

4-1-2020

## Foreword: Lawyers and Linguists Collaborate in Using Corpus Linguistics to Produce New Insights Into Original Meaning

Clark D. Cunningham

Georgia State University College of Law, [cdcunningham@gsu.edu](mailto:cdcunningham@gsu.edu)

Follow this and additional works at: <https://readingroom.law.gsu.edu/gsulr>



Part of the [Law Commons](#)

---

### Recommended Citation

Clark D. Cunningham, *Foreword: Lawyers and Linguists Collaborate in Using Corpus Linguistics to Produce New Insights Into Original Meaning*, 36 GA. ST. U. L. REV. vi (2020).

Available at: <https://readingroom.law.gsu.edu/gsulr/vol36/iss5/3>

This Frontmatter is brought to you for free and open access by the Publications at Reading Room. It has been accepted for inclusion in Georgia State University Law Review by an authorized editor of Reading Room. For more information, please contact [gfowke@gsu.edu](mailto:gfowke@gsu.edu).

## FOREWORD: LAWYERS AND LINGUISTS COLLABORATE IN USING CORPUS LINGUISTICS TO PRODUCE NEW INSIGHTS INTO ORIGINAL MEANING

Clark D. Cunningham\*

On August 23, 2019, the Supreme Court of Idaho published a decision that sent an explicit signal to lawyers in its jurisdiction: “We . . . reference the use of corpus linguistic tools as a support for our analysis . . . and **as an motivation for counsel** to consider this . . . tool . . . for statutory interpretation . . . when called for in the future.”<sup>1</sup> The Idaho decision occupied the midpoint in a forty-five-day period during which two other appellate courts also explicitly used research based on “corpus linguistics” to support statutory interpretation decisions. The first of these cases was decided by the United States Court of Appeals for the Third Circuit on August 1, 2019,<sup>2</sup> and the third decision issued in this short period was published by the Supreme Court of Utah on September 13, 2019.<sup>3</sup>

At the end of May 2019, a three-judge panel of the United States Court of Appeals for the Sixth Circuit went a step further than the “motivation” given by the Idaho Supreme Court. The panel explicitly directed the United States Department of Justice and the appointed counsel in a federal prisoner habeas case to file supplemental briefs addressing “how does” a specific linguistic corpus (the *Corpus of Founding Era American English*) “help inform” a determination “of the original meaning of the Article III Cases or Controversies requirement.”<sup>4</sup> Although the court in that case did not end up using

---

\* Professor of Law and W. Lee Burge Chair in Law & Ethics, Georgia State University College of Law, cdcunningham@gsu.edu, www.clarkcunningham.org.

1. *State v. Lantis*, 447 P.2d 875, 880 (Idaho 2019) (emphasis added) (internal quotation marks and citation omitted) (using data from *Corpus of Historical American English* regarding use of “disturbing the peace” in 1887).

2. *Caesars Entm’t Corp. v. Int’l Union of Operating Eng’rs*, 932 F.3d 91, 95 (3d Cir. 2019) (using data from *Corpus of Historical American English* regarding use of “previously”).

3. *Richards v. Cox*, 450 P.3d 1074, 1079–81 (Utah 2019) (using *Corpus of Contemporary American English* and *Corpus of Historical American English* to investigate the meaning of “employment” in the Utah Constitution). For further citations to published judicial opinions specifically discussing corpus-based research, see *Resources on Law & Linguistics*, CLARKCUNNINGHAM.ORG, <http://www.clarkcunningham.org/Law-Linguistics.html> [https://perma.cc/U666-KSRT] (last visited Mar. 8, 2019).

4. Letter to Counsel at 1, *Wright v. Spaulding*, 939 F.3d 695 (6th Cir. 2019) (No. 17-4257).

the linguistic research that it received,<sup>5</sup> a member of that panel wrote a long concurring opinion in another case decided in July 2019, explaining that “corpus linguistics is a powerful tool for discerning how the public would have understood a statute’s text at the time it was enacted.”<sup>6</sup>

Corpus linguistics has not only recently attracted judicial attention but also has been generating considerable legal scholarship. Over twenty-five law review articles have been written on the subject since 2016,<sup>7</sup> including, in the past two years, several published in some of the country’s leading law reviews.<sup>8</sup>

“Corpus linguistics” is not, as recently suggested by a tongue-in-cheek newspaper column, the name of “a singer in a punk band.”<sup>9</sup> A digitized data set representing actual language—typically a very large data set—is called by the science of linguistics a “corpus” (plural: corpora).<sup>10</sup> “Corpus linguistics” is a short-hand term applied to a wide variety of ways linguists use computer technology to acquire, store, and process such data and then analyze it to document and describe patterns of natural language usage, often to solve real-world problems.<sup>11</sup> The recent virtual explosion of interest in using linguistic analysis of corpus data for legal interpretation can be traced back<sup>12</sup> to a 2010 law review student note by Stephen Mouritsen<sup>13</sup> and a 2011 concurring opinion by Utah Supreme Court

---

5. See Haoshan Ren et al., “Questions Involving National Peace and Harmony” or “Injured Plaintiff Litigation”? *The Original Meaning of “Cases” in Article III of the Constitution*, *infra* 36 GA. ST. U. L. REV. 535, 547–48 (2020).

6. *Wilson v. Safelite Grp., Inc.*, 930 F.3d 429, 438–45 (6th Cir. 2019) (Thapar, J., concurring).

7. For citations, see *Resources on Law & Linguistics*, *supra* note 3.

8. See Thomas R. Lee & Stephen C. Mouritsen, *Judging Ordinary Meaning*, 127 YALE L.J. 788 (2018); Thomas R. Lee & James C. Phillips, *Data-Driven Originalism*, 167 U. PA. L. REV. 261 (2019); Jennifer L. Mascott, *Who Are “Officers of the United States”?*, 70 STAN. L. REV. 443 (2018); Kevin P. Tobia, *Testing Ordinary Meaning: An Experimental Assessment of What Dictionary Definitions and Linguistic Usage Data Tell Legal Interpreters*, 133 HARV. L. REV. (forthcoming 2020).

9. Scott Martelle, Opinion, *Word Nerds Take on ‘Emoluments,’ and Deliver a Verdict Sure to Anger Trump*, L.A. TIMES (Jan. 30, 2019, 1:05 PM), <https://www.latimes.com/opinion/enterthefray/la-ol-trump-emoluments-constitution-20190130-story.html> [<https://perma.cc/5F3N-J2CV>].

10. Clark D. Cunningham & Jesse Egbert, *Using Empirical Data to Investigate the Original Meaning of “Emolument” in the Constitution*, *infra* 36 GA. ST. U. L. REV. 465, 473 (2020).

11. *Id.*

12. Clark D. Cunningham & Jesse Egbert, *Corpora and Analyzing Legal Discourse in the United States*, in ROUTLEDGE HANDBOOK OF CORPUS APPROACHES TO DISCOURSE ANALYSIS (Eric Friginal & Jack Hardy eds., forthcoming 2020), pre-publication draft available at [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3554023](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3554023).

13. Stephen C. Mouritsen, *The Dictionary Is Not a Fortress: Definitional Fallacies and a Corpus-Based Approach to Plain Meaning*, 2010 BYU L. REV. 1915 (2010).

Justice Thomas Rex Lee,<sup>14</sup> written while Mouritsen was clerking for Justice Lee.

To my knowledge, with the exception of that 2011 concurrence by Justice Lee, none of the judicial opinions published referencing linguistic analysis of corpus data were developed with the assistance of someone with formal training in linguistics. Likewise, with a few notable exceptions, the many law review articles discussing corpus linguistics published prior to 2020 do not include an author who holds a Ph.D. in linguistics or an academic post in that field.<sup>15</sup> To promote and demonstrate the value of law-linguistic collaboration in applying corpus linguistics to legal interpretation, the College of Law and the Department of Applied Linguistics and English as a Second Language at Georgia State University jointly sponsored a full-day workshop on October 18, 2019.<sup>16</sup> Four of the five papers presented at that workshop were coauthored by one or more internationally-known linguists,<sup>17</sup> and the two lawyers who coauthored the fifth paper received mentoring from one of the

---

14. *In re Adoption of Baby E.Z.*, 266 P.3d 702, 715–32 (Utah 2011) (Lee, J., concurring).

15. Exceptions include Edward Finegan, *Comments on 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. 1297 (2017); Stefan Th. Gries & Brian G. Slocum, *Ordinary Meaning and Corpus Linguistics*, 2017 BYU L. REV. 1417 (2017); 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); Lawrence M. Solan, *Can Corpus Linguistics Help Make Originalism Scientific?*, 126 YALE L.J.F. 57 (2016); Lawrence M. Solan, *Patterns in Language and Law*, 6 INT'L J. LANGUAGE & L. 46 (2017); and Lawrence M. Solan & Tammy Gales, *Corpus Linguistics as a Tool in Legal Interpretation*, 2017 BYU L. REV. 1311 (2018).

16. Video of the workshop as well as PowerPoints presented by authors and commentators are available at *Workshop on Law & Linguistics*, CLARKCUNNINGHAM.ORG, <http://www.clarkcunningham.org/Workshop-Law-Linguistics.html> [https://perma.cc/26GK-Z9J4] (last visited Mar. 8, 2019) [hereinafter *Workshop Website*]. The workshop was supported by grants from the Faculty Development Committee of the Georgia State University College of Law and from the Office of the Provost, Georgia State University, through the Study in a Second Discipline fellowship program.

17. Jesse Egbert, Associate Professor of Applied Linguistics, Northern Arizona University, coauthored *Using Empirical Data to Investigate the Original Meaning of “Emolument” in the Constitution* and *“Questions Involving National Peace and Harmony” or “Injured Plaintiff Litigation”?* *The Original Meaning of “Cases” in Article III of the Constitution*; Ute Römer, Associate Professor of Applied Linguistics and English as a Second Language, Georgia State University, also coauthored *Original Meaning of Cases* and coauthored *“We the Citizens?”: A Corpus Linguistic Inquiry into the Use of “People” and “Citizens” in the Founding Era*; and Tammy Gales, Associate Professor of Comparative Literature, Languages, and Linguistics, Hofstra University, coauthored *Revisiting a Classic Problem in Statutory Interpretation: Is a Minister a Laborer?* with Lawrence Solan, a law professor who also holds a Ph.D. in linguistics. For information about the academic accomplishments of these linguists, see Author Biographies, *infra*.

founding figures in the field.<sup>18</sup> The five papers presented at this workshop were then developed into the five articles comprising this Special Issue on Law and Linguistics. The United States Constitution prohibits federal officials from receiving any “present, Emolument, Office or Title” from a foreign state without the consent of Congress.<sup>19</sup> However, in trying to determine the meaning of *emolument*, courts are confronted with a term that might as well be a foreign word from an unknown language because the word *emolument* has virtually vanished from contemporary American English. At the time of writing, cases are pending in three federal circuits alleging that President Donald Trump is violating the Constitution by receiving business-based revenue generated in part from payments by foreign states. These cases prompted Jesse Egbert and me to undertake a corpus linguistics research project that asked: “Is there evidence that Americans in the Founding Era could have used the word *emolument* to describe revenue derived from ownership of a hotel?” Our analysis of data from the *Corpus of Founding Era American English*, reported in this issue’s first article, *Using Empirical Data to Investigate the Original Meaning of “Emolument” in the Constitution*,<sup>20</sup> found that *emolument* was frequently used with modifying adjectives and prepositional phrases, indicating that eighteenth-century writers often needed to constrain or specify the word’s broad meaning. We further discovered that *emolument* often functioned as a catch-all term of inclusion and that its use in lists ending “and other emoluments” produced evidence that at least twenty-five different nouns could be described as a type of emolument. We conclude that Founding Era Americans could indeed have used *emolument* to describe revenue derived from ownership of a hotel and cite several actual examples where *emolument* was used to refer to revenue from ownership interest in a business.

In this issue’s second article, *Revisiting a Classic Problem in Statutory Interpretation: Is a Minister a Laborer?*,<sup>21</sup> Tammy Gales and Lawrence Solan present a new analysis of an iconic 1892 U.S.

---

18. During the research and writing process, Edward Finegan, Professor of Linguistics and Law, Emeritus, University of Southern California, reviewed and commented on Eleanor Miller & Heather Obelgoner, *Effective but Limited: A Corpus Linguistic Analysis of the Original Public Meaning of Executive Power*, *infra* 36 GA. ST. U. L. REV. 605 (2020).

19. U.S. CONST. art. I, § 9, cl. 8.

20. Cunningham & Egbert, *supra* note 10.

21. Tammy Gales & Lawrence M. Solan, *Revisiting a Classic Problem in Statutory Interpretation: Is a Minister a Laborer?*, *infra* 36 GA. ST. U. L. REV. 491 (2020).

Supreme Court case—*Holy Trinity Church v. United States*.<sup>22</sup> The question in *Holy Trinity Church* concerned whether a law making it illegal to pay the transportation of a person entering the U.S. under contract to perform “labor or service of any kind” applied to a Manhattan church that had paid to bring its new rector from England to New York. The Supreme Court unanimously ruled that the law did not apply to the church’s contract, relying first on the ordinary meaning of “labor” and second on the legislative history of the phrase “labor or service.” Highlighting the use of corpus linguistic methods in corpora of general and statutory language, the study by Gales and Solan demonstrates that the phrase “labor or service” did appear in the nineteenth century to be a legal term of art with a narrow interpretation that would exclude clergy; but, within a few decades of the *Holy Trinity* decision, the phrase began to be applied to contexts with broader meaning: first, in the general corpus, and, eventually, in the statutory corpus. Such changes were marked by changes in pluralization and modification to the terms in question. Additionally, when examining “labor” as an independent term, Gales and Solan found that those who “labored” were generally not clergy and the activities of clergy were typically not described as “labor.”

If a federal official is deliberately violating the Constitution, is it possible that no federal court has the power to halt that conduct? Federal judges have been answering “yes” for more than a century—dismissing certain kinds of lawsuits alleging unconstitutional conduct by ruling that the lawsuits were not “cases” as meant in the phrase “The judicial Power shall extend to all cases”<sup>23</sup> in Article III, Section Two of the Constitution. The empirical research reported in this issue’s third article, “*Questions Involving National Peace and Harmony*” or “*Injured Plaintiff Litigation*”? *The Original Meaning of “Cases” in Article III of the Constitution*,<sup>24</sup> suggests that the Court’s current interpretation of the meaning of “cases” may be more narrow—perhaps indeed entirely different—than how the word in its Article III context would have been used and understood by those who drafted and ratified the Constitution. The authors—a seven-member team comprised of five linguists, myself, and a practicing lawyer—conclude that (1) during the drafting process the

---

22. 144 U.S. 457 (1892).

23. U.S. CONST. art. III, § 2, cl. 1.

24. Ren et al., *supra* note 5. For a thoughtful review and critique by Susan Smelcer, a Georgia State law professor specializing in legal analytics, see video and PowerPoint at *Workshop Website*, *supra* note 16.

Article III phrase “cases arising under laws” was recurrently understood as one type of a general and expansive jurisdiction described in earlier drafts of the Constitution as questions involving the national peace and harmony; and (2) “cases” as used in various parts in Article III did not have a stable, inherent meaning such as “litigation in court”—instead “cases” in each different Article III context would have been read as having a different meaning, constructed through its combination with accompanying words.

Article I of the Constitution begins by providing that “[a]ll legislative Powers **herein granted** shall be vested in a Congress of the United States,”<sup>25</sup> whereas Article II simply states “[t]he executive Power shall be vested in a President of the United States.”<sup>26</sup> As Eleanor Miller and Heather Obelgoner point out in the fourth article of this issue, *Effective but Limited: A Corpus Linguistic Analysis of the Original Public Meaning of Executive Power*,<sup>27</sup> advocates of a strong presidency have pointed to the different phrasing of Articles I and II to argue that, unlike congressional authority, which is expressly limited to the powers specifically enumerated in Article I, presidential authority extends to the full reach of whatever is meant by “the executive power.” Miller and Obelgoner combine corpus linguistics and historical research to marshal evidence that the original meaning of the “executive power” was probably much more limited than contemporary conceptions. For example, modern descriptions of the President as the “Chief Executive Officer” suggest that the President is “in charge” of the country in the same way that a corporate CEO runs a company. However, while their study of data from the *Corpus of Contemporary American English* shows that 96% of uses of “chief executive officer” in a random sample of modern English referred to the leader of a business enterprise, the same phrase is **never** found in the *Corpus of Founding Era American English* to refer to a business leader but only referred to a leader of a governmental body. Thus, it seems unlikely that those who drafted and ratified the Constitution would have thought of “the executive power” vested in the President as comparable to the authority exercised by a business executive over a private company. Even

---

25. U.S. CONST. art. I, § 1 (emphasis added).

26. U.S. CONST. art. II, § 1, cl. 1 (emphasis added).

27. Miller & Obelgoner, *supra* note 18. See commentary on the working paper version of this article by Michigan law professor Julian Davis Mortenson, who has written extensively on the meaning of “executive power,” and Edward Finegan, Professor of Linguistics and Law, Emeritus, University of Southern California at *Workshop Website*, *supra* note 16.

more fundamentally, their linguistic and historical research converge on the finding that “executive power” in the Founding Era likely had only the limited meaning of “power to execute laws.”

The final article in this issue, “*We the Citizens?*”: *A Corpus Linguistic Inquiry into the Use of “People” and “Citizens” in the Founding Era*,<sup>28</sup> began as a graduate research project by Abigail Stout and Diana Coetzee using corpus linguistics to study the original meaning of “powers . . . reserved . . . to the people”<sup>29</sup> in the Tenth Amendment. As they expanded their research into this article and were joined by linguistics professor Ute Römer, the interdisciplinary team became intrigued by the prevalence of the phrase “the people” in the Constitution in contrast to “the citizens.” The Constitution famously begins with the words “We the People,” and while the First, Second, Fourth, Ninth, and Tenth Amendment all refer to “the people,” the word *citizen* appears nowhere in the Bill of Rights. The team specifically examined both collocate words and fixed phrases found surrounding the words *people* and *citizens*, using both the *Corpus of Founding Era American English* and a corpus consisting of the public papers of James Madison. This study allowed them to discover more about the “roles” that each party (i.e., people and citizens) played and the kinds of actions each party was associated with. They found that although the verbs *elected*, *chosen*, and *made* were used with both words (e.g., “elected by the people/citizens,” “chosen by the people/citizens,” and “made by the people/citizens”), these verbs were much more commonly combined with *people* than with *citizens*. Their findings thus suggest that (1) *people* predominates over *citizens* when referring to governing in the Founding Era; (2) the actions of people—*electing*, *choosing*, and *making*—were wider in their scope than those of citizens; and, ultimately, (3) *the people* were understood as the ground-up source of power for the establishment of the Constitution.

---

28. Abigail Stout, Diana Coetzee & Ute Römer, “*We the Citizens?*”: *A Corpus Linguistic Inquiry into the Use of “People” and “Citizens” in the Founding Era*, *infra* 36 GA. ST. U. L. REV. 663 (2020).

29. U.S. CONST. amend. X.