firearms in collective storage units that people could then use for individual purposes, such as hunting or self-defense.[458] His 2000 book, *Arming America: The Origins of a National Gun Culture*, was awarded the Bancroft Prize.[459] That prestigious award was later rescinded, however, due to "evidence of falsification" and "serious failures of and carelessness in the gathering and presentation of archival records and the use of quantitative analysis."[460]

Fifth, we asked if the sense of *keep* /6 *arms* involved an "ambiguous" sense. That is, the instance did not fit into any of the previous four categories. Only one of the three coders found any ambiguous instances.

For these queries, our coders were largely in agreement as to the percentage of each sense. In light of this high level of agreement, we only used three coders.

TABLE 4. INTERCODER AGREEMENT FOR *KEEP* /6 *ARMS*

Coder Pairing	**Percent Agreement**
Coder 1 – Coder 2	84.6%
Coder 1 – Coder 3	87.2%
Coder 2 – Coder 3	79.5%

Figure 3 reports the results from the three coders we used for this query.

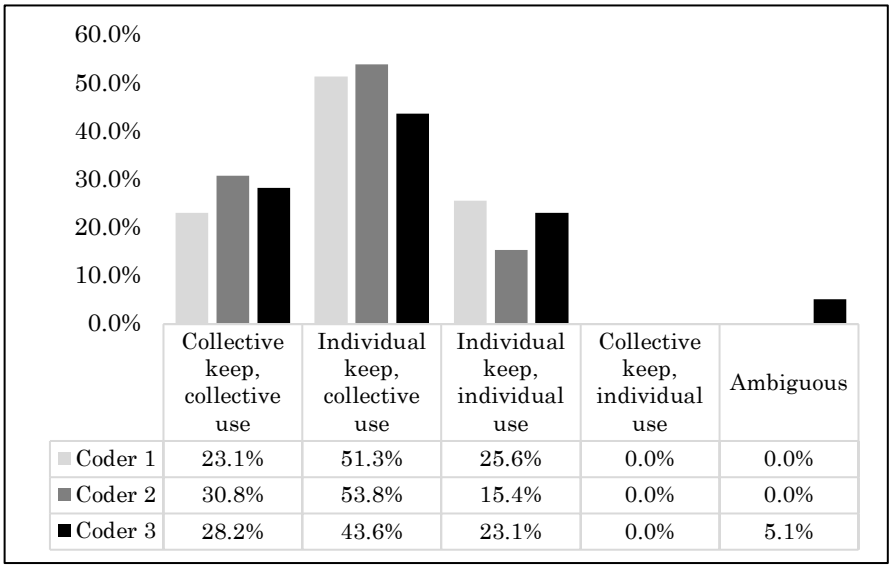
457. H.B. 51, 1785 Gen. Assemb., (Va. 1785) (emphasis added).

458. MICHAEL BELLESILES, ARMING AMERICA: THE ORIGINS OF A NATIONAL GUN CULTURE 73 (2000).

459. *The Bancroft and Bellesiles*, COLUMBIAN COLL. ARTS & SCIS.: HIST. NEWS NETWORK (Dec. 13, 2002), http://hnn.us/articles/1157.html.

460. *Id.*; *see also* James Lindgren, *Fall from Grace:* Arming America *and the Bellesiles Scandal*, 111 YALE L.J. 2195, 2197, 2201 (2002) (reviewing BELLESILES, *supra* note 458).

FIGURE 3. LEMMA OF *KEEP* /6 *ARMS* (N = 39)



	Collective keep, collective use	Individual keep, collective use	Individual keep, individual use	Collective keep, individual use	Ambiguous
Coder 1	23.1%	51.3%	25.6%	0.0%	0.0%
Coder 2	30.8%	53.8%	15.4%	0.0%	0.0%
Coder 3	28.2%	43.6%	23.1%	0.0%	5.1%

The first category involved "collective-keep/collective-use." Our coders selected between 23.1%–30.8% of the instances for this category.

The second category involved "individual-keep/collective-use." Two of our three coders found a majority sense a smidge above 50%. The third coder selected 43.6% of the instances for this second category. Instances coded as falling in the second category occurred more frequently than instances coded as falling in the other categories.

The third category involved "individual-keep/individual-use." Our coders selected between 15.4%–25.6% instances for this third category.

The fourth category involved "collective-keep/individual-use." Our coders found zero such results. Based on our corpus analysis, we conclude that this fourth sense was not used.

The fifth category involves "ambiguous" instances. Two of our coders found no ambiguous lines. The third coder found 5.1% of the lines were ambiguous.

So far, we have considered each category in isolation. But we can also combine some of the results. If we merge the first and second categories, about three-quarters of the instances referred to "collective-use" (71.8%–84.6%). If we merge the second and third categories, about three-quarters of the instances referred to "individual-keep" (69.2%–76.9%). But the references to "individual use" only ranged from 15.4%–25.6%.

We acknowledge one possible limitation on our research. The four categories we selected may reflect distinct senses, or they may reflect different contexts. Consider two examples involving the word *keep*: First, a person can *keep* books in his house. This phrase reflects an individual sense, not a collective sense. Second, a town can *keep* books in a library. This same word, in a different phrase, could reflect a collective sense of keeping for individual use. Or perhaps the library example does not reflect either an individual sense or a collective sense. Rather, the context is altogether different. There are different senses of *keep* and different ways to *keep*.

Consider a hypothetical: Perhaps people were allowed to keep a gun in their homes for militia service. But when the militia was not in service, the people were also free to use the gun for personal purposes, such as hunting or self-defense. In other words, the guns were not locked up and could legally be used for multiple uses. COFEA would not indicate if the phrase *keep arms* was limited *solely* to a collective purpose. Based on the context, the phrase could have different meanings.

E. "Bear Arms"

Our final query involved the lemma of *bear* within six words of *arms*. We will refer to this query as *bear* /6 *arms*. COFEA identified seven variations of *bear*: *born*, *bore*, *bearing*, *bears*, *borne*, and *beared*. We also searched COFEA for *bear weapons* and *bear guns*. But those queries did not generate any results. The phrase *bear arms* may have taken on a specialized meaning that is more than the sum of the meaning of *bear* and *arms*.

Our search of *bear* /6 *arms* yielded 718 results. Josh Jones looked at a similar data set and found 727 results.[461] We suspect he used either an older or newer version of COFEA than we used. We then randomly sampled fifty results from the Hein corpus (legal materials), and another fifty results from the Evans corpus (ordinary materials). These instances generated a sample size of one hundred. We conducted multiple samples to eliminate duplicates. We conservatively estimated that there were approximately 600 unique search results. Jones, by our estimates, located approximately 500 unique results.[462]

We were confident with this sample size of one hundred—especially because our results turned out to be so lopsided. Our sampling has a margin of error of 7% at a 90% confidence level, or 9% at a 95% confidence level. (Our margin of error would be even smaller if we used Jones's count of 500 unique, relevant results.) Ultimately,

461. Jones, *supra* note 249, at 161.

462. *Id.* (rejecting 31.9% of his sample due to duplication, quotes of the Second Amendment, or irrelevancy).

these margins of error did not change our results. We encourage other scholars to perform the same search and review as large of a sample as they deem appropriate.

For this query, we asked our coders to categorize each result based on four possible senses. First, we asked if the sense of *bear* /6 *arms* was "collective/military." That is, bearing arms for the *common defense*. Here are two examples of instances that our coders placed in the first category:

> "To complete their Security, Succours of all Kinds had been thrown into Quebec, and a numerous Body of regular Troops, joined to the Troops of the Colony, filled up with every Canadian that was able to *bear Arms*, besides several Nations of Savages, had taken the Field in a very advantageous Situation."[463]

> "Their religion, it seems, will not suffer them to *bear arms*, What can be more ridiculous than this principle, to a man who knows human nature, except the people who indulge them in this humour? What right have any set of men to the protection of a government in times of peace, who will not assist with every power they possess to defend their country in times of war?"[464]

Second, we asked if the sense of *bear* /6 *arms* was "individual." For example, a person using arms for self-defense or for hunting. Here is an example of an instance that our coders placed in the second category:

> "A soldier, according to [a British Army officer's] directions, sold an old rusty musket to a countryman for three dollars, who brought vegetables to market. This could be no crime in the market-man, who had an undoubted right to purchase, and *bear arms*. He was, notwithstanding, immediately seized by [the British officer], and conveyed to the guard-house, where he was confined all night. Early the next morning they stripped him naked, covered him with warm tar, and then with feathers, and conducted him to the north end of the town, then to the south end, and as far as liberty-tree, where they dismissed the man, through fear of the people . . . ."[465]

Third, we asked if the sense of *bear* /6 *arms* was both "collective/militia" and "individual." Here is an instance that our coders placed in the third category:

---

463. JOHN PRINGLE, THE LIFE OF GENERAL JAMES WOLFE, THE CONQUEROR OF CANADA: OR, THE ELOGIUM OF THAT RENOWNED HERO, ATTEMPTED ACCORDING TO THE RULES OF ELOQUENCE 21–22 (n.p., 1760) (emphasis added).

464. Letter from Batista Angeloni to Manzoni (Lancaster County, 1764) (on file with Evans Early American Imprint Collection) (emphasis added).

465. JOHN TRUMBULL, M'FINGAL: A MODERN EPIC POEM IN FOUR CANTOS 51 (1812) (emphasis added).

> "That the people have a right to *bear arms* for the defence of themselves and the State . . . ."[466]

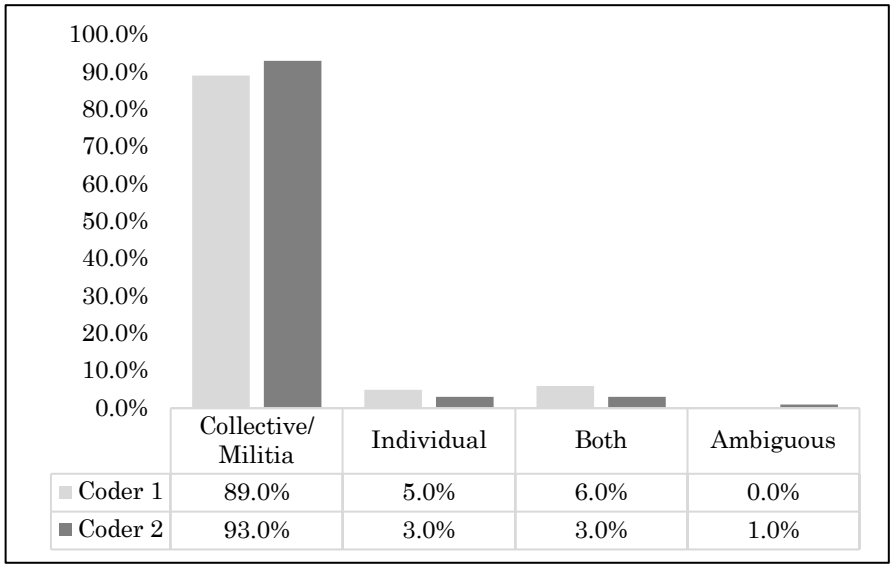
Fourth, we asked if the sense of *bear* /6 *arms* was "ambiguous." Only one coder placed one line in this category.

For this query, there was a clear predominant usage. Roughly 90% of the instances were in the first category. The breakdown was not even close. The overwhelming majority of *bear arms* was the "collective/militia" sense.

For some lines, *bear arms* would have surrounding language that indicated the collective/militia sense, such as *bear arms in the militia*, or *bear arms for the publick defence*. Even without these modifiers, however, the context was generally clear enough to suggest that *bear arms* concerned a "collective/militia" sense.

Our coders found very few instances that referred to an "individual" sense (3%–5%). And our coders found roughly the same number of instances (3%–6%) that reflected a combination of both a "collective/militia" sense and an "individual" sense. One coder found zero ambiguous instances; the other found a single instance that was ambiguous. Finally, the intercoder agreement between our first two coders was 93%. This rate was so high that we did not deem it necessary to ask a third coder to assist. Our results are depicted in

FIGURE 4. LEMMA OF *BEAR* /6 *ARMS* (N = 100)



	Collective/ Militia	Individual	Both	Ambiguous
Coder 1	89.0%	5.0%	6.0%	0.0%
Coder 2	93.0%	3.0%	3.0%	1.0%

466. VT. CONST. of 1777, ch. 1, art. XV.

V. GRADING *HELLER*

In Part I, we identified four linguistic claims that Justices Scalia and Stevens made about the operative clause. Here, we grade each of these four claims. First, we consider whether the phrase *right of the people* in the Second Amendment refers to an individual right or a collective right. Second, we discuss whether *keep and bear arms* was a "unitary right" or consisted of two separate rights. Third, we focus on whether *keep arms* involved an individual right or a collective right. Fourth, we turn to the phrase *bear arms*, and we resolve whether it predominantly referred to an individual right or a collective right. We conclude by grading the operative clause itself.

A. Claim #1: "Right of the People"

In *Heller*, Justice Scalia wrote that where the phrase *the people* is coupled with a *right*, then the Constitution speaks to an individual right.[467] Justice Scalia overstated his case. Our research suggests that the phrase *the people* coupled with a *right* referred to a collective right about 60% of the time. This coupling referred to an individual right about 20% of the time.

By contrast, Justice Stevens determined that the phrase *right of the people* in the Second Amendment referred to a collective right.[468] In light of our research in COFEA, Justice Stevens selected the predominant sense. But Justice Stevens was perhaps a bit too confident in reaching this conclusion. There is still a distinct possibility that this provision of the Second Amendment did not follow the predominant sense.

In the founding era, the collective sense and the individual sense were used at roughly equal rates when speaking about rights in the context of arms. Our research suggests that the phrase *right of the people* could have implicated both a collective right and an individual right.

Moreover, Justice Stevens's debate with Justice Scalia over the First and Fourth Amendments reflects the potential dual nature of *right of the people*, as used in the Bill of Rights. As explained in Part I, Justice Scalia's majority opinion used an intratextualist analysis that looked to the First and Fourth Amendments.[469] The Assembly and Petition Clause provides: "Congress shall make no law . . . abridging . . . the *right of the people* peaceably to assemble, and to petition the Government for a redress of grievances."[470] The Search and Seizure Clause states: "The *right of the people* to be secure

467. District of Columbia v. Heller, 554 U.S. 570, 580 (2008).

468. *Id.* at 637 (Stevens, J., dissenting).

469. *See* Amar, *supra* note 62, at 747 (examining the intratextualist approach to legal analysis).

470. U.S. CONST. amend. I (emphasis added).

in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated . . . ."[471] Justice Scalia thought the phrase *right of the people* in both provisions was individual in nature.[472]

Justice Stevens disagreed. He observed that both the rights under the Assembly and Petition Clause "contemplate collective action."[473] Justice Stevens acknowledged that "the Fourth Amendment describes a right that need not be exercised in any collective sense."[474] But he thought the Fourth Amendment differed from the rights in the First and Second Amendments: "[T]he Fourth Amendment describes a right *against* governmental interference rather than an affirmative right to engage in protected conduct, and so refers to a right to protect a purely individual interest."[475] Justice Scalia disagreed on both fronts. He wrote that the First and Fourth Amendments "refer to individual rights, not 'collective' rights, or rights that may be exercised only through participation in some corporate body."[476]

Justice Scalia had strong arguments, as did Justice Stevens. Perhaps the way to resolve this debate over the phrase *right of the people* in the First and Fourth Amendments is to characterize them as *both* individual and collective rights. Groups can petition collectively—for example, many people can sign a petition. Alternatively, an individual can send a standalone petition. The right to assemble, in contrast, could only be exercised in a group—unless you consider a single person on a soapbox to be an assembly. And the right to be free from searches and seizures could be viewed as a right that individuals possess or as a right that society has against the government.

Justices Scalia's and Stevens's positions are not self-evidently correct. Rights do not easily reduce to a modernistic binary frame. And the fact that the language could be used in both fashions suggests that intratextualism may not be the best method to interpret the phrase *right of the people* in the Second Amendment. It is possible that the Second Amendment used the phrase *right of the people* in a different fashion than the First and Fourth Amendments did.

Ultimately, the results from COFEA do not help us establish, with a sufficient degree of confidence, the meaning of *right of the people* in the Second Amendment.

On this first claim, Justice Scalia gets a B, and Justice Stevens gets a C. Justice Scalia gets a higher grade because he entertained

471. U.S. CONST. amend IV (emphasis added).
472. *Heller*, 554 U.S. at 579.
473. *Id.* at 645 (Stevens, J., dissenting).
474. *Id.*
475. *Id.* at 646.
476. *Id.* at 579 (majority opinion).

the possibility that the *right of the people* could in some cases refer to a collective right and in other cases an individual right. The linguistic evidence provides more support for Justice Scalia's position. Justice Stevens confidently concluded that *right of the people*, as an "affirmative right to engage in protected conduct," can *only* refer to a collective right.[477] On that front, he is wrong.

B. Claim #2: "Keep and Bear Arms"

Neither Justice Scalia nor Justice Stevens performed any linguistic analysis to determine whether *keep and bear arms* was a single linguistic unit. Justice Scalia dissected the phrase into two components: *keep arms* and *bear arms*.[478] Justice Stevens disagreed. He stated unequivocally that *keep and bear arms* "describe[s] a unitary right."[479] He added that the Second Amendment "protects only one right, rather than two."[480] Indeed, Justice Stevens criticized the majority opinion's "atomistic, word-by-word approach to construing the [Second] Amendment."[481] In response, Justice Scalia responded that Justice Stevens's "single right" argument "suggests that 'keep and bear Arms' was some sort of term of art, presumably akin to 'hue and cry' or 'cease and desist.'"[482] Here, Justice Scalia was referring to the concept of a binomial.[483]

COFEA reflects that Justice Scalia was right and that Justice Stevens was wrong. *Keep and bear* was neither a "term of art" nor a binomial.

On this second claim, Justice Stevens receives an F. He took a swing and missed everything. Justice Scalia receives a C. Even though Justice Scalia got the right answer, we still dock some points. Like many bright students, he did not show his work. He offered no explanation as to why *keep and bear arms* was not a "term of art." But his intuition allowed him to stumble upon the correct answer.

C. Claim #3: "Keep Arms"

In *Heller*, Justice Scalia concluded that "'[k]eep arms' was simply a common way of referring to possessing arms, for militiamen *and everyone else*."[484] That is, he concluded that the phrase *keep arms* was sometimes used in a military sense and sometimes used in a personal sense. In contrast, Justice Stevens's dissenting opinion asserted that

477. *Id.* at 646 (Stevens, J., dissenting).
478. *Id.* at 582 (majority opinion).
479. *Id.* at 646 (Stevens, J., dissenting).
480. *Id.* at 651.
481. *Id.* at 652 n.14.
482. *Id.* at 591 (majority opinion).
483. *See supra* notes 409–10 and accompanying text.
484. *Heller*, 554 U.S. at 583.

the "[Second] Amendment's use of the term 'keep' in no way contradicts the military meaning conveyed by the phrase 'bear arms' and the [Second] Amendment's preamble."[485] Who was right?

COFEA suggests that the phrase *keep arms* often, though not primarily, referred to private, non-militia use of firearms. And far more often, the phrase *keep arms* referred to private *keeping* of firearms—and that keeping may have been used for militia purposes or for personal purposes.

But we are not confident about the results. We found only thirty-nine unique and relevant instances of *keep* within six words of *arms*. This small population does not let us draw broad conclusions about usage in the late eighteenth century. Yet, our dataset of thirty-nine instances is far greater the dataset Justices Scalia and Stevens used in 2008.

Here, Justice Stevens was more right than Justice Scalia about the predominant sense of *keep arms*. But Justice Stevens was completely wrong that the phrase *keep arms* "in no way contradicts the military meaning."[486] Our findings show that the phrase *keep arms* does contradict the military meaning about 20% of the time. Neither the majority nor the dissent performed much analysis of *keep arms*. For different reasons, both Justices Stevens and Scalia score a C here.

D. Claim #4: "Bear Arms"

In *Heller*, Justice Scalia acknowledged that, "at the time of the founding," the phrase *bear arms* had a military connotation, but "it *unequivocally* bore that idiomatic meaning only when followed by the preposition 'against.'"[487] For example, he wrote that the Declaration of Independence charged that the King of England had "constrained our fellow Citizens taken Captive on the high Seas to bear Arms *against* their Country."[488] Justice Scalia added, "[e]very example" provided by the linguistics professors of *bear arms* with a military context "includes the preposition 'against,' or is not clearly idiomatic."[489]

It is true that the word *against* sometimes follows *bear arms*. In our sample, we found that construction in 9% of the results. We think Justice Scalia may have been led astray by this fact. He identified several instances where the phrase *bear arms*, absent the preposition *against*, had a non-military context. And he identified many more instances where the phrase *bear arms*, with *against*, had a military

485. *Id.* at 650 (Stevens, J., dissenting).
486. *Id.*
487. *Id.* at 586 (majority opinion).
488. *Id.* (emphasis added).
489. *Id.*

context. From this small survey he drew an inference: *bear arms*, without *against*, generally referred to an individual right.

A collocate search provides some support for Justice Scalia's position. We performed a query for the collocates that appear within four words to the left and right of the phrase *bear arms*. The six most frequent collocates are *to*, *the*, *and*, *of*, *not*, and *be*. These common words are collocates of nearly every word due to the nature of English grammar and are referred to as *stop words*.[490] Many corpora allow researchers to exclude these stop words in a collocate search. But the seventh most common collocate for this query is *against*.

Justice Scalia's inference, however, was misplaced. In 2008, he did not accurately represent the evidence presented by the linguistics professors.[491] And now COFEA proves, with a high degree of certainty, that Justice Scalia's conclusion about *bear arms against* was in error. The phrase *bear arms*, with *against*, is sufficient to make the phrase military. But the phrase *bear arms*, without *against*, can still invoke the military sense. In other words, *against* is a sufficient but not a necessary modifier to move a phrase into the military sense category.

The fact that most references to bearing arms in late eighteenth-century writings were in the military context likely means that the military-only sense of *bear arms* was more common than an individual sense. Or the data may show *only* that, when people used the phrase *bear arms*, they did so most frequently in a military setting. Justice Scalia seems to adopt the latter explanation. He wrote that it was not surprising that founding-era documents from this period, which referenced war and the militia, would show more instances of a militia as opposed to an individual sense of bearing arms.[492] But if we operationalize the original public meaning to correspond to the dominant sense, then, in general use, the dominant sense of *bear arms* is its militia/collective meaning.

How do the Justices stack up? Justice Stevens contended *bear arms* was "a familiar idiom" having a military meaning.[493] Here, Justice Stevens's linguistic analysis gets a B+. He missed an A for two reasons. First, he partially relied on etymology, which is a poor indicator of contemporaneous meaning.[494] Second, he seemed to take

490. *See* JURE LESKOVEC ET AL., MINING OF MASSIVE DATASETS 9 (3d ed. 2020) ("The most frequent words will most surely be the common words such as 'the' or 'and,' which help build ideas but do not carry any significance themselves. In fact, the several hundred most common words in English (called *stop words*) are often removed from documents before any attempt to classify them.").

491. *See supra* notes 140–46 and accompanying text.

492. *See Heller*, 554 U.S. at 587–88.

493. *Id.* at 646 (Stevens, J., dissenting).

494. *See id.* at 646 (Stevens, J., dissenting); *see also* Lee & Phillips, *supra* note 90, at 288; *supra* notes 176–77 and accompanying text.

the position that *bear arms* only takes on an idiomatic military meaning, which is not true. Justice Scalia, on the other hand, barely ekes out a D. He recognized that *bear arms* could have an idiomatic military meaning, but he relied on an incorrect argument about the meaning of *bear arms against*. That mistake caused him to reject the overwhelming weight of the evidence.

E. Grading the Operative Clause

We recognize that corpus linguistics may have a limited utility for certain constitutional questions. And this utility can be conceptualized as a function of two factors. First, how textually (or grammatically) simple or complex is the interpretive inquiry? In the academic literature, corpus linguistics has been most useful when the linguistic inquiry is straightforward; for example, determining which of two competing senses of a word or phrase is the most common one *in the appropriate context*. We find searches that simply count the number of senses that appear in all possible contexts to be less useful. Second, is the inquiry into meaning more linguistic or behavioral in nature? In some cases, the scope of a constitutional phrase can be determined based solely on the meaning of words. In other cases, the meaning of those words incorporates shared understandings of behavior by the relevant actors.

Consider these two factors in practice. The Commerce Clause is textually simple. Most of the debate over its meaning has hinged solely on the linguistic meaning of the word "commerce."[495] There is seemingly no relevant behavioral inquiry to define "commerce."[496] For this inquiry, the utility of corpus linguistics is high. By contrast, determining the meaning of "judicial power" in Article III with corpus linguistics is not so straightforward.[497] This phrase is not a simple linguistic term that one could turn to a dictionary to fully understand. Rather, the exercise of the "judicial power" reflects an inquiry into founding-era, colonial, and English judicial behavior.[498] In that context, the utility of corpus linguistics is low.

---

495. *See generally, e.g.*, Randy E. Barnett, *New Evidence of the Original Meaning of the Commerce Clause*, 55 ARK. L. REV. 847 (2003).

496. *See, e.g.*, Lee & Phillips, *supra* note 90, at 282 n.87 (explaining that the authors did not conduct a behavioral inquiry when analyzing "commerce" and "public use" because "such evidence would have only indirect—circumstantial—evidentiary significance").

497. *See* Randy E. Barnett, *The Original Meaning of the Judicial Power*, 12 SUP. CT. ECON. REV. 115, 120 (2004) ("Original meaning interpretation is not, however, always sufficient to yield a rule of law to apply to a case or controversy.").

498. *See generally, e.g.*, William N. Eskridge, Jr., *All About Words: Early Understandings of the "Judicial Power" in Statutory Interpretation, 1776–1806*,

The Second Amendment lies in the middle of this spectrum. The Second Amendment is textually complex: it has four verbs and four subjects, and all are linked by a complex grammatical structure.[499] We think, therefore, that corpus linguistics has moderate utility for understanding the Second Amendment.

In light of the linguistic complexity of the Second Amendment, we will consider two approaches. First, we consider a model that we dubbed the "Summation Approach." When dealing with a complex grammatical constitutional phrase, we can add up the various meanings of each word or phrase to determine an overall meaning. Stated differently, each phrase or word is initially viewed in isolation from the others. Then, we add up the meaning of (1) *the right of the people* (majority "collective" sense), (2) *keep arms* (no clear majority sense), and (3) *bear arms* (majority "collective" sense). Through this approach, we would likely conclude that the operative clause only protected a collective sense.

We also consider a second approach, which we refer to as the "Interactive Approach." With this method, each word or phrase will be influenced by the meanings of the surrounding constitutional text, regardless of its "independent" meaning in the greater corpus. To follow this method, we will arrange the different parts of the Second Amendment in different combinations—as if we were assembling constitutional text like a Lego kit. Consider six counterfactuals, presented below, in light of our corpus linguistics queries.

1. "The right to bear Arms shall not be infringed."

The right at issue is the "right to bear Arms." We know that the phrase *right* in the context of *arms* was slightly more likely to refer to a "collective" right than to an "individual" right. But there was a sizeable number of instances where rights, in the context of arms, could refer to both a collective right and an individual right. How do we break this logjam? The word *bear* serves that role. We know that the phrase *bear arms* overwhelmingly referred to a collective right.[500] The provision uses *right*, followed by *bear arms*. We can be fairly confident that this provision secures some sort of collective right. Here, we assume that the dominant sense in the relevant context is the original public meaning rather than a reflection of the type of documents in the corpus. And we also assume that the dominant sense is not based on the frequency of certain facts in the real world (such as war) driving the corpus frequency numbers.

101 COLUM. L. REV. 990 (reviewing the approach of both English judges and founding-era judges to determine the meaning of "judicial power").

499. *See supra* Part IV.

500. *See supra* Part IV.E.

2. "The right to keep Arms shall not be infringed."

Once again, the word *right* in the context of *arms* could be collective or individual, but there was no majority sense. However, now the verb is different. Instead of *bear Arms*, the provision refers to *keep Arms*, and there is no majority sense for *keep Arms*. Moreover, we found that the phrase *keep Arms* could refer to collective-keep/collective-use, individual-keep/collective-use, or individual-keep/individual-use. There is no obvious way to resolve this logjam with the linguistic data we found. Because this provision could reasonably refer to either a collective right or an individual right, it would likely have been understood to cover both types of rights.

3. "The right to keep and bear Arms shall not be infringed."

This right now includes two verbs: *keep* and *bear*. COFEA and COEME tells us that *keep and bear* was not a single linguistic unit. This text could be restated as "The rights (plural) to keep and to bear Arms shall not be infringed." Or it could be restated as, "The right to keep Arms and the right to bear Arms shall not be infringed." Given this construction, we must consider the phrase *bear arms* (overwhelmingly collective) and *keep arms* (no majority sense), in the context of rights. We think this provision is arguably more likely to refer to a collective right than an individual right, but not by much. It could go either way, and it is difficult to pick a dominant reading.

4. "The right of the people to bear Arms shall not be infringed."

Now, we have added the phrase *right of the people*. This phrase had a slight majority "collective" sense, about 60%. If we combined the overwhelmingly collective sense of *bear arms* and the slightly collective sense of *right of the people*, then we are left with a strong indication this provision secures a collective right.

5. "The right of the people to keep Arms shall not be infringed."

This provision combines *right of the people* and *keep arms*. Version #5, more than version #4, is more likely to refer to an "individual" right. But it could still go either way.

6. "The right of the people to keep and bear Arms shall not be infringed."

Version #6 is the ratified version of the operative clause. It combines *right of the people* with *keep* and *bear*. Consider two scenarios. First, if *right of the people*, *keep*, and *bear* each implicate a "collective" sense, then on balance, the operative clause refers to a collective right. Second, if *right of the people* and *keep* implicate an "individual" sense—we presume that *bear* will only implicate a collective sense—then the operative clause could, on balance, refer to an "individual" sense. Here, different parts of the operative clause

pull at each other between the "collective" and the "individual" senses, as if the Framers created a perpetual tug-of-war. The first reading is more likely than the second reading, but we cannot rule out the possibility that the second reading is reasonable.

***

The Summation Approach is intuitive enough, but we find it problematic. This approach seems to violate the principle of non-compositionality. As we discussed earlier,[501] this principle holds that a phrase is not always the linguistic sum of its parts—the very type of grammatical math that the Summation Approach attempts.

The Interactive Approach may better reflect how people would have read and understood the Second Amendment. With this second model, we can consider two factors: (1) the lack of a clear majority sense for *keep arms* and (2) the similar lack of a clear majority sense in the most relevant context of arms and rights. And these factors would, in turn, influence how to read *the right of the people* and *bear arms*, both of which can have a "collective" sense and an "individual" sense. Under the Interactive Approach, the operative clause is more likely to cover both senses.

Because we find the Interactive Approach more useful than the Summation Approach, our best sense of the original public meaning of the Second Amendment, based solely on this linguistic data, is that it covered both a collective and an individual right to keep and bear arms. Ultimately, given the mixed utility of corpus linguistic analysis, multiple methods are likely necessary to determine the meaning of the Second Amendment through "thick originalism."[502] Again, this Article does not consider other historical methodologies to inform the original public meaning of the Second Amendment. Here, we limit our inquiry to corpus linguistics.

CONCLUSION

Our corpus linguistic examination of the Second Amendment demonstrates that both Justice Scalia and Justice Stevens made errors in *Heller*. In the past, our preliminary work has been cited to cast doubt on Justice Scalia's majority opinion.[503] Going forward, we

501. *See supra* note 92 and accompanying text.

502. *See generally* John O. McGinnis, *Thick Originalism as a Constraint on Ideology*, LAW & LIBERTY (Apr. 7, 2014), https://lawliberty.org/thick-orginalism-as-a-constraint-on-ideology/.

503. *See e.g.*, State v. Misch, No. 2019-266, 2021 WL 650366, *6 (Vt. Feb. 19, 2021) (citing Josh Blackman & James C. Phillips, *Corpus Linguistics and the Second Amendment*, HARV. L. REV. BLOG (Aug. 7, 2018), https://blog.harvardlawreview.org/corpus-linguistics-and-the-second-amendment/); *see also* Darrell A. H. Miller, *Owning* Heller, 30 U. FLA. J.L. & PUB. POL'Y 153, 161 (2020) ("To their credit, individuals usually characterized as more gun-rights leaning

hope critics of Justice Scalia also acknowledge our critiques of Justice Stevens. Scholars, advocates, and judges alike have a tendency to only focus on the errors made by the side they oppose. But Justice Scalia's critics need to acknowledge that Justice Stevens's analysis was also not immune from mistake.

Finally, our research highlights some of the limitations of applying corpus linguistics to constitutional interpretation: Like all data-driven inquiries, sometimes the data do not provide clear answers, and sometimes there is not enough data to have full confidence in the answers. But like an opinion poll that shows a fifty-fifty tie for a presidential race, we would much rather have some data than be left in the dark. And there is a little less darkness on the question of what the Second Amendment means after our analysis, even if there may not yet be a crystal-clear answer on its original public meaning.

concede the new evidence places *Heller*'s historical justification in doubt. Professors Josh Blackman and James C. Phillips have conducted their own research on the corpus and have found that 'the overwhelming majority of instances of "bear arms" was in the military context.'").