

Michigan Law Review

Volume 116 | Issue 2

2017

Corpus Linguistics: Misfire or More Ammo for the Ordinary - Meaning Canon?

John D. Ramer
University of Michigan Law School

Follow this and additional works at: <https://repository.law.umich.edu/mlr>



Part of the [Courts Commons](#), [Judges Commons](#), [Legislation Commons](#), and the [State and Local Government Law Commons](#)

Recommended Citation

John D. Ramer, *Corpus Linguistics: Misfire or More Ammo for the Ordinary - Meaning Canon?*, 116 MICH. L. REV. 303 (2017).

Available at: <https://repository.law.umich.edu/mlr/vol116/iss2/3>

<https://doi.org/10.36644/mlr.116.2.corpus>

This Note is brought to you for free and open access by the Michigan Law Review at University of Michigan Law School Scholarship Repository. It has been accepted for inclusion in Michigan Law Review by an authorized editor of University of Michigan Law School Scholarship Repository. For more information, please contact mlaw.repository@umich.edu.

NOTE

CORPUS LINGUISTICS: MISFIRE OR MORE AMMO FOR THE ORDINARY-MEANING CANON?

John D. Ramer*

Scholars and judges have heralded corpus linguistics—the study of language through collections of spoken or written texts—as a novel tool for statutory interpretation that will help provide an answer in the occasionally ambiguous search for “ordinary meaning” using dictionaries. In the spring of 2016, the Michigan Supreme Court became the first to use corpus linguistics in a majority opinion. The dissent also used it, however, and the two opinions reached different conclusions. In the first true test for corpus linguistics, the answer seemed to be just as ambiguous as before.

This result calls into question the utility of corpus linguistics. If the Michigan Supreme Court could reach opposite conclusions about the ordinary meaning of a word when both opinions used the same database, can one really consider corpus linguistics an effective aid to statutory interpretation? This Note first argues that the Michigan Supreme Court’s majority opinion conducted a flawed search of the database and that the dissent’s contrary conclusion was correct.

In addition, this Note contends that transparency—the greatest strength of corpus linguistics—outweighs the risk that judges may fail to use a corpus-linguistics database accurately because that transparency permits advocates, other judges, and legal scholars to review the court’s analysis, aiding appellate review. Lastly, this Note includes recommended steps to guide the use of a corpus-linguistics database in the search for ordinary meaning.

TABLE OF CONTENTS

INTRODUCTION	304
I. THE BACKGROUND OF CORPUS LINGUISTICS IN THE LAW	309
A. <i>The Legacy of Corpus Linguistics in Statutory Interpretation</i>	309
B. <i>The Michigan Supreme Court’s Use of the COCA</i>	311
1. Factual Background	311
2. The Majority Opinion’s Use of the COCA	312
3. The Dissent’s Use of the COCA	314
II. REVIEWING THE MICHIGAN SUPREME COURT’S COCA ANALYSIS	316

* J.D. Candidate, December 2017, University of Michigan Law School. I would first like to thank my wife Whitney for her love, patience, and encouragement during the experience that is law school. I would also like to thank my parents, James and Jane, and my sister Libby for their unwavering support. Finally, I want to thank everyone at the *Michigan Law Review* for their excellent editing.

A.	<i>Search Methods</i>	316
B.	<i>Contextual Review</i>	317
III.	THE COCA AFTER HARRIS	321
A.	<i>Criticisms of the COCA</i>	321
B.	<i>Responses to the Criticism</i>	321
1.	Experts	322
2.	<i>Sua Sponte Use</i>	323
C.	<i>The Value of Transparency</i>	326
D.	<i>Steps to Use the COCA</i>	326
CONCLUSION	328

INTRODUCTION

The judiciary says what a statute means¹ in order “to give effect to the law” a legislature enacted.² To this end, courts “read the statute according to its text.”³ If the legislature did not define a particular word, courts generally look to the ordinary meaning of the word used⁴ at the time the legislature adopted the statute.⁵ This interpretive method—commonly called the “ordinary-meaning canon”⁶—assumes that a word’s ordinary meaning most “accurately expresses the legislative purpose.”⁷

This inquiry sounds simple enough: courts merely ask how a word is (or was) ordinarily understood. As one professor put it, the ordinary meaning is “that which an ordinary speaker of the English language—twin sibling to the common law’s reasonable person—would draw from the statutory text.”⁸ But putting the theory into practice can be difficult.

1. *Rivers v. Roadway Express, Inc.*, 511 U.S. 298, 312–13 (1994) (“A judicial construction of a statute is an authoritative statement of what the statute meant before as well as after the decision of the case giving rise to that construction.”); *Bank of Hamilton v. Lessee of Dudley*, 27 U.S. (2 Pet.) 492, 524 (1829) (“The judicial department of every government is the rightful expositor of its laws . . .”).

2. *Lewis v. City of Chicago*, 560 U.S. 205, 217 (2010).

3. *Hui v. Castaneda*, 559 U.S. 799, 812 (2010); *see Levin v. United States*, 133 S. Ct. 1224, 1231 (2013).

4. *E.g.*, *Asgrow Seed Co. v. Winterboer*, 513 U.S. 179, 187 (1995) (“When terms used in a statute are undefined, we give them their ordinary meaning.”).

5. *See, e.g.*, *Baker Botts L.L.P. v. ASARCO LLC*, 135 S. Ct. 2158, 2165 & n.2 (2015) (using dictionaries from 1933 and 1934 to determine the meaning of “services” because Congress added the relevant language to the federal bankruptcy law in 1934); *Cuomo v. Clearing House Ass’n*, 557 U.S. 519, 526 (2009) (determining the meaning of “visitation” in 1864, the year Congress enacted the National Bank Act). *Contra* William N. Eskridge, Jr., *Dynamic Statutory Interpretation*, 135 U. PA. L. REV. 1479, 1479 (1987) (“Statutes, however, should—like the Constitution and the common law—be interpreted ‘dynamically,’ that is, in light of their present societal, political, and legal context.”).

6. *E.g.*, ANTONIN SCALIA & BRYAN A. GARNER, *READING LAW: THE INTERPRETATION OF LEGAL TEXTS* 69 (2012).

7. *Gross v. FBL Fin. Servs. Inc.*, 557 U.S. 167, 175 (2009) (quoting *Engine Mfrs. Ass’n v. S. Coast Air Quality Mgmt. Dist.*, 541 U.S. 246, 252 (2004)).

8. William N. Eskridge, Jr., *Textualism, the Unknown Ideal?*, 96 MICH. L. REV. 1509, 1511 (1998) (reviewing ANTONIN SCALIA, *A MATTER OF INTERPRETATION: FEDERAL COURTS*

In the last few decades, lawyers and judges have increasingly used dictionaries to determine a word's ordinary meaning.⁹ For example, the ordinary-meaning canon was on display in *Muscarello v. United States*,¹⁰ when the U.S. Supreme Court consolidated two criminal cases.¹¹ In one case, Frank Muscarello was arrested after driving to a drug deal with a handgun locked in his truck's glove compartment.¹² In the other case, which had similar facts, the defendants had stored their guns in the car's trunk.¹³ In both lower court opinions, the courts of appeals found that the defendants had violated 18 U.S.C. § 924(c)(1) because they "'carrie[d]' the guns during and in relation to a drug trafficking offense."¹⁴ The issue was whether the ordinary meaning of the phrase "carries a firearm" included storing handguns in a vehicle's locked compartment.¹⁵

The Supreme Court looked to a number of sources in its attempt to discern the phrase's ordinary meaning. First, the Court looked to the *Oxford English Dictionary*.¹⁶ Then it consulted *Webster's Third New International Dictionary*, the *Random House Dictionary of the English Language Unabridged*, the *Barnhart Dictionary of Etymology*, the *Oxford Dictionary of English Etymology*, the King James Bible, *Robinson Crusoe*, and *Moby Dick*.¹⁷ The Court also cited articles from the *New York Times*, the *Boston Globe*, the *Colorado Springs Gazette Telegraph*, and the *Arkansas Gazette*.¹⁸ After reviewing all these sources (and others), the Court determined that the ordinary meaning of "carrying a firearm" included driving a car with a gun in it.¹⁹ The dissent also conducted an extensive search for the word's ordinary meaning by consulting a legal dictionary, several translations of the Bible, poems, and scripts from the film *The Magnificent Seven* and the television

AND THE LAW (1997)); see also Oliver Wendell Holmes, *The Theory of Legal Interpretation*, 12 HARV. L. REV. 417, 417–18 (1899) ("Thereupon we ask, not what [the author] meant, but what those words would mean in the mouth of a normal speaker of English, using them in the circumstances in which they were used . . .").

9. James J. Brudney & Lawrence Baum, *Oasis or Mirage: The Supreme Court's Thirst for Dictionaries in the Rehnquist and Roberts Eras*, 55 WM. & MARY L. REV. 483, 494–95 (2013) (discussing the Supreme Court's increasing use of dictionaries starting in the Rehnquist era); e.g., *Taniguchi v. Kan Pac. Saipan, Ltd.*, 566 U.S. 560, 566–67 (2012) (describing the issue presented as determining the ordinary meaning of "interpreter" and then consulting ten different dictionaries).

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

11. *Muscarello*, 524 U.S. at 127.

12. *Id.*

13. *Id.*

14. *Id.* (alteration in original) (first quoting *United States v. Muscarello*, 106 F.3d 636 (5th Cir. 1997), and then quoting *United States v. Cleveland*, 106 F.3d 1056 (1st Cir. 1997)).

15. *Id.* at 127–28.

16. *Id.* at 128.

17. *Id.* at 128–29.

18. *Id.* at 129–30.

19. *Id.* at 131.

shows *M*A*S*H* and *Sesame Street*.²⁰ The ordinary-meaning analysis of both opinions seemed somewhat haphazard.

To bring more statistical analysis to the ordinary-meaning canon, then—law student Stephen Mouritsen,²¹ who held a master of arts in linguistics with an emphasis in corpus linguistics,²² thought of a better way. He proposed using corpus linguistics²³—the study of language through collections of spoken or written texts, called corpora.²⁴ These collections typically take the form of digitized databases that are accessible through an online interface.²⁵ A user can search the corpus-linguistics database for a particular word or phrase to study how that word or phrase has actually been used in the texts collected in the corpus.²⁶

One corpus-linguistics database is the Corpus of Contemporary American English (“COCA”), which “is the largest freely-available corpus of English.”²⁷ The COCA contains more than 520 million words and is equally divided among transcriptions of spoken language,²⁸ fiction publications,²⁹

20. Stephen C. Mouritsen, Comment, *The Dictionary Is Not a Fortress: Definitional Fallacies and a Corpus-Based Approach to Plain Meaning*, 2010 BYU L. REV. 1915, 1916–17 (summarizing the dissent’s sources).

21. Mouritsen is now an associate at Cravath, Swaine, & Moore LLP in New York, an adjunct professor at Brigham Young University, and a former law clerk to the Honorable Justice Thomas Rex Lee of the Utah Supreme Court. *Stephen Mouritsen Faculty Biography*, BYU LAW, http://www.law2.byu.edu/faculty/profile_2016.php?id=234 [<https://perma.cc/664H-QFCC>].

22. Mouritsen, *supra* note 20, at 1970 n.*.

23. *Id.* at 1951–70.

24. MARTIN WEISSER, PRACTICAL CORPUS LINGUISTICS: AN INTRODUCTION TO CORPUS-BASED LANGUAGE ANALYSIS 13 (2016).

25. *See id.*

26. *See* HANS LINDQUIST, CORPUS LINGUISTICS AND THE DESCRIPTION OF ENGLISH 1, 5 (2009) (“At the centre of [corpus linguistics] is that the rules of language are usage-based . . . [And] if you are interested in the workings of a particular language, like English, it is a good idea to study English in use.”).

27. CORPUS CONTEMP. AM. ENG., <http://corpus.byu.edu/coca/> [<https://perma.cc/GT9J-9NNR>].

28. *What Kind of Texts Are in the Corpus?*, CORPUS CONTEMP. AM. ENG., <https://corpus.byu.edu/coca/help/texts.asp> [<https://perma.cc/CV5E-2DND>] (“Transcripts of unscripted conversation from more than 150 different TV and radio programs (examples: *All Things Considered* (NPR), *Newshour* (PBS), *Good Morning America* (ABC), *Today Show* (NBC), *60 Minutes* (CBS), *Hannity and Colmes* (Fox), *Jerry Springer*, etc[.]).”).

29. *Id.* (“Short stories and plays from literary magazines, children’s magazines, popular magazines, first chapters of first edition books 1990–present, and movie scripts.”).

popular magazines,³⁰ newspapers,³¹ and academic texts.³² Professor Mark Davies, a corpus linguist, developed the COCA through his academic work at Brigham Young University.³³ The COCA is a “tagged corpus,”³⁴ which means that each word is labeled in reference to its particular part of speech—for example, a noun or a verb.³⁵ The COCA does not, however, function like a search engine. The user cannot simply ask what the word means. Instead, the database allows a user to review many instances of the use of a word or phrase in the database’s collected texts.³⁶

Mouritsen argued that the COCA’s utility in determining ordinary meaning exceeds that of a dictionary³⁷ because dictionaries provide a range of *permissible* meanings only³⁸: “A dictionary, it is vital to observe, never says what meaning a word must bear in a particular context. Nor does it ever purport to say this.”³⁹ Mouritsen described a hypothetical, seemingly modeled after *Muscarello*, where Judge A and Judge B disagree about the ordinary meaning of a word, and both judges find support in different dictionaries.⁴⁰ As Mouritsen explained, assuming no additional evidence, the

30. *Id.* (“Nearly 100 different magazines, with a good mix (overall, and by year) between specific domains (news, health, home and gardening, women, financial, religion, sports, etc.). A few examples are *Time*, *Men’s Health*, *Good Housekeeping*, *Cosmopolitan*, *Fortune*, *Christian Century*, *Sports Illustrated*, etc.”).

31. *Id.* (“Ten newspapers from across the US, including: *USA Today*, *New York Times*, *Atlanta Journal Constitution*, *San Francisco Chronicle*, etc. In most cases, there is a good mix between different sections of the newspaper, such as local news, opinion, sports, financial, etc.”).

32. *Id.* (“Nearly 100 different peer-reviewed journals. These were selected to cover the entire range of the Library of Congress classification system (e.g. a certain percentage from B (philosophy, psychology, religion), D (world history), K (education), T (technology), etc.), both overall and by number of words per year[.]”).

33. Davies created “[t]he underlying corpus architecture and web interface In most cases, he also designed, collected, edited, and annotated the corpora as well,” but he also received help from his BYU students to scan some of the sources. *Frequently Asked Questions*, BYU CORPORA, <http://corpus.byu.edu/faq.asp#x1> [<https://perma.cc/N37M-ZC7X>].

34. This is also called a “lemmatized” corpus. See *State v. Rasabout*, 356 P.3d 1258, 1281 n.28 (Utah 2015) (Lee, J., concurring).

35. Stephen C. Mouritsen, *Hard Cases and Hard Data: Assessing Corpus Linguistics as an Empirical Path to Plain Meaning*, 13 COLUM. SCI. & TECH. L. REV. 156, 192 (2012).

36. See SUSAN HUNSTON, CORPORA IN APPLIED LINGUISTICS 3 (2002) (“Strictly speaking, a corpus by itself can do nothing at all, being nothing other than a store of used language. Corpus access software, however, can rearrange that store so that observations of various kinds can be made.”).

37. Mouritsen, *supra* note 20, at 1956–57.

38. *Id.* at 1923 (“The implicit claim made when a given definition is included in a dictionary is not that a particular meaning is correct or even common, but that its use, in a given context, is verifiable.”); see also *Rasabout*, 356 P.3d at 1273 (Lee, J., concurring) (“[T]he dictionary generally cannot provide the basis for preferring one of its definitions over another.”).

39. Mouritsen, *supra* note 20, at 1923 (emphasis omitted) (quoting HENRY M. HART, JR. & ALBERT M. SACKS, *THE LEGAL PROCESS: BASIC PROBLEMS IN THE MAKING AND APPLICATION OF LAW* 1190 (William N. Eskridge, Jr. & Phillip P. Frickey eds., 1994)).

40. *Id.* at 1969.

dispute is intractable because it is based on the judges' differing intuitions about the word's ordinary meaning.⁴¹ In contrast, if Judge *A* uses a corpus-based analysis, then Judge *B* "will have numerous 'overt' bases" to challenge Judge *A*'s corpus analysis.⁴² In the first scenario, the disagreement is a metaphysical debate; in the second scenario, the disagreement is an empirical debate.⁴³ This empiricism is the chief benefit of using a corpus-linguistics database as a supplement to a dictionary. The corpus is "essentially tell[ing] us what a language is like, and the main argument in favour of using a corpus is that it is a more reliable guide to language use than native speaker intuition."⁴⁴

Since the publication of Mouritsen's note,⁴⁵ lawyers and judges have occasionally used corpus linguistics to aid in statutory interpretation.⁴⁶ Most recently, the Michigan Supreme Court became the first state supreme court to do so.⁴⁷ Interestingly, the dissent also used it but disagreed with the majority's corpus analysis.⁴⁸ Some judges are driving forward into the frontier with corpus linguistics.⁴⁹ Some judges think the databases are too complicated to use and should only be considered when the parties cite to a database in their brief.⁵⁰ Others think judicial recognition of corpus linguistics will force parties to hire linguistics experts, which will unduly add to the costs of litigation.⁵¹

This Note analyzes the judicial use of corpus linguistics in statutory interpretation through the lens of the Michigan Supreme Court's decision in *People v. Harris*.⁵² Part I explains the evolution of using corpus linguistics—specifically the COCA—in statutory interpretation and examines the *Harris* opinions' corpus-linguistics analyses. Part II contends that neither opinion in *Harris* conducted a complete analysis of the COCA but that the dissent reached the correct result. And Part III recites some common criticism of the COCA, addresses this criticism, and briefly suggests practical steps for using the COCA.

41. *Id.*

42. *Id.* at 1969–70 (noting that Judge *B* could attack the corpus selection, the construction of the corpus, or the analysis of the corpus).

43. *Id.* at 1970.

44. HUNSTON, *supra* note 36, at 20.

45. *See* Mouritsen, *supra* note 20.

46. *See infra* Sections I.A–I.B.

47. Gordon Smith, *Michigan Supreme Court Embraces Corpus Linguistics*, CONGLOMERATE (June 28, 2016), <http://www.theconglomerate.org/2016/06/michigan-supreme-court-embraces-corpus-linguistics.html> [<https://perma.cc/5SUK-DMAD>].

48. *See* *People v. Harris*, 885 N.W.2d 832, 850 n.14 (Mich. 2016) (Markman, J., concurring in part and dissenting in part).

49. *See, e.g.,* *State v. Rasabout*, 356 P.3d 1258, 1271–90 (Utah 2015) (Lee, J., concurring).

50. *Id.* at 1269–71 (Durrant, C.J., concurring).

51. *Id.* at 1265 (majority opinion) (stating that the use of corpus linguistics could impose "such a significant financial burden on so many of the litigants coming through the doors of our courts [that it] would be tantamount to locking those doors for all but the most affluent").

52. 885 N.W.2d 832.

I. THE BACKGROUND OF CORPUS LINGUISTICS IN THE LAW

Part I traces the history of corpus linguistics in statutory interpretation. Section I.A discusses the emergence of corpus linguistics in scholarship, advocacy, and judicial decisionmaking. Section I.B examines the Michigan Supreme Court's use of corpus linguistics in *People v. Harris* by detailing the different types of searches that both the majority and the dissent used in their respective analyses.

A. *The Legacy of Corpus Linguistics in Statutory Interpretation*

Before the COCA appeared in any judicial opinions, it debuted in an amicus brief before the U.S. Supreme Court. In *FCC v. AT&T Inc.*,⁵³ the Court had to determine whether corporations had “personal privacy” or if the phrase applied only to humans.⁵⁴ The Project on Government Oversight (“POGO”) submitted an amicus brief that used multiple corpora to determine the ordinary meaning of “personal privacy.”⁵⁵ Specifically, POGO used tagged corpora,⁵⁶ including the COCA, which allowed POGO to search for nouns that the adjective “personal” most often modified.⁵⁷ POGO included the corpora's data in its brief.⁵⁸ At oral argument, Justice Ginsburg asked AT&T's attorney to respond specifically to POGO's amicus brief.⁵⁹ Corpus linguistics had arrived.

In 2015, Justice Lee of the Utah Supreme Court used corpus linguistics (and specifically the COCA) in a judicial opinion.⁶⁰ In *State v. Rasabout*, the Utah Supreme Court had to determine whether the phrase “discharge of a firearm” constituted firing one bullet only or included firing one sequence of several bullets.⁶¹ The defendant, Andy Rasabout, had fired twelve shots

53. 562 U.S. 397 (2011).

54. *AT&T*, 562 U.S. at 399–400. The case arose out of a FOIA request for correspondence that the FCC had on file from an earlier investigation of AT&T. *Id.* at 400. The FCC withheld some of the requested information under FOIA Exemption 7(C), which protects “‘records or information’ . . . that ‘could reasonably be expected to constitute an unwarranted invasion of personal privacy.’” *Id.* at 401 (quoting 5 U.S.C. § 552(b)(7)(C) (2006)).

55. See Brief for the Project on Government Oversight et al. as Amici Curiae Supporting Petitioners at 13–20, *AT&T*, 562 U.S. 397 (No. 09-1279), 2010 WL 5069527, at *13–20 [hereinafter POGO Amicus Brief] (concluding that only humans could have “personal privacy” in the ordinary sense of the phrase).

56. See *supra* notes 34–35 and accompanying text.

57. POGO Amicus Brief, *supra* note 55, at 16 (searching the COCA; the TIME Corpus of American English, a 100 million-word corpus of texts taken from *Time* magazine; and the Corpus of Historical American English).

58. *Id.* at 16–20.

59. Transcript of Oral Argument at 37, *AT&T*, 562 U.S. 397 (No. 09-1279) (“JUSTICE GINSBURG: Mr. Klineberg, you have read the brief of the Project on Government Oversight, where they give dozens and dozens of examples to show that, overwhelmingly, ‘personal’ is used to describe an individual, not an artificial being.”).

60. *State v. Rasabout*, 356 P.3d 1258, 1281–82 (Utah 2015) (Lee, J., concurring).

61. *Id.* at 1262 (majority opinion).

during a drive-by shooting.⁶² He was charged with twelve felony counts of “unlawful discharge of a firearm”—one charge for each bullet.⁶³ On appeal, Rasabout argued that the phrase “discharge of a firearm” ordinarily means one shooting sequence, regardless of how many bullets were fired.⁶⁴ Justice Lee searched the COCA to locate all the instances where the word “discharge” appeared within five words of either “firearm,” “firearms,” “gun,” or “weapon.”⁶⁵ His search returned eighty-six instances.⁶⁶ And the returns overwhelmingly suggested that “discharge of a firearm” refers to the firing of a single bullet.⁶⁷ In fact, he found only one instance that unambiguously supported Rasabout’s argument.⁶⁸ Therefore, Justice Lee concluded that “discharge of a firearm” ordinarily means firing of only one shot.⁶⁹

Justice Lee’s colleagues did not embrace his use of the COCA. The majority stated that it refused to follow his approach because the parties did not brief it and because it was *sua sponte* scientific research with no scientific review.⁷⁰ The majority then wrote that no judge would be able to analyze the COCA reliably because most judges are generalists and do not have the “scientific expertise” to use the COCA.⁷¹ The *Rasabout* majority also thought that permitting the use of the COCA would force parties to hire linguistics experts, which would “[i]mpos[e] such a significant financial burden on so many of the litigants coming through the doors of our courts [that it] would be tantamount to locking those doors for all but the most affluent.”⁷²

Chief Justice Durrant concurred in part, but he took a more conciliatory tone with regard to Justice Lee’s corpus-linguistics analysis.⁷³ The chief justice wrote that, although corpus linguistics could be a great aid to statutory interpretation, the parties—not the court—should conduct the analysis because then the adversaries could challenge one another’s database, methodology, and conclusion.⁷⁴ The Utah Supreme Court’s differing views raised a number of questions about the use of corpus linguistics for statutory interpretation, especially by judges *sua sponte*.

62. *Id.* at 1260.

63. *See id.* at 1261.

64. *Id.* at 1264.

65. *Id.* at 1281–82 (Lee, J., concurring).

66. *Id.* But five of the instances were dismissed as duplicates. *Id.* at 1282 n.30.

67. *Id.* at 1282.

68. *See id.*

69. *Id.*

70. *Id.* at 1264–65 (majority opinion).

71. *Id.* at 1265 (“Indeed, we are aware of almost no one sitting on the bench or practicing law in this state who has the scientific expertise required to reliably conduct the research Justice Lee requires.”).

72. *Id.*

73. *Id.* at 1269 (Durrant, C.J., concurring) (“I applaud Justice Lee for his thoughtful exploration of corpus linguistics as a potential additional tool for our statutory interpretation tool box.”).

74. *Id.* at 1270.

B. *The Michigan Supreme Court's Use of the COCA*

In the spring of 2016, the Michigan Supreme Court became the first state supreme court to use the COCA in a majority opinion when it decided *People v. Harris*.⁷⁵ The dissent also used the COCA but executed a different search method and ultimately disagreed with the majority's analysis.⁷⁶ This Section reviews these two opinions. Section I.B.1 discusses the facts of the case. Section I.B.2 examines the *Harris* majority's COCA analysis. And Section I.B.3 examines the *Harris* dissent's COCA analysis.

1. Factual Background

Harris arose out of an alleged assault at a Detroit gas station.⁷⁷ Dajuan Hodges-Lamar was sitting in his parked car at the gas station when Detroit Police Officer Hughes approached him.⁷⁸ After Hughes asked Hodges-Lamar some questions, Hughes suddenly began beating him, while Hughes's fellow officers, Harris and Little, watched.⁷⁹ After the attack, the Detroit Police Department's Office of the Chief Investigator opened an internal investigation that required all three officers (Hughes, Harris, and Little) either to testify at a *Garrity* hearing⁸⁰ or potentially face termination.⁸¹ At the hearing, the investigators asked about the altercation, and all three officers denied using any more force than was necessary.⁸² The Detroit Police Department's Internal Affairs Section later obtained the gas station's video-surveillance footage,⁸³ which showed the officers had made false statements.⁸⁴ As a result, the State charged all three officers with obstruction of justice.⁸⁵

75. Smith, *supra* note 47.

76. See *People v. Harris*, 885 N.W.2d 832, 850 n.14 (Mich. 2016) (Markman, J., concurring in part and dissenting in part).

77. *Id.* at 834 (majority opinion).

78. *Id.*

79. *Id.*

80. *Id.* A *Garrity* hearing is named for the Supreme Court's decision in *Garrity v. New Jersey*, 385 U.S. 493 (1967). See *People v. Brown*, 755 N.W.2d 664, 683 (Mich. Ct. App. 2008) ("Essentially, a *Garrity* hearing allows the interviewee to answer questions with the knowledge that any statements elicited therein will not be used against him in criminal proceedings.").

81. *Harris*, 885 N.W.2d at 834 (stating that the officers received an advice-of-rights form warning the officers that if they refused to answer questions they would be subject to departmental charges, potentially including dismissal from the police force).

82. *Id.* at 835.

83. Plaintiff-Appellee's Amended Brief on Appeal at 5, *Harris*, 885 N.W.2d 832 (Nos. 149872, 149873 & 150042), 2015 WL 9874034, at *5.

84. *Harris*, 885 N.W.2d at 835. The video showed Hughes "striking Hodges-Lamar with an open hand in the throat, punching him again, pushing him to the ground, picking him up by the collar several times, slamming him onto the car, and pushing him back toward Harris and Little." *Id.* at 835 n.4; see also WXYZ-TV Detroit Channel 7, *DPD Beating Caught on Tape*, YouTube (June 30, 2011), <https://www.youtube.com/watch?v=AZSoihW83k4> [<https://perma.cc/2C35-WXGZ>].

85. *Harris*, 885 N.W.2d at 835.

Despite the dramatic circumstances, the case's outcome came down to the ordinary meaning of the word "information." In the Michigan trial court, the officers argued that their statements were protected by Michigan's Disclosures by Law Enforcement Officers Act,⁸⁶ which states that "any information derived from [an] involuntary statement, shall not be used against the law enforcement officer in a criminal proceeding."⁸⁷ Section 15.391 of the Michigan Compiled Laws defines "involuntary statement" as "information provided by a law enforcement officer, if compelled under threat of dismissal from employment."⁸⁸ The officers argued that their statements, though false, still constituted "information," so the Act prohibited the State from using the statements to support the obstruction of justice charges.⁸⁹ The Michigan trial court agreed and dismissed the charges.⁹⁰ The Michigan Court of Appeals later reversed, mostly basing its decision on the dictionary definition of "information."⁹¹

The issue, once it reached the Michigan Supreme Court, was whether the definition of "information" includes false statements.⁹² All four Michigan appellate opinions—two majorities and two dissents—stated that the goal was to find the plain meaning of the word "information."⁹³ And at the Michigan Supreme Court, both the majority and the dissent used the COCA.⁹⁴

2. The Majority Opinion's Use of the COCA

The majority described its task as interpreting "information" "according to the common and approved usage of the language" and turned to the

86. *Id.*

87. MICH. COMP. LAWS ANN. § 15.393 (West 2004 & Supp. 2017).

88. *Id.* § 15.391 (West 2004 & Supp. 2017).

89. *Harris*, 885 N.W.2d at 835–36.

90. *Id.* at 835.

91. *People v. Hughes*, 855 N.W.2d 209, 216 (Mich. Ct. App. 2014), *rev'd sub nom. Harris*, 885 N.W.2d 832.

92. *Harris*, 855 N.W.2d at 838.

93. *Id.* at 837 ("The plain language of the [Act] controls our resolution of this dispute . . ."); *id.* at 845 (Markman, J., concurring in part and dissenting in part) ("I agree with the Court of Appeals that false statements do not constitute 'information.'"); *Hughes*, 855 N.W.2d at 216 ("In sum, the plain language of MCL 15.391(a) establishes that an 'involuntary statement' includes only truthful and factual information."); *id.* at 219 (Wilder, J., dissenting) ("We make every effort to interpret clear and unambiguous language in accordance with its plain meaning . . .").

94. *Harris*, 885 N.W.2d at 838–39 (majority opinion); *id.* at 850 n.14 (Markman, J., concurring in part and dissenting in part).

COCA as a resource.⁹⁵ The opinion then used the COCA to conduct a collocation search.⁹⁶ This type of search calculates how often a specific word appears near other words.⁹⁷ The COCA search box has a specific tab entitled “Collocates,”⁹⁸ where the user enters a “node word” (the word that is the basis of the search). The user then selects the range of words to search—for example, four words to the right or to the left of the node word.⁹⁹ The user can also modify the context of the word—such as searching for only a word’s verb form¹⁰⁰—or the user can search for all the examples where one particular word appeared within the specified range to the user’s node word.¹⁰¹ The COCA will then generate a list of the results.¹⁰²

In *Harris*, the majority used “information” as its node word and selected a range of four words to the right and to the left of the word “information.”¹⁰³ As a result, the COCA generated a list that ranked words by how often they appeared within four words of “information.”¹⁰⁴ For example, “provide” is the word that most frequently appeared within four words of “information,” appearing 4,223 times.¹⁰⁵

95. *Id.* at 838 (majority opinion) (quoting MICH. COMP. LAWS ANN. § 8.3a (West 2004)).

96. *Id.* at 839 n.33.

97. Mouritsen, *supra* note 20, at 1962 (“‘Collocation is the tendency of words to be biased in the way they co-occur,’ that is, the tendency of certain words to be used in the same semantic environment as other words.” (footnote omitted) (quoting HUNSTON, *supra* note 36, at 68)).

98. CORPUS CONTEMP. AM. ENG., *supra* note 27 (follow “Collocates” hyperlink above the search box).

99. Mouritsen, *supra* note 35, at 200.

100. See Mouritsen, *supra* note 20, at 1962–63.

101. See, e.g., *State v. Rasabout*, 356 P.3d 1258, 1281–82 (Utah 2015) (Lee, J., concurring) (using the COCA to search for the instances where the nouns “firearm,” “firearms,” “gun,” and “weapon” appeared within five words of the verb “discharge”).

102. Mouritsen, *supra* note 20, at 1963.

103. *People v. Harris*, 885 N.W.2d 832, 839 n.33 (Mich. 2016) (“The collocation search for ‘information’ is available at Corpus of Contemporary American English, “Information” Frequency <<http://corpus.byu.edu/coca/?c=coca&q=47913597>> (accessed June 6, 2016).”).

104. “Information” Collocation Frequency, CORPUS CONTEMP. AM. ENG., <http://corpus.byu.edu/coca/?c=coca&q=47913597> (follow the “Find collocates” hyperlink) [<https://perma.cc/GT9J-9NNR?type=image>].

105. *Id.*

FIGURE 1. RETURNS FOR THE MAJORITY'S COLLOCATION SEARCH

RANK	CONTEXT	FRQ	ALL	%	MI
1	PROVIDE	4223	93081	4.54	4.17
2	TECHNOLOGY	2819	79997	3.52	3.81
3	AVAILABLE	2685	85438	3.14	3.64
4	PROVIDED	2392	53412	4.48	4.15
5	ACCESS	2194	56241	3.90	3.95
6	CONTACT	1914	41050	4.66	4.21
7	SYSTEMS	1911	63837	2.99	3.57
8	SOURCES	1720	38910	4.42	4.13
9	ADDITIONAL	1627	41095	3.96	3.97
10	SOURCE	1431	48432	2.95	3.55
11	PROVIDES	1413	38520	3.67	3.86
12	SHARE	1323	62183	2.13	3.08
13	REGARDING	1304	25967	5.02	4.32
14	PROVIDING	1228	29305	4.19	4.06
15	USEFUL	1215	24157	5.03	4.32

The majority then scanned the ranked list of returns and looked for adjectives that related to “truth or falsity.”¹⁰⁶ The majority used this approach to determine that the word “accurate” was the most common adjective “that refers to truth or falsity” located within four words of “information”; the words “false” and “inaccurate” were also commonly collocated with “information.”¹⁰⁷ This data, the majority reasoned, showed that in common usage, “‘information’ is regularly used in conjunction with adjectives suggesting that it may be both true and false,”¹⁰⁸ so the majority found that the unmodified word “information” could describe either true or false statements.¹⁰⁹ As a result, the majority concluded that the officers’ false statements constituted “information” and thus could not be used in a criminal prosecution under the Act.¹¹⁰ The court then reversed the court of appeals’ ruling and reinstated the district court’s order dismissing the obstruction-of-justice charges for all three officers.¹¹¹

3. The Dissent’s Use of the COCA

The dissent used a simpler approach than the majority and reached the opposite conclusion. Justice Markman, joined by Justice Viviano, used the

106. See *Harris*, 885 N.W.2d at 839 n.33.

107. *Id.*

108. *Id.* at 839.

109. *Id.*

110. *Id.* at 842–43.

111. *Id.* at 845.

COCA to search for the number of times the word “information” appeared in the COCA’s sources.¹¹² This search turned up 168,187 hits.¹¹³

FIGURE 2. RETURN FOR THE DISSENT’S FREQUENCY SEARCH

The screenshot shows the COCA interface with the 'FREQUENCY' tab selected. The search results table is as follows:

	CONTEXT	FREQ
1	INFORMATION	168187

Additional interface elements include: 'SEE CONTEXT: CLICK ON WORD OR SELECT WORDS + [CONTEXT] [HELP...]', 'COMPARE', and a loading indicator '0.875 seconds'.

The dissent then looked to the same collocation-frequency list that the majority used and found that the adjectives bearing meaning to truth or falsity rarely modified “information.”¹¹⁴ Specifically, the dissent pointed out that “truthful” only modified “information” 28 times, “true” only 18 times, “accurate” 508 times, “inaccurate” 112 times, and “false” 271 times.¹¹⁵ This totaled 937 times that “information” was modified by these adjectives.¹¹⁶ “The other 167,250 times that the word ‘information’ is used it is unmodified by one of these adjectives. That is, 99.44% of the time ‘information’ in the COCA is unmodified by any of these adjectives related to veracity.”¹¹⁷

Most importantly, the dissent did something the majority opinion did not: it looked at the results in context.¹¹⁸ After every search, the COCA displays a list of “concordance lines,” which display the word with its surrounding context.¹¹⁹ A concordance line gives a preview of the source material by showing the node word—in this case, “information”—and the surrounding text on each side of that word.¹²⁰ A user can then ask the COCA to display a random sample of the concordance lines in the search results, which allows the user to review the node word’s use in context to determine the word’s ordinary meaning.¹²¹

112. *Id.* at 850 n.14 (Markman, J., concurring in part and dissenting in part) (noting the total number of times the word “information” appears in the COCA).

113. *Id.*; see also “Information” Frequency, CORPUS CONTEMP. AM. ENG., <http://corpus.byu.edu/coca/?c=coca&q=58969614> [<https://perma.cc/S4TX-VL36?type=image>].

114. See *Harris*, 885 N.W.2d at 850 n.14 (Markman, J., concurring in part and dissenting in part).

115. *Id.*

116. *Id.*

117. *Id.*

118. *Id.* (listing examples of the unmodified word “information” from the COCA).

119. Mouritsen, *supra* note 35, at 197; see also discussion *infra* Section II.B.

120. Mouritsen, *supra* note 35, at 197 (stating that the display of concordance lines is referred to as a “Key Word in Context or KWIC display”).

121. See Mouritsen, *supra* note 20, at 1958; see also “Information” Frequency, *supra* note 113 (allowing a user to click “information” and display the concordance lines under the “Context” tab; a user can then select a random sample of 100, 200, 500, or 1000 concordance lines).

The dissent cited multiple concordance lines in support of its conclusion but did not specify exactly how many lines it reviewed.¹²² The dissent even invited the reader “to peruse at random any number of the 167,250 uses of ‘information’ in the COCA and assess whether the term was reasonably used and understood as indistinguishably referring to true and false information.”¹²³ Ultimately, the dissent concluded that the plain meaning of the word “information” does not include false statements,¹²⁴ so the statute allowed the State to use the false statements in its obstruction-of-justice case against the three officers.¹²⁵

The *Harris* court was the first to use corpus linguistics, and specifically the COCA, in a state supreme court majority opinion. Unfortunately for proponents of using corpus linguistics in statutory interpretation, the very first decision to use the COCA revealed a potential flaw: reasonable judges trying their best to use the COCA can reach opposite conclusions.

II. REVIEWING THE MICHIGAN SUPREME COURT’S COCA ANALYSIS

This Part asserts that neither opinion in *Harris* conducted a complete analysis of the COCA.¹²⁶ Section II.A contends that the majority erred by using the wrong search method to determine the meaning of a noun. Section II.B argues that, even though the dissent selected the correct search method, the opinion declined to state how many search returns it reviewed before arriving at its conclusion, which reduced the dissent’s analytical transparency.

A. Search Methods

The majority erred by conducting a collocation search¹²⁷ within the COCA. It is possible that the *Harris* majority copied the search technique used by POGO in its amicus brief for *FCC v. AT&T Inc.*¹²⁸ In that case, however, the question was the meaning of an adjective, not a noun.¹²⁹ POGO described its mission as follows: “Our method will be to find out what nouns *personal* most often modifies, and we will do this by querying each corpus so that it returns the nouns that appear most frequently in the

122. See *Harris*, 885 N.W.2d at 850 n.14 (Markman, J., concurring in part and dissenting in part) (listing examples from the COCA that use the unmodified word “information”).

123. *Id.*

124. *Id.* at 855.

125. *Id.* at 861.

126. See *infra* Section III.D (explaining how to conduct a complete COCA analysis).

127. See *supra* notes 99–104 and accompanying text.

128. See POGO Amicus Brief, *supra* note 55, at 16.

129. *Id.* at 2–3.

position immediately following *personal*.¹³⁰ This is nearly identical to what the *Harris* majority did with the word “information.”¹³¹

For the *Harris* majority, however, POGO’s method was inadequate because POGO was trying to determine the meaning of an adjective, while the *Harris* majority was trying to determine the meaning of a noun.¹³² By definition, an adjective *always* modifies something,¹³³ so it made sense for POGO merely to rank the words that the adjective “personal” modified to determine its meaning (because the adjective would always be modifying another word). In contrast, the *Harris* majority tried to use this same method to determine the meaning of the noun “information.” But as the dissent made clear, other words modified “information” only 0.56% of the time.¹³⁴ A user must conduct different types of searches for different parts of speech. The *Harris* majority used the wrong search method in the COCA to find the ordinary meaning of “information.”

B. Contextual Review

From the descriptions of the searches that each opinion provided, the *Harris* dissent used the COCA more effectively. This conclusion is based on the familiar principle that the meaning of a word must be drawn from context—not by considering the word in isolation.¹³⁵ Ultimately, “[b]ecause common words typically have more than one meaning, you must use the context in which a given word appears to determine its aptest, most likely sense.”¹³⁶ The COCA conveniently provides a tool for this very purpose, called the Keyword-in-Context or “KWIC.”¹³⁷

After a user conducts a search, the KWIC displays the search results as a list of “concordance lines.”¹³⁸ These lines allow the user to see a twenty-eight-word snapshot of the source that contains the word that the user searched.¹³⁹

130. *Id.* at 16.

131. See *People v. Harris*, 885 N.W.2d 832, 839 n.33 (Mich. 2016) (“In conducting a COCA search, the word ‘accurate’ is the most common adjective collocated with ‘information’ to bear a meaning that refers to truth or falsity.”).

132. Compare *id.* at 838–39, with POGO Amicus Brief, *supra* note 55, at 8.

133. BRYAN GARNER, *GARNER’S MODERN ENGLISH USAGE* 19 (4th ed. 2016) (“An adjective is a word that modifies a noun.”).

134. *Harris*, 885 N.W.2d at 850 n.14 (Markman, J., concurring in part and dissenting in part).

135. *Deal v. United States*, 508 U.S. 129, 132 (1993) (citing *King v. St. Vincent’s Hosp.*, 502 U.S. 215, 221 (1991); *Davis v. Mich. Dep’t of Treasury*, 489 U.S. 803, 809 (1989); *United States v. Morton*, 467 U.S. 822, 828 (1984)); see also SCALIA & GARNER, *supra* note 6, at 56 (“[W]ords are given meaning by their context.”).

136. SCALIA & GARNER, *supra* note 6, at 418.

137. Mouritsen, *supra* note 35, at 197.

138. *Id.*; see, e.g., Figure 3 (showing an example of twenty-five random concordance lines displayed in the KWIC).

139. Mouritsen, *supra* note 35, at 197.

FIGURE 3. A RANDOM SAMPLE OF 25 CONCORDANCE LINES FOR THE WORD “INFORMATION”

Corpus of Contemporary American English			
SEARCH	FREQUENCY	CONTEXT	HELP
TeachLibrar	A B C	challenge for today's schools. It is hoped this article might stimulate debate about information literacy and transfer in schools. # REFERENCES Charraz, K. (2006).	
PhysicalEduc	A B C	population is in need of skill-building and education regarding reducing the risk of osteoporosis. Information on vitamins and minerals that build bone, wise food choices, the impo	
USNWR	A B C	# Obsessed. When people become aware of a threat, they typically seek out information on it and start figuring out what to do. Some might obsess: internet	
Fortune	A B C	, PGP's director of consulting. Not only can PKF sift through Sidebar financial information in the here and now, but its accountants can do so in collaboration with	
ReadingTeacher	A B C	gives children a rich opportunity to engage in learning the vocabulary, concepts, and information in many different contexts. " # " These standards may be transformative in enhan	
CNN_Zahn	A B C	were something like 2,500, 3,000 Web sites to track all the business and the information for which that place is responsible. So who knows why somebody at probably a	
Environmental	A B C	claimed an undue financial burden and argued that their First Amendment right not to disclose information had been violated. The lawsuit was later thrown out by two federal app	
EmergingInfectious	A B C	have a major role in strengthening specific areas of health systems, such as health information systems and surveillance, laboratory capacity, workforce skills, operational research	
CSMonitor	A B C	to a conflict at the family business. He confirmed that investigators were still seeking information from the armed forces. In October, Mangula, the defense minister, and	
SocialStudies	A B C	. The Polish teacher team was as successful as the student teams in sharing that information orally. Extending Students' Learning # I was pleased with the results of the	
CollegeStud	A B C	. # Students seemed particularly concerned about the cost of healthcare and spoke of googling information as a way of determining whether the cost of a doctor's visit is warrant	
MusicEduc	A B C	of the band's purpose was to gain some visibility for the concept and gather information that would help others create their own version of the model. Nine years later	
Atlantic	A B C	page was one of several released to the National Security Archive through the Freedom of Information Act: the memo itself was not released.) " Look, they came	
SanFranChron	A B C	the city characterized as a suspected carcinogen, insisted that officials were working with outdated information when it assessed their product. Ron Benjamin, the company's manu	
PhysicalEduc	A B C	to examine the prevalence of burnout among Korean PE teacher-athletic directors. Table 2 provides information relative to this purpose. In Table 2, means and standard deviation	
InstrPsych	A B C	It comes to intervening with the student. This simple case study highlights the way information from the functional analysis assessment was used to implement a simple, effective,	
TeachLibrar	A B C	has a question that will be a good guide for research. # For more information , see the Background to Question model from Ban Those Bird Units (Loertscher,	
TeachingExceptional	A B C	. It is important to remember, though, that educational personnel should gather additional information about students before changing their educational program. For example, Al	
CSMonitor	A B C	no conclusions about the death toll among children and the elderly. The lack of information about whether the dead were bystanders or combatants also leaves open the debate c	
TeachLibrar	A B C	WITH MIDDLE SCHOOL STUDENTS. # Results of the unit show that giving students background information on the internet and showing Internet site examples to change their pre	
Education	A B C	the movement concepts and creatively perform the movement in different ways, she presented the information by prompting children's intuitive knowledge base of cloud shapes.	
BioScience	A B C	publishers, professional societies, and the editorial boards of journals to provide standardized geographic information in their publications and with their publicly shared metada	
NPR_TalkNation	A B C	mother had e-mailed a friend of hers about some details related to the event - information I needed to share with my husband, so I forwarded it onto him without	
EmergingInfectious	A B C	the need to continually monitor isolates obtained from mammal reservoirs for genetic variation. This information may help guide clinical treatment and assessment of epidemiolog	
NYTimes	A B C	It helps to place them in a strainer or pasta pot insert). Nutritional information per serving (6 servings): 340 calories; 4 grams fat 0	

The KWIC would have displayed these concordance lines for the dissent's search returns of "information."¹⁴⁰ As discussed previously, from the KWIC display, the user can generate a random sample of 100, 200, 500, or 1,000 concordance lines and then review that sample to avoid the process of reviewing every single line.¹⁴¹

There is an open question about how many concordance lines a judge should review.¹⁴² If judges were forced to review too many lines, the COCA would lose its viability as a helpful tool for statutory interpretation. In *Rasabout*, Justice Lee's search returned only eighty-six lines total,¹⁴³ which made it easy to review all of them. But in *Harris*, there were 168,187 concordance lines.¹⁴⁴

140. "Information" Frequency, *supra* note 113 (follow "INFORMATION" hyperlink).

141. *Id.* (offering the user the option to "find [a] sample" for 100, 200, 500, or 1000 concordance lines in the upper left corner, below the link for "search").

142. See, e.g., LINDQUIST, *supra* note 26, at 56 ("How many concordance lines a lexicographer needs to write an entry for a word depends on how many different meanings the word in question has, but twenty is seldom enough."); Mouritsen, *supra* note 35, at 197 & n.203 (reviewing 1,000 concordance lines for "thoroughness," but noting that one scholar has suggested that reviewing 500 concordance lines per 10,000 could be sufficient).

143. State v. Rasabout, 356 P.3d 1258, 1281–82 (Utah 2015) (Lee, J., concurring).

144. People v. Harris, 885 N.W.2d 832, 850 n.14 (Mich. 2016) (Markman, J., concurring in part and dissenting in part).

One linguistics scholar has proposed a solution to analyzing common words that yield thousands of concordance lines.¹⁴⁵ His method takes a sample of twenty-five to thirty concordance lines, which the COCA provides for with its “Find Sample” button.¹⁴⁶ The judge would examine this sample of twenty-five to thirty lines and classify them.¹⁴⁷ In *Harris*, the classifications could have been whether the source used the word “information” (1) to include both true and false statements, (2) to include only true statements, or (3) in an ambiguous manner. Then the judge would select another random twenty-five to thirty lines and repeat the process.¹⁴⁸ The judge would continue this process until the next set of twenty-five to thirty lines did not reveal any changes in the observed pattern.¹⁴⁹ For example, in *Harris*, a judge could have ceased reviewing concordance lines when the new set of lines did not change the previously observed ratio of possible meanings of “information”—among both truthful and false meanings, only truthful meanings, and ambiguous meanings.

Based on the majority opinion’s description of its analysis, the majority reviewed zero concordance lines¹⁵⁰ and merely ranked the *number* of times that the COCA displayed the word “information” being modified by an adjective suggesting truth or falsity.¹⁵¹ Reliance on gross numbers, like the majority did here, can lead to misleading results¹⁵² because “[a] particular word may have a very broad semantic range and may collocate with numerous terms.”¹⁵³ “Information” is such a word. Although the majority’s calculation was correct, as the dissent pointed out, simply ranking the adjectives that relate to veracity by the number of times they modified the word “information” missed the mark.¹⁵⁴ If a word is only modified 0.56% of the time, then

145. John Sinclair, *A Way with Common Words*, in *OUT OF CORPORA: STUDIES IN HONOUR OF STIG JOHANSSON* 166 (Hilde Hasselgard & Signe Oksefjell eds., 1999).

146. See *supra* note 141 and accompanying text. Sinclair describes his method as using the “every *nth*” convention, meaning that the user selects a concordance line from an evenly spaced sequence through the range of concordance lines. Sinclair, *supra* note 145, at 166. For example, if there are 300 concordance lines, the reader chooses numbers 1, 11, 21, 31, and so on until 30 concordance lines have been selected. See *id.*

147. See Sinclair, *supra* note 145, at 166.

148. See *id.*

149. See *id.* (“This process continues until each new set of instances adds little or nothing to the description . . .”).

150. *People v. Harris*, 885 N.W.2d 832, 839 n.33 (Mich. 2016).

151. *Id.* (“In conducting a COCA search, the word ‘accurate’ is the most common adjective collocated with ‘information’ to bear a meaning that refers to truth or falsity.”).

152. HUNSTON, *supra* note 36, at 79 (“It is tempting, when looking at a list of collocates, to draw conclusions about the overall frequency of compounds and phrases that may not be justified.”).

153. Mouritsen, *supra* note 35, at 201.

154. *Harris*, 885 N.W.2d at 850 n.14 (Markman, J., concurring in part and dissenting in part) (“99.44% of the time ‘information’ in the COCA is unmodified by any of these adjectives related to veracity.”).

the statistics suggest that the ordinary usage of the word is in an unmodified form.¹⁵⁵

Judges should always review the concordance lines to read the word in context.¹⁵⁶ Justice Lee did this in his *Rasabout* concurrence: “By examining the instances of *discharge* in connection with these nearby nouns, I confirmed that the single shot sense of this verb is overwhelmingly the ordinary sense of the term in this context.”¹⁵⁷ He detailed his analysis of the eighty-one unique concordance lines that the COCA returned for his search.¹⁵⁸ The dissenting opinion in *Harris* reviewed concordance lines, but it did not indicate exactly how many lines it reviewed¹⁵⁹ or whether it used a principled process to analyze the thousands of concordance lines.¹⁶⁰ The dissent wrote only that the word “information” was “overwhelmingly used to refer to truthful information.”¹⁶¹ And although the dissent did list two concordance lines,¹⁶² the opinion lacked the detailed description of its concordance-line review that Justice Lee included in his *Rasabout* concurrence.¹⁶³

The *Harris* dissent should have stated how many concordance lines it reviewed and what process it used to sift through the thousands of concordance lines. It could have used the random-sample technique to review groups of twenty-five to thirty random concordance lines until the pattern of usage for “information” became clear. Thus, the *Harris* dissent used the correct search method in the COCA for a noun, but the opinion should have presented an analysis similar to Justice Lee’s in *Rasabout* so that the reader could see exactly what the concordance-line review revealed.¹⁶⁴

155. *Id.*

156. See Mouritsen, *supra* note 35, at 201 (“[C]ollolocations can be helpful as a guide to the semantic field that a given word may occupy, but they cannot serve as a substitute for reviewing concordance lines.”).

157. *State v. Rasabout*, 356 P.3d 1258, 1282 (Utah 2015) (Lee, J., concurring).

158. *Id.* (“Twelve of the eighty-one hits in the COCA search clearly linked *discharge* to a single bullet. In sixteen other hits, the discharge was accidental. . . . Fifteen other hits were a bit more ambiguous; but on closer examination, the *discharge* in question seemed to imply a single shot In thirty-six other instances I concluded there was insufficient detail In all, I found only one instance of *discharge* of a weapon that seemed consistent with the firing of multiple shots.”).

159. See *Harris*, 885 N.W.2d at 850 n.14 (Markman, J., concurring in part and dissenting in part) (listing concordance lines, but not identifying how many the dissent reviewed).

160. See *supra* notes 147–151 and accompanying text.

161. *Harris*, 885 N.W.2d at 850 n.14 (Markman, J., concurring in part and dissenting in part).

162. See *id.*

163. See *supra* note 158 and accompanying text.

164. This would change the disagreement from a metaphysical debate about what is the most ordinary meaning to an empirical debate about what the concordance lines actually say. See Mouritsen, *supra* note 20, at 1969–70.

III. THE COCA AFTER *HARRIS*

This Part argues that, even though judges can misuse the COCA, it is still a valuable tool that judges should use *sua sponte*, within the limits of the database. Section III.A outlines criticism of judges using the COCA *sua sponte*. Section III.B addresses this criticism. Section III.C contends that *Harris* highlighted the COCA's greatest asset: transparency. And Section III.D briefly sets out guidelines for using the COCA effectively.

A. *Criticisms of the COCA*

Despite the COCA's value for determining a word's ordinary meaning,¹⁶⁵ use of the COCA has been criticized; indeed, the split in *Harris*¹⁶⁶ potentially justifies some of this criticism. Underlying the *Rasabout* majority's criticism was the court's belief that advocates and judges would find the database too complicated to use. First, the majority in *Rasabout* expressed concern about the cost imposed on parties if the court permitted the use of the COCA.¹⁶⁷ But the COCA is free to all users,¹⁶⁸ so the majority's concern about cost could not be related to the cost of using the database. Instead, the majority's criticism of cost must have been related to its prediction that COCA analysis "could only be reliably conducted by dueling linguistics experts."¹⁶⁹

Relatedly, the *Rasabout* majority argued that this difficulty also counseled against allowing judges to use the COCA *sua sponte*.¹⁷⁰ According to the majority, this was a "scientific" endeavor, and judges lack the training to use the COCA.¹⁷¹ The split in *Harris* potentially supports this concern by showing that judges could use incorrect search methods if they are unfamiliar with the database.¹⁷²

B. *Responses to the Criticism*

Section III.B.1 addresses the criticism that parties will need to hire dueling linguistics experts if courts permit the use of the COCA for statutory

165. See *supra* notes 38–45 and accompanying text.

166. See discussion *supra* Sections I.B.2–I.B.3.

167. *State v. Rasabout*, 356 P.3d 1258, 1265 (Utah 2015).

168. See *CORPUS CONTEMP. AM. ENG.*, *supra* note 27.

169. *Rasabout*, 356 P.3d at 1265.

170. *Id.* at 1264–65.

171. *Id.* at 1265.

172. See *supra* Part II. No party in *Harris* cited to the COCA in its brief. See Plaintiff-Appellee's Amended Brief on Appeal, *supra* note 83; Defendant-Appellant Hughes' Brief on Appeal, *People v. Harris*, 855 N.W.2d 832 (Mich. 2016) (No. 150042), 2015 WL 9874057; Appellants' Brief, *Harris*, 855 N.W.2d 832 (Nos. 149872, 149873), 2015 WL 9874037.

interpretation. And Section III.B.2 refutes the argument that judges should be prohibited from using the COCA *sua sponte*.¹⁷³

1. Experts

The *Rasabout* majority's concern that litigants will need to hire "linguistics experts" to use the COCA seems misguided. First, the corpus analysis is "something of a last resort."¹⁷⁴ It is used only when (1) the legislature is not using a word in a specialized sense and (2) the court cannot reject either party's definition based on the context or structure of the statute.¹⁷⁵ Therefore, even if the COCA required experts—which it does not¹⁷⁶—using the database would not lock the courtroom doors "for all but the most affluent."¹⁷⁷

More importantly, the COCA is free and accessible to anyone. It includes a tour that describes the basic functions and features of the database.¹⁷⁸ And it offers sample searches that serve as a tutorial for users.¹⁷⁹ For each type of search, the website offers additional guidance.¹⁸⁰ It is unclear whether the *Rasabout* majority attempted to use the COCA; in light of the COCA's tutorials and help guides, however, the concern that parties would be forced to hire experts to use the database seems unfounded. A party or advocate may need to spend time to learn how the site functions, but this time alone does not raise use of the website to the level of scientific, technical, or specialized knowledge that forms the basis of expert opinion.¹⁸¹ As Justice Lee stated, "Corpus analysis, in all events, is not rocket science."¹⁸² Because of the COCA's extensive, free online training guides and the fact that corpus linguistics is not required in every case of dueling interpretations, the COCA does not close the courtroom doors to the poor.

173. The *Rasabout* majority did not consider that judges using the COCA *sua sponte* could actually reduce the need for parties to conduct their own analysis (because the court would be doing it anyway), but this Section will address both concerns.

174. *Rasabout*, 356 P.3d at 1286 (Lee, J., concurring).

175. *Id.*

176. See discussion *infra* notes 182–183 and accompanying text.

177. *Contra Rasabout*, 356 P.3d at 1265 (majority opinion).

178. *Brief Tour*, CORPUS CONTEMP. AM. ENG., <http://corpus.byu.edu/coca/help/tour.asp> [<https://perma.cc/EHZ8-MUSE>].

179. *Id.* (allowing a user to click any of the hyperlinked words to see a sample search with the word).

180. CORPUS CONTEMP. AM. ENG., *supra* note 27 (allowing a user to select the "List," "Chart," "Collocates," "Compare," or "KWIC" hyperlinks above the search box to reveal a help guide on the right side of the display).

181. See, e.g., FED. R. EVID. 702 ("A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify . . .").

182. *Rasabout*, 356 P.3d at 1286 (Lee, J., concurring).

2. Sua Sponte Use

The justification for using dictionaries sua sponte justifies using the COCA sua sponte. In 1875, the U.S. Supreme Court stated in dicta that judicial notice is appropriate for “the meaning of words in the vernacular language.”¹⁸³ In a later case, the Court used a dictionary to define terms because they were “words of common speech, and, as such, their interpretation is within the judicial knowledge, and therefore a matter of law.”¹⁸⁴ A judge’s duty undeniably includes interpreting statutes.¹⁸⁵ As Justice Lee wrote, “In performing the core function of deciding what the law is or should be, we cannot properly be restricted from consulting sources that inform our understanding.”¹⁸⁶ This notion also finds support in the advisory committee note to Federal Rule of Evidence 201.¹⁸⁷

When writing opinions, judges are not limited to using only the cases cited by the parties.¹⁸⁸ More importantly, Justice Lee stressed that the rationale of his *Rasabout* opinion was the same as what the parties had briefed: the ordinary meaning of a phrase in the statute.¹⁸⁹ This meant that the parties *had* briefed the issue, and Justice Lee’s use of the COCA did not violate the principle of the adversary system.

U.S. Supreme Court precedent instructs judges to interpret statutes according to their ordinary meaning,¹⁹⁰ so judges are already obligated to engage in linguistic analysis. The majority in *Rasabout* wrote multiple paragraphs criticizing Justice Lee’s sua sponte use of the COCA.¹⁹¹ But the

183. Samuel A. Thumma & Jeffrey L. Kirchmeier, *The Lexicon Has Become a Fortress: The United States Supreme Court’s Use of Dictionaries*, 47 *BUFF. L. REV.* 227, 246 (1999) (quoting *Brown v. Piper*, 91 U.S. 37, 42 (1875)).

184. *Id.* at 247 (quoting *Marvel v. Merritt*, 116 U.S. 11, 12 (1885)).

185. *See Rivers v. Roadway Express, Inc.*, 511 U.S. 298, 312 (1994).

186. *Rasabout*, 356 P.3d at 1284 (Lee, J., concurring).

187. *FED. R. EVID.* 201(a) advisory committee’s note (“Professor Morgan gave the following description of the methodology of determining domestic law: ‘In determining the content or applicability of a rule of domestic law, the judge is unrestricted in his investigation and conclusion. He may reject the propositions of either party or of both parties. He may consult the sources of pertinent data to which they refer, or he may refuse to do so. He may make an independent search for persuasive data or rest content with what he has or what the parties present.’ . . . This is the view which should govern judicial access to legislative facts.” (citation omitted) (quoting Edmund M. Morgan, *Judicial Notice*, 57 *HARV. L. REV.* 269, 270–71 (1944)).

188. *See, e.g., Rasabout*, 356 P.3d at 1284 n.35 (Lee, J., concurring) (“On issues that are squarely presented . . . we regularly do take it upon ourselves to conduct independent legal research. No party would be surprised to read an opinion citing authority not presented in the briefs, or analysis taking a somewhat different angle than the parties.”).

189. *Id.* at 1283.

190. *See, e.g., Asgrow Seed Co. v. Winterboer*, 513 U.S. 179, 187 (1995).

191. *See supra* Section III.A.

Supreme Court's decision in *Muscarello* shows that judges already use materials outside of dictionaries to find ordinary meaning.¹⁹² The COCA, however, looks at a far greater range of material than a particular justice's handpicked book.¹⁹³ The question, then, is whether to use novel tools to solve an age-old problem.

In response to the *Rasabout* majority's fears, Justice Lee somewhat humorously recalled when online databases—such as Westlaw and Lexis—were first introduced.¹⁹⁴ As he put it, “[S]ome even predicted that computer-based research would undermine the lawyer’s craft.”¹⁹⁵ He proceeded to cite three law review articles that warned of the supposed impending professional crisis that would be ushered in by online legal databases.¹⁹⁶ For example, one warned that computer-aided legal research would “undermine the ability to think like a lawyer.”¹⁹⁷ Another argued that computer-aided legal research “may be harmful to the process of legal reasoning.”¹⁹⁸ And lastly, Justice Lee stated that “the title says it all” for Scott P. Stolley’s article *Shortcomings of Technology: The Corruption of Legal Research*.¹⁹⁹ All of this is to say that there will often be skeptics when a new tool with unfamiliar technology is used to tackle an old problem.²⁰⁰

Although judges may err as they learn to use a new tool for statutory interpretation, judges can also err when they use old tools. In *Muscarello*, the majority began its entire analysis by stressing the order of the possible definitions within the dictionary.²⁰¹ The opinion listed each dictionary’s “first” definition and italicized the word to emphasize the seeming importance of a particular definition being listed first.²⁰² But this is not how dictionaries work.²⁰³ The *Oxford English Dictionary* and *Webster’s Third New International Dictionary* state explicitly that they rank their definitions historically, meaning that older definitions are listed first and more recent definitions are

192. See *Muscarello v. United States*, 524 U.S. 125, 129–30 (1998) (citing to *Robinson Crusoe*, *Moby Dick*, and the *New York Times*, among other sources).

193. See *supra* notes 29–33 and accompanying text.

194. *Rasabout*, 356 P.3d at 1282 (Lee, J., concurring).

195. *Id.*

196. *Id.* at 1282 n.31.

197. *Id.* (citing Barbara Bintliff, *From Creativity to Computerese: Thinking Like a Lawyer in the Computer Age*, 88 LAW LIBR. J. 338, 339 (1996)).

198. *Id.* (quoting Molly Warner Lien, *Technocentrism and the Soul of the Common Law Lawyer*, 48 AM. U. L. REV. 85, 85–86 (1998)).

199. *Id.* (citing Scott P. Stolley, *Shortcomings of Technology: The Corruption of Legal Research*, FOR DEF., Apr. 2004, at 39, 39).

200. *Id.* at 1282.

201. See *Muscarello v. United States*, 524 U.S. 125, 128 (1998)

202. *Id.*

203. See Mouritsen, *supra* note 20, at 1926–38 (describing this error as the “sense-ranking fallacy”).

listed last.²⁰⁴ The error in *Muscarello* shows that judges can misuse even established tools. The possibility that judges could misuse the COCA is not a sufficient reason for prohibiting them from conducting the analysis *sua sponte*. Judges can surely learn how to use the COCA consistently, and they will benefit from the effort.²⁰⁵

In the field of statutory interpretation, some have suggested that a principle or canon of construction is useless if two judges can reach different conclusions when applying the principle or canon in good faith. Karl Llewellyn famously wrote that “there are two opposing canons on almost every point” before listing twenty-eight proposed examples.²⁰⁶ Judge Posner (not a proponent of the canons of construction)²⁰⁷ wrote that some observers think the canons “are figleaves for decisions reached on other grounds.”²⁰⁸

That a tool can potentially be misused is no reason to disregard the tool entirely. Rather, as Justice Scalia and Bryan Garner wrote in response to Llewellyn’s law review article:

Principles of interpretation are guides to solving the puzzle of textual meaning, and as in any good mystery, different clues often point in different directions The skill of sound construction lies in assessing the clarity and weight of each clue and deciding where the balance lies.²⁰⁹

The same principle applies to the use of corpus linguistics. The risk that judges can possibly misuse the COCA is not a reason to cease using what has the potential to become a powerful tool in statutory interpretation’s search for ordinary meaning. Justice Scalia and Garner’s defense of the canons of construction applies with equal force to the use of the COCA.²¹⁰ Corpus linguistics should not be abandoned, and courts should be free to use the method *sua sponte*.²¹¹

204. See *id.* at 1933 & n.101. The Supreme Court is not alone in making this mistake. See *id.* at 1928 n.79 (citing two federal circuit opinions that mistakenly analyze a dictionary’s first listed definition as the primary definition).

205. See Lawrence M. Solan, *The New Textualists’ New Text*, 38 *LOY. L.A. L. REV.* 2027, 2060 (2005) (arguing that computers make it easy for judges to research the lexicon and that “[i]f they perform that task seriously, they stand to learn more about how words are ordinarily used, than by today’s method of fighting over which dictionary is the most authoritative”).

206. Karl N. Llewellyn, *Remarks on the Theory of Appellate Decision and the Rules or Canons About How Statutes Are to Be Construed*, 3 *VAND. L. REV.* 395, 401–06 (1950).

207. See Richard A. Posner, *Statutory Interpretation—in the Classroom and in the Courtroom*, 50 *U. CHI. L. REV.* 800, 806, 816 (1983) (calling the canons “[v]acuous and inconsistent”).

208. *Cont’l Cas. Co. v. Pittsburgh Corning Corp.*, 917 F.2d 297, 300 (7th Cir. 1990).

209. SCALIA & GARNER, *supra* note 6, at 59.

210. See *id.* at 61 (“Yes, they can be abused, as every useful tool can be abused. But we should hardly abandon them.”).

211. In light of this, advocates should include a corpus-based analysis in their briefs when arguing about the ordinary meaning of a word.

C. *The Value of Transparency*

The split in *Harris* reveals the COCA's greatest strength: transparency. As Mouritsen wrote, "[T]he corpus method removes the determination of ordinary meaning from the black box of the judge's mental impression and renders the discussion of ordinary meaning one of tangible and quantifiable reality."²¹² This transparency serves two goals. First, it provides for meaningful appellate review because the COCA allows a judge to describe her methodology and analysis.²¹³ Judges can save their searches and include the links in their opinions so that reviewing courts, advocates, or scholars can replicate the search, conduct the analysis for themselves, and determine whether the court got it right or wrong.²¹⁴ Second, this review process could resolve one of the criticisms of the COCA because it allows the judiciary to refine its use of the COCA by learning from published opinions that describe search methods and analyses.

Therefore, courts should describe their analysis in order to maximize the COCA's benefit to transparency. The COCA is better than a dictionary because the COCA's transparency prevents a judge from asserting that a particular definition from a particular dictionary is *the* ordinary meaning, based on nothing more than the judge's intuition.²¹⁵ The COCA's transparency prohibits this intuitive assertion and forces judges to support their determination of ordinary meaning with data—which provides something tangible to refute or endorse.²¹⁶ Thus, the *Harris* split is not evidence of corpus linguistics' inability to help resolve questions of ordinary meaning. Instead, we are now able to review the court's analysis of what constitutes ordinary meaning and actually comment on whether it was correct.

D. *Steps to Use the COCA*

As *Harris* shows, the legal community may take time to learn how to use the COCA effectively for statutory interpretation. Part of this Note's aim is to provide some basic guidance for future use of the COCA. The first step is to select the proper corpus. This selection will often be determined by the year of enactment for the statute being interpreted because courts generally look to the meaning of the words at the time they were enacted.²¹⁷ Users can

212. Mouritsen, *supra* note 20, at 1970; *see also* State v. Rasabout, 356 P.3d 1258, 1286 (Utah 2015) (Lee, J., concurring) ("I am convinced that [corpus-based analysis] is essential to a more reliable, transparent fulfillment of [interpretation].").

213. *See Rasabout*, 356 P.3d at 1283 ("A judge who presents a transparent corpus analysis opens the curtain in a manner allowing the parties to review and analyze his approach, and even challenge it . . .").

214. *Id.* at 1281.

215. *See supra* notes 37–43 and accompanying text.

216. Mouritsen, *supra* note 35, at 203.

217. *See, e.g., Baker Botts L.L.P. v. ASARCO LLC*, 135 S. Ct. 2158, 2165 & n.2 (2015) (using dictionaries from 1933 and 1934 to determine the meaning of "services" because Congress added the language to the federal bankruptcy law in 1934); *Cuomo v. Clearing House Ass'n*, 557 U.S. 519, 526–29 (2009) (determining the meaning of "visitation" in 1864, the year

choose from a number of corpora. The COCA allows users to search texts published between 1990 and 2015.²¹⁸ Another corpus, the Corpus of Historical American English (“COHA”) allows users to search for words by decade from texts published between 1810 and 2009.²¹⁹ And the forthcoming Corpus of Founding-Era American English (“COFEA”) will cover the period from 1760 to 1799.²²⁰

Once users select the best corpus for the job, they must determine the type of word they are trying to define because users should use different searches for different parts of speech.²²¹ This was the *Harris* majority’s mistake.²²² If a user searches for the meaning of an adjective—like in POGO’s amicus brief²²³—then a collocation search will show the nouns that the adjective most commonly modifies and thus what the adjective ordinarily means. But when the question is the meaning of a noun, like in *Harris*,²²⁴ a collocation search will fail to determine a word’s ordinary meaning if the word ordinarily appears alone. The COCA does not function like a search engine where the user can simply ask what the word means.²²⁵ Therefore, a user must be certain to select the proper search method.

After users conduct the search, they must determine how to review the concordance lines. If the number of concordance lines is manageable—as it was with *Rasabout*’s eighty-six concordance lines²²⁶—the user should review all of them. If, like *Harris*, the search returns thousands of concordance

Congress enacted the National Bank Act). *But see* Eskridge, *supra* note 5, at 1479 (“Statutes, however, should—like the Constitution and the common law—be interpreted ‘dynamically,’ that is, in light of their present societal, political, and legal context.”).

218. See CORPUS CONTEMP. AM. ENG., *supra* note 27 (stating that the COCA dates back to 1990).

219. Mouritsen, *supra* note 20, at 1967. The COHA contains over 400 million words from works that were published between 1810 and 2009. James C. Phillips et al., *Corpus Linguistics & Original Public Meaning: A New Tool to Make Originalism More Empirical*, 126 YALE L.J.F. 21, 26, 30–31 (2016), http://www.yalelawjournal.org/pdf/10.Phillips.MACROEDFINAL_izrkk2s8.pdf [<https://perma.cc/79RG-JZHN>]. Some scholars hope to begin using corpus linguistics to aid in originalist constitutional interpretation as well. See generally Lawrence M. Solan, *Can Corpus Linguistics Help Make Originalism Scientific?*, 126 YALE L.J.F. 57 (2016), http://www.yalelawjournal.org/pdf/Solanforwebsitepdf_tryz9ksn.pdf [<https://perma.cc/XAL2-ZYJY>] (concluding that using corpus linguistics “should bring the practice of originalist analysis closer to its goal of discovering original public meaning”).

220. Phillips et al., *supra* note 219, at 31.

221. See *supra* Section II.A.

222. See *supra* Section II.A.

223. See POGO Amicus Brief, *supra* note 55.

224. See *Harris*, 885 N.W.2d at 837–38 (clarifying that the plain language of the word “information” controls the resolution of the case).

225. See HUNSTON, *supra* note 36, at 3 (“Strictly speaking, a corpus by itself can do nothing at all, being nothing other than a store of used language. Corpus access software, however, can re-arrange that store so that observations of various kinds can be made.”).

226. *State v. Rasabout*, 356 P.3d 1258, 1281–82 (Utah 2015) (Lee, J., concurring).

lines,²²⁷ the user should apply the method described in Section II.B.²²⁸ Then users must utilize the KWIC²²⁹ to read through the search returns and observe how the word is used in context. This process is the heavy lifting of a COCA search, but it is necessary to conduct a complete analysis because data alone do not determine the meaning of a word.

To maximize the COCA's attribute of transparency, judges and lawyers should include the details of their searches and analyses in their opinions and briefs.²³⁰ Legal writers can even include the links to their searches.²³¹ These details will permit reviewing judges to scrutinize the analysis on appellate review.

CONCLUSION

The Michigan Supreme Court recognized the COCA's contribution to the search for ordinary meaning—every single justice signed onto an opinion using the COCA to interpret the statute. And the shortcomings of the *Harris* majority's analysis revealed the COCA's contribution to judicial transparency. This Note was made possible only because both opinions in *Harris* described their search method, linked to their search results, and detailed their analysis. This transparency permits appellate judges, lawyers, scholars, and, yes, law students to review the court's work. That the Michigan Supreme Court reached different conclusions after using the COCA does not show that the COCA's results are ambiguous; rather, this Note argues, one side got it right and the other got it wrong. This type of review will increase the value of the COCA as a tool for statutory interpretation as the legal community learns to use the database more effectively.

227. *People v. Harris*, 885 N.W.2d 832, 850 n.14 (Mich. 2016) (Markman, J., concurring in part and dissenting in part) (stating that the dissent's search returned 168,187 concordance lines).

228. See *supra* notes 145–149 and accompanying text.

229. See Mouritsen, *supra* note 35, at 197.

230. See, e.g., POGO Amicus Brief, *supra* note 55, apps. B–G (listing POGO's corpora analyses).

231. See, e.g., *Harris*, 885 N.W.2d at 839 n.33 (majority opinion); *Rasabout*, 356 P.3d at 1281 n.29 (Lee, J., concurring).