Washington University Law Review

Volume 93 | Issue 6

2016

A Quantitative Analysis of Writing Style on the U.S. Supreme Court

Keith Carlson

Michael A. Livermore

Daniel Rockmore

Follow this and additional works at: http://openscholarship.wustl.edu/law_lawreview

Part of the Law Commons

Recommended Citation

Keith Carlson, Michael A. Livermore, and Daniel Rockmore, *A Quantitative Analysis of Writing Style on the U.S. Supreme Court*, 93 Wash. U. L. Rev. 1461 (2016).
Available at: http://openscholarship.wustl.edu/law_lawreview/vol93/iss6/6

This Article is brought to you for free and open access by the Law School at Washington University Open Scholarship. It has been accepted for inclusion in Washington University Law Review by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.

A QUANTITATIVE ANALYSIS OF WRITING STYLE ON THE U.S. SUPREME COURT

KEITH CARLSON
MICHAEL A. LIVERMORE
DANIEL ROCKMORE*

ABSTRACT

This Article presents the results of a quantitative analysis of writing style for the entire corpus of US Supreme Court decisions. The basis for this analysis is the measure of frequency of function words, which has been found to be a useful "stylistic fingerprint" and which we use as a general proxy for the stylistic features of a text or group of texts. Based on this stylistic fingerprint measure, we examine temporal trends on the Court, verifying that there is a "style of the time" and that contemporaneous Justices are more stylistically similar to their peers than to temporally remote Justices. We examine potential "internal" causes of stylistic changes, and conduct an in-depth analysis of the role of the modern institution of the judicial clerk in influencing writing style on the Court. Using two different measures of stylistic consistency, one measuring intra-year consistency on the Court and the other examining inter-year consistency for individual Justices, we find evidence that the writing styles of individual Justices have become less consistent as clerks have taken on a greater role on the Court.

TABLE OF CONTENTS

INTRODUCTION 1462
I. COMPUTATIONAL ANALYSIS OF LEGAL TEXTS 1465
  A. *Content Analysis of Judicial Opinions* 1466
  B. *Non-Content Analysis: Why Style?* 1469
II. MEASURING STYLE 1473
  A. *Data* 1474

---

* Keith Carlson is a Ph.D. candidate (Computer Science) at Dartmouth College. Michael A. Livermore is an Associate Professor at the University of Virginia School of Law. Daniel Rockmore is the William H. Neukom 1964 Professor of Computational Science and Professor of Mathematics and Computer Science, Dartmouth College; and Member, External Faculty, The Santa Fe Institute. Our thanks to comments from Quinn Curtis, Michael Gilbert, Andrew Hayashi, Richard Hynes, Jed Stiglitz, and David Zaring.

Washington University Open Scholarship

B. Preliminary Analyses 1475
1. Productivity 1475
2. Friendliness 1478
3. Defensiveness 1481
C. The Stylistic Fingerprint and Similarity Scores 1483
III. TIME TRENDS IN JUDICIAL STYLE 1486
IV. POTENTIAL MECHANISMS 1491
A. Prior Decisions 1492
B. Partisan Affiliation 1495
C. Substantive Factors 1496
V. CLERK INFLUENCE 1497
A. Previous Studies 1498
B. Comparing Inter-Year Variability 1502
CONCLUSION 1509

INTRODUCTION

The written word is the medium through which the law travels: courts, agencies, and legislatures create law by producing text. In recent years, these legal texts have increasingly become available to the public in digital form. Together with advances in processing power, data storage, machine learning, and computational text analysis, the digitization of the law has opened a new frontier in empirical legal scholarship. A number of researchers have eagerly crossed over into this unexplored territory in search of new insights, methods, and questions.

This Article contributes to this exciting new enterprise by undertaking the first general quantitative investigation of writing style on the US Supreme Court. While judicial writing style often serves as fodder for commentary, it has rarely been subject to systematic study.[1] Systematic qualitative analysis is made difficult by the sheer bulk of the corpus, which prevents a human reader from digesting any more than a tiny sample. Perhaps for this reason, qualitative analysis of style tends to focus on the gems in judicial writing, examining the prominent writings of prominent Justices and neglecting the mine-run of workaday opinions.[2] Scholars have only relatively recently combined accessible digital versions of the corpus

1. See *infra* Part I.B for some notable examples.
2. *See, e.g.*, Richard H. Weisberg, *Law, Literature and Cardozo's Judicial Poetics*, 1 CARDOZO L. REV. 283 (1979) (providing qualitative discussion of the judicial writings of Justice Cardozo).

of judicial writing with the tools offered by computational text analysis to undertake quantitative analysis of style.[3]

Prior attempts to analyze judicial writing style in a quantitatve way have typically been based on relatively small datasets, and are limited to a small number of specific stylistic features.[4] This Article relies on a corpus of all cases in the US Supreme Court in the period from 1792 to 2008, compiled from publically available raw textual data that has been augmented with identifying information concerning the year, author, and opinion type.[5] In addition to examining specific stylistic features culled from the prior literature on judicial writing style, we deploy a general stylistic measure that serves as a proxy for what Judge Posner refers to as "style as signature."[6] This general style proxy is first used to examine the gradual change in writing style over time, and then to investigate potential hypotheses about sources of stylistic variation over time. Perhaps most controversially, we find that the institution of the modern clerkship appears to have had an important effect on judicial writing style on the Court, both in the consistency of writing style in individual chambers and in the consistency of writing style of the Court as an institution.

Our primary analysis relies on a commonly used measure of writing style based on the frequency of use of content-free words (also called "function words"). This measure provides a "useful stylistic fingerprint" and was used for a large-scale study of literary style executed by Hughes, Foti, Krakauer, and Rockmore.[7] This stylistic approach has its roots in statistical methods to address the problem of author attribution.[8] As will be discussed more thoroughly, our stylistic fingerprint measure allows for analysis of the similarity between texts or a group of texts, including texts that are grouped by time and by author. In our analysis, we use the stylistic fingerprint as a means of developing descriptive statistics and as the basis

3. *See infra* Part II.B.

4. *See infra* Parts II.B, IV.A. A study by Black and Spriggs of opinion length in the Supreme Court notably uses data covering the entire Court's existence. *See* Ryan C. Black & James F. Spriggs II, *An Empirical Analysis of the Length of U.S. Supreme Court Opinions*, 45 HOUS. L. REV. 621 (2008).

5. Jonathan Ashley, research librarian at the University of Virginia, was primarily responsible for identifying resources, collecting cases, and providing the markup needed for analysis. We are extremely grateful for his efforts.

6. Richard A. Posner, *Judges' Writing Styles (And Do They Matter?)*, 62 U. CHI. L. REV. 1421, 1425 (1995). Our stylistic measure is based on function word frequency. *See infra* Part II.C.

7. *See* James M. Hughes, Nicholas J. Foti, David C. Krakauer & Daniel N. Rockmore, *Quantitative Patterns of Stylistic Influence in the Evolution of Literature*, 109 PROC. NAT. ACAD. SCI. 7682, 7682 (2012) [hereinafter HFKR (2012)].

8. *Id.*

Washington University Open Scholarship

for testing a number of hypotheses concerning the evolution of judicial writing style in the Supreme Court.

We first address the general relationship between style and time. The starting place for this analysis is the intuitive hypothesis that there is a "style of a time" in the Court.[9] Stated somewhat more formally, the hypothesis is that as the distance in time between judicial writings increases, there is a lower likelihood that they will be stylistically similar. Our analysis finds that stylistic similarity decreases with distance in time, as expected.

We also examine potential mechanisms that could drive this robust temporal trend. We examine the possibility that the writing style of particularly influential Justices propagates over time, so that the most read and cited Justices tend to project style forward. Perhaps surprisingly, we do not generally find that being widely cited increases the stylistic similarity between a past Justice and members of the current Court.[10] We also examine the potential for party affiliation to have played a role in stylistic evolution. While initial analysis reveals some differences between Democratic-appointed and Republican-appointed Justices, these differences appear to be the result of the temporal trend (alongside the changing partisan balance on the Court over time) rather than the cause. We then examine whether substantive features of opinions influence their style. We find that there are robust stylistic differences between majority opinions and dissents, even when comparing the writings of the same Justice. The growth of dissents, and dissent-like writing styles, may account for some of the drift in writing style on the Court over time.

We finally investigate the influence that the modern institution of the judicial clerkship has on writing style on the Court. For each Justice, we define a measure of consistency as the similarity between a Justice's writings in one year and in all other years.[11] We then test the hypothesis that the number of law clerks that a Justice employs is negatively associated with intra-Justice stylistic consistency. Overall, we find evocative evidence that the substantive role that clerks now play on the Court has led to *decreasing* inter-year intra-Justice stylistic consistency,

9. *Id*. at 7685 (finding quantitative support for a "style of the time" in Western literature).

10. This analysis of the "current court" included Justice Scalia because it was carried out before his death on February 13, 2016. For our measure of influence, we use the tables produced in Montgomery N. Kosma, *Measuring the Influence of Supreme Court Justices*, 27 J. LEGAL STUD. 333 (1998).

11. Our measure is described in more detail in Part II.C.

while leading to *increasing* intra-year institutional stylistic consistency on the Court.

The methodology and findings presented here are meant to serve as a starting place for future quantitative text-based analysis of the law and legal writing style. In particular, this Article serves as an invitation for researchers to engage in a broader examination of the causal factors that affect judicial writing style; of stylistic trends in other governmental institutions, including Congress and agencies; and of the dynamic relationship between writing style in political institutions and the broader culture. As discourse continues its move into the digital realm—as new content is created and earlier content is digitized—it opens the door for a rich inquiry into the relationship between literature, law, and mass media, and the relative influence (stylistic and otherwise) of each of these domains on the others.

This Article proceeds in five Parts. Part I explains how our research contributes to existing literature on judicial writing style and the growing literature using computational text analysis as the basis for legal scholarship. Part II describes our data and methods and conducts some preliminary computational style analysis on our new data. Part III examines broad stylistic time trends on the Court to test the general hypothesis that contemporaneous Justices tend to exhibit more similarity in their writing styles. Part IV discusses potential mechanisms that may account for changes in style over time. Part V focuses on the influence of clerks on writing style on the Court, and specifically on the consistency of writing style exhibited by Justices. The conclusion summarizes and provides potential future directions for research using computational tools to analyze judicial writing style.

I. COMPUTATIONAL ANALYSIS OF LEGAL TEXTS

Formalized analysis of the content of judicial decisions in both the legal and social sciences literatures has experienced substantial growth in recent years, and new computational tools have contributed to that growth. Some initial forays into systematic analysis of judicial writing style are part of this broader trend. This Part discusses the reasons for taking judicial writing style as an object of inquiry as well as recent advances in using quantitative tools to analyze judicial writings.

Washington University Open Scholarship

A. Content Analysis of Judicial Opinions

The ability to analyze the content of legal writing, and often judicial writing, is the stock-in-trade of the legal profession.[12] Identifying relevant law, extracting legal principles, and applying those principles to factual circumstances is the basic work that lawyers are called on to do in their daily practice. Learning how to do that work is the basic goal of a legal education, and law students spend their time practicing and (hopefully) mastering this skill.

While this type of qualitative content analysis is as old as the legal profession, quantitative analysis of legal content is a newer animal. Hall and Wright provide an overview of the growth in systematic content analysis, explaining that the goal of this research is to ground empirical conclusions about law in "the community's ability to reproduce . . . findings rather than the author's rhetorical power."[13] Such analysis begins by defining a relevant set of cases, which are then analyzed by human readers who "code" the case for pre-identified content elements.[14] This information then serves as data for whatever analysis will be carried out, from analysis of ideology space on the Supreme Court to observations of relevant legal factors in classes of cases.[15]

Systematic content analysis of legal opinions is now a mainstay of legal and political science scholarship. The Supreme Court Database[16] and the U.S. Appeals Courts Database[17] together have generated scores of research projects.[18] Perhaps the most frequently discussed research concerns the relationship between judicial decision-making and "extra-legal" or "ideological" factors,[19] but content analysis has been deployed to

---

12. Law students may start with the sensation that reading judicial writing is akin to "stirring concrete with [their] eyelashes," but most of them ultimately gain some degree of fluency in their newly acquired language. SCOTT TURROW, ONE L 31 (2d ed. 1988).

13. Mark A. Hall & Ronald F. Wright, *Systematic Content Analysis of Judicial Opinions*, 96 CALIF. L. REV. 63, 66 (2008).

14. *Id.*

15. *See generally* DOUGLAS LAYCOCK, THE DEATH OF THE IRREPARABLE INJURY RULE (1991) (conducting exhaustive search of cases concerning specific legal doctrine); Andrew D. Martin & Kevin M. Quinn, *Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953–1999*, 10 POL. ANALYSIS 134 (2002) (developing novel measure of judicial ideology).

16. THE SUP. CT. DATABASE, http://supremecourtdatabase.org (last visited Apr. 5, 2016).

17. *U.S. Appeals Courts Database*, THE JUD. RESEARCH INITIATIVE, http://artsandsciences.sc.edu/poli/juri/appct.htm (last visited Mar. 31, 2016).

18. *See* Lee Epstein, Jack Knight & Andrew D. Martin, *The Political (Science) Context of Judging*, 47 ST. LOUIS U. L.J. 783, 812 n.85 (2003) (noting reliance on Spaeth databases).

19. *See generally* LEE EPSTEIN ET AL., THE BEHAVIOR OF FEDERAL JUDGES: A THEORETICAL AND EMPIRICAL STUDY OF RATIONAL CHOICE (2013); Harry T. Edwards & Michael A. Livermore,

examine questions as far-flung as judicial deference to agency decisions[20] and the differences between state and federal court interpretation of the US Constitution.[21] Systematic legal content analysis also extends beyond courts, and has examined, for example, the effect of public comments on the decisions of administrative agencies.[22]

Some newer approaches to systematic content analysis substitute computational approaches for human readers.[23] Oldfather, Bockhorst, and Dimmer use an automated approach to investigate the "responsiveness" of appellate court decisions to briefs.[24] Calvin, Collins, and Corley use computational methods to examine the influence of parties' briefs and lower court decisions on Supreme Court opinions.[25] Smith uses machine-reading techniques to attempt to distinguish fact-intensive from law-intensive cases.[26] Cheng et al. use automated content analysis to identify the contested values in congressional testimony and agency hearings over net neutrality regulation.[27] Katz et al. apply an n-gram analysis to examine word frequency within the Supreme Court corpus, creating an online tool

---

Pitfalls of Empirical Studies That Attempt to Understand the Factors Affecting Appellate Decisionmaking, 58 DUKE L.J. 1895 (2009).

20. Christopher H. Schroeder & Robert L. Glicksman, Chevron, State Farm, *and EPA in the Courts of Appeals During the 1990s*, 31 ENVTL. L. REP. 10,371 (2001).

21. Brett Christopher Gerry, *Parity Revisited: An Empirical Comparison of State and Lower Federal Court Interpretations of* Nollan v. California Coastal Commission, 23 HARV. J.L. & PUB. POL'Y 233 (1999).

22. Keith Naughton et al., *Understanding Commenter Influence During Agency Rule Development*, 28 J. POL'Y ANALYSIS & MGMT. 258 (2009).

23. *See generally* An-Shou Cheng, Kenneth R. Fleischmann, Ping Wang & Douglas W. Oard, *Advancing Social Science Research by Applying Computational Linguistics*, 45 PROC. AM. SOC'Y INFO. SCI. & TECH. 1 (2008); Michael Evans, Wayne McIntosh, Jimmy Lin & Cynthia Cates, *Recounting the Courts? Applying Automated Content Analysis to Enhance Empirical Legal Research*, 4 J. EMPIRICAL LEGAL STUD. 1007 (2007). Computational approaches to content analysis also, of course, occur outside of the legal context. *See, e.g.*, David Lazer et al., *Computational Social Science*, 323 SCIENCE 721 (2009) (discussing potential for computational analysis to provide new methods of social science research); Rada Mihalcea, Courtney Corley & Carlo Stapparava, *Corpus-based and Knowledge-based Measures of Text Semantic Similarity*, 1 PROC. 21ST NAT'L CONF. ON ARTIFICIAL INTELLIGENCE 775, 778 (2006) (applying computational analysis to database of summaries of news stories).

24. Chad M. Oldfather, Joseph P. Bockhorst & Brian P. Dimmer, *Triangulating Judicial Responsiveness: Automated Content Analysis, Judicial Opinions, and the Methodology of Legal Scholarship*, 64 FLA. L. REV. 1189, 1189 (2012).

25. Pamela C. Corley, Paul M. Collins, Jr. & Bryan Calvin, *Lower Court Influence on U.S. Supreme Court Opinion Content*, 73 J. POL. 31 (2011); Pamela C. Corley, *The Supreme Court and Opinion Content: The Influence of Parties' Briefs*, 61 POL. RES. Q. 468 (2008).

26. Joseph L. Smith, *Law, Fact, and the Threat of Reversal from Above*, 42. AM. POL. RES. 226 (2014).

27. An-Shou Cheng, Emi Ishita, Douglas W. Oard, Kenneth R. Fleischmann & Ping Wang, *The Role of Innovation and Wealth in the Net Neutrality Debate: A Content Analysis of Human Values in Congressional and FCC Hearings*, 63 J. AM. SOC'Y INFO. SCI. & TECH. 1360 (2012).

Washington University Open Scholarship

that allows any interested person to examine word frequency by year.[28] Stiglitz analyzes word frequency in the Federal Register to test a rule's level of controversy.[29] Rice tests how well statistical "topic models" compare to the human coders for the Supreme Court Database.[30] Livermore, Riddell, and Rockmore exploit topic models to test hypotheses about agenda formation on the Supreme Court.[31] Together, these scholars are placing computational methods firmly in the toolkit of legal studies.

The study of judicial citation practices is also related to content analysis, although citation does not cleanly fall on one side of the style-substance divide.[32] Citation studies have examined the use of foreign law[33] and legislative history,[34] and there is a growing body of literature that examines case citation.[35] In particular, citation has been used to examine the relative influence of judges and judicial opinions.[36] New computational tools have allowed judicial citations to serve as the basis for network analysis.[37]

---

28. Daniel Martin Katz et al., *Legal N-Grams? A Simple Approach to Track the Evolution of Legal Language*, 235 FRONTIERS IN ARTIFICIAL INTELLIGENCE & APPLICATIONS 167 (2011).

29. Edward H. Stiglitz, *Unaccountable Midnight Rulemaking? A Normatively Informative Assessment*, 17 LEGIS. & PUB. POL'Y 137 (2014).

30. Douglas Rice, Measuring the Issue Content of Supreme Court Opinions Through Probabilistic Topic Models (unpublished manuscript) (presented at the 2012 Midwest Political Science Association Conference, Chicago, Ill., Apr. 13, 2012), *available at* http://static1.squarespace.com/static/5557a550e4b0443afbea6783/t/55dc7c17e4b05465d92e94ad/1440513070923/issueMeasures.pdf.

31. Michael A. Livermore, B. Allen Riddell & Daniel Rockmore, *Agenda Formation and the U.S. Supreme Court: A Topic Model Approach* (Va. Law & Econ. Research Paper No. 2, 2016), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2740126. *See also* David S. Law, *Constitutional Archetypes*, 95 TEX. L. REV. (forthcoming 2017) (using topic models to examine constitutional preambles).

32. Using Judge Posner's formulation, *see* Posner, *supra* note 6, at 1425, it is not clear whether and how much citation would need to be included in a "paraphrase" of a legal decision for it to be accurate. *See generally* Tom S. Clark & Benjamin Lauderdale, *Locating Supreme Court Opinions in Doctrine Space*, 54 AM. J. POL. SCI. 871 (2010) (using citation to develop more nuanced ideological variables for decisions).

33. David Zaring, *The Use of Foreign Decisions by Federal Courts: An Empirical Analysis*, 3 J. EMPIRICAL LEGAL STUD. 297 (2006).

34. David S. Law & David Zaring, *Law Versus Ideology: The Supreme Court and the Use of Legislative History*, 51 WM. & MARY L. REV. 1653 (2010) (quantitative analysis of citation to legislative history).

35. *See, e.g.*, Stephen J. Choi & G. Mitu Gulati, *Bias in Judicial Citations: A Window into the Behavior of Judges?*, 37 J. LEGAL STUD. 87 (2008); Yonatan Lupu & James H. Fowler, *Strategic Citations to Precedent on the U.S. Supreme Court*, 42 J. LEGAL STUD. 151 (2013); Richard A. Posner, *An Economic Analysis of the Use of Citations in the Law*, 2 AM. L. & ECON. REV. 381 (2000).

36. *See, e.g.*, Frank B. Cross, *The Ideology of Supreme Court Opinions and Citations*, 97 IOWA L. REV. 693 (2012); Frank B. Cross & James F. Spriggs II, *The Most Important (and Best) Supreme Court Opinions and Justices*, 60 EMORY L.J. 407 (2010); *see also* Kosma, *supra* note 10.

37. *See, e.g.*, Thomas A. Smith, *The Web of Law*, 44 SAN DIEGO L. REV. 309 (2007).

B. Non-Content Analysis: Why Style?

Just as analysis of the content of judicial decisions is the routine work of the legal profession, judges, lawyers, legal academics, and law students have frequently turned their attention to non-content, stylistic features of legal writing as well.[38] Practicing lawyers are often called on to persuade through the written word, and stylistic features of a text can contribute to (or detract from) its persuasive force. Guides on legal writing, geared toward law students and practicing attorneys, often pay substantial attention to non-content textual characteristics.[39] A host of stylistic conventions distinguish legal writing from standard written English,[40] and a lawyer's competence is judged, in part, by the degree to which his or her individual stylistic voice conforms to this particular "professional discourse community."[41]

In literary analysis, style refers to the distinctive characteristics of an author's writings.[42] The juxtaposition of style and substance is sometimes used to define style as all of the characteristics of a text other than the substance.[43] This is the sense that Judge Posner invokes when he defines

---

38. For example, the adoption of legal writing courses at American law schools evidences a desire to teach students appropriate writing style, in addition to facilitating a mastery of legal content. Of course, legal writing courses serve many purposes simultaneously, some of which can be content oriented. *See* David S. Romantz, *The Truth About Cats and Dogs: Legal Writing Courses and the Law School Curriculum*, 52 U. KAN. L. REV. 105, 127–36 (2003) (providing a brief overview of the history of legal writing courses at American law schools); *cf.* MARK MCGURL, THE PROGRAM ERA: POSTWAR FICTION AND THE RISE OF CREATIVE WRITING 77–125 (2011) (linking the post-World War II rise of creative writing programs to the development of American literature in the same period).

39. *See generally* BRYAN A. GARNER, LEGAL WRITING IN PLAIN ENGLISH: A TEXT WITH EXERCISES (2d ed. 2013); BRYAN A. GARNER, THE WINNING BRIEF: 100 TIPS FOR PERSUASIVE BRIEFING IN TRIAL AND APPELLATE COURTS (3d ed. 2014); ROSS GUBERMAN, POINT MADE: HOW TO WRITE LIKE THE NATION'S TOP ADVOCATES (2011); *see also* Bret Rappaport, *Using the Elements of Rhythm, Flow, and Tone to Create a More Effective and Persuasive Acoustic Experience in Legal Writing*, 16 J. LEGAL WRITING INST. 65 (2010). For earlier examples of stylistic advice for judges in drafting opinions, see Griffin B. Bell, *Style in Judicial Writing*, 15 J. PUB. L. 214 (1966); Robert A. Leflar, *Some Observations Concerning Judicial Opinions*, 61 COLUM. L. REV. 810 (1961).

40. *Cf.* Marilyn Warren, *The Obligation to Communicate: The Interaction Between Language and the Law*, 15 DEAKIN L. REV. 117, 127 (2010) ("Judges do not use Powerpoint presentations, signs, billboards, e-mail, blogs or Twitter to communicate their decisions.").

41. Andrea McArdle, *Teaching Writing in Clinical, Lawyering, and Legal Writing Courses: Negotiating Professional and Personal Voice*, 12 CLINICAL L. REV. 501, 501 (2006).

42. Stylistic analysis can allow for the attribution of a piece to an individual when authorship is in doubt. There is a long lineage of quantitative analysis of style for attribution purposes. *See* HFKR (2012), *supra* note 7, at 7682 (citing quantitative methods of stylistic analysis that date from the nineteenth century for use in dating Plato's dialogues).

43. "There are two things wrong with almost all legal writing. One is its style. The other is its content. That, I think, about covers the ground." Fred Rodell, *Goodbye to Law Reviews*, 23 VA. L. REV. 38, 38 (1936).

Washington University Open Scholarship

style "as the range of options for encoding the paraphrasable content of a writing."[44]

There is substantial stylistic variation within legal writing. At the level of institutions, although a legal question might be resolved through a statute, opinion, or regulation, each lawmaking form follows particular conventions of structure, length, and organization that allow the reader to readily distinguish the writing of a judge from the writing of an agency rule-writer or legislative drafter. The structure of Supreme Court opinions has evolved from the seriatim opinions of earlier days to the multipart, multiple concurrence opinions of more recent years.[45] Style also differs between judges. Some Justices are lionized for their writing style, others are mocked, and most are ignored.[46] The writing style of individual Justices can sometimes be identified by human readers, and quantitative tools for attributing writings to specific Justices have been developed.[47]

Although style is distinct from substance, it can still matter. Style may serve as an indicator of judicial temperament or disposition.[48] Stylistic norms may constrain judicial writing in ways that ultimately affect judicial reasoning, and in turn, legal outcomes.[49] The evolution of writing style may indicate broader substantive trends on the Court. Style can affect the comprehensibility and usability of the law.[50] Finally, style may be deserving of study simply as an empirical feature of an important cultural artifact.

---

44. Posner, *supra* note 6, at 1423.

45. *See generally* WILLIAM D. POPKIN, EVOLUTION OF THE JUDICIAL OPINION: INSTITUTIONAL AND INDIVIDUAL STYLES (2007) (conducting comprehensive analysis of evolution of form of judicial writing in the United States).

46. *See, e.g.*, Jeffrey Rosen, *Strong Opinions*, NEW REPUBLIC, Aug. 2011, *available at* http://www.newrepublic.com/article/politics/magazine/92773/elena-kagan-writings (praising Justice Kagan's writing style).

47. *See, e.g.*, William Li et al., *Using Algorithmic Attribution Techniques to Determine Authorship in Unsigned Judicial Opinions*, 16 STAN. TECH. L. REV. 503 (2013).

48. *See, e.g.*, Posner, *supra* note 6, at 1421 (suggesting a link between "pure" and "impure" writing styles and "formalist" and "pragmatist" judicial approaches); *see also* Mary L. Dudziak, *Oliver Wendell Holmes as a Eugenic Reformer: Rhetoric in the Writing of Constitutional Law*, 71 IOWA L. REV. 833, 834 (1986) ("[E]xamin[ing] the role of language as an expression of ideology . . . ."); Warren D. TenHouten, *Inferring Cognitive Structure from Text*, 6 J. CONTEMP. LEGAL ISSUES 97, 100–02 (1995) (discussing research on how neurological conditions affect stylistic choices in language use); Philip E. Tetlock, Jane Bernzweig & Jack L. Gallant, *Supreme Court Decision Making: Cognitive Style as a Predictor of Ideological Consistency of Voting*, 48 J. PERSONALITY & SOC. PSYCHOL. 1227, 1230–31 (1985) (attempting to link opinion language to personality characteristics of authoring Justices).

49. Chad M. Oldfather, *Writing, Cognition, and the Nature of the Judicial Function*, 96 GEO. L.J. 1283, 1286 (2008) (discussing benefits and problems of the practice of writing opinions for the quality of judicial reasoning).

50. *See, e.g.*, Frederick Schauer, *Opinions as Rules*, 62 U. CHI. L. REV. 1455, 1455–57 (1995).

Judicial writing style has long been the object of qualitative analysis.[51] Commentators bring different standards to this enterprise, with some focused on the literary quality of a judge's work while others focus on more quotidian characteristics.[52] Observers of the contemporary Supreme Court lament the lost "play of intelligence"[53] and the "decline and fall of the American judicial opinion."[54] Others have proposed ethical limitations on judicial writing style.[55] The writing styles of specific judges have been the subject of careful qualitative analysis.[56] Linguists have examined "legal English as a sub-language with its own style, syntax and terminology" for several decades.[57] Specific rhetorical characteristics of judicial opinions have been parsed in close detail.[58]

In the field of literary studies, quantitative tools have been brought to bear on the analysis of style for decades.[59] Early work in this area was

---

51. *See* Robert F. Blomquist, *Playing on Words: Judge Richard A. Posner's Appellate Opinions, 1981-82—Ruminations on Sexy Judicial Opinion Style During an Extraordinary Rookie Season*, 68 U. CIN. L. REV. 651, 673–83 (2000) (surveying earlier discussions of judicial writing style); *see also* Ray Forrester, *Supreme Court Opinions—Style and Substance: An Appeal for Reform*, 47 HASTINGS L.J. 167, 173–80 (1995).

52. A University of Chicago symposium on the subject drew various candidates for the appropriate standard to judge judicial writing. *See* Martha C. Nussbaum, *Poets as Judges: Judicial Rhetoric and the Literary Imagination*, 62 U. CHI. L. REV. 1477 (1995); Posner, *supra* note 6; Schauer, *supra* note 50; Patricia M. Wald, *The Rhetoric of Results and the Results of Rhetoric: Judicial Writings*, 62 U. CHI. L. REV. 1371 (1995); James Boyd White, *What's an Opinion For?*, 62 U. CHI. L. REV. 1363 (1995); *see also* WILLIAM DOMNARSKI, IN THE OPINION OF THE COURT (1996) (treating judicial opinions as literary genre); Robert A. Ferguson, *The Judicial Opinion as Literary Genre*, 2 YALE J.L. & HUMAN. 201 (1990); Walker Gibson, *Literary Minds and Judicial Style*, 36 N.Y.U. L. REV. 915 (1961) (providing reflections on commonality between legal writing and literary writing, and the benefits to scholars in both areas of conversing across disciplines).

53. Daniel A. Farber, *Missing the "Play of Intelligence,"* 36 WM. & MARY L. REV. 147, 148 (1994). For a more recent lament along the same lines, see Michael Serota, *Intelligible Justice*, 66 U. MIAMI L. REV. 649 (2012).

54. Jeffrey A. Van Detta, *The Decline and Fall of the American Judicial Opinion, Part I: Back to the Future from the Roberts Court to Learned Hand—Context and Congruence*, 12 BARRY L. REV. 53, 53 (2009).

55. *See, e.g.*, Gerald Lebovits, Alifya V. Curtin & Lisa Solomon, *Ethical Judicial Opinion Writing*, 21 GEO. J. LEGAL ETHICS 237, 237 (2008); David McGowan, *Judicial Writing and the Ethics of the Judicial Office*, 14 GEO. J. LEGAL ETHICS 509, 513–15 (2001).

56. *See, e.g.*, ROBERT F. BLOMQUIST, THE QUOTABLE JUDGE POSNER 4–7 (2010); RICHARD A. POSNER, CARDOZO: A STUDY IN REPUTATION 33–57 (1990).

57. Davide Simone Giannoni & Celina Frade, *Introduction*, *in* RESEARCHING LANGUAGE AND THE LAW: TEXTUAL FEATURES AND TRANSLATION ISSUES 7, 8 (Davide Simone Giannoni & Celina Frade eds., 2010); *see also* Judith N. Levi, *"What Is Meaning in a Legal Text?" A First Dialogue for Law and Linguistics*, 73 WASH. U. L.Q. 771 (1995).

58. *See, e.g.*, HAIG BOSMAJIAN, METAPHOR AND REASON IN JUDICIAL OPINIONS (1992); James Boyd White, *Law as Rhetoric, Rhetoric as Law: The Arts of Cultural and Communal Life*, 52 U. CHI. L. REV. 684 (1985).

59. Richard W. Bailey, *The Future of Computational Stylistics*, *in* LITERARY COMPUTING AND LITERARY CRITICISM 3, 5 (Rosanne G. Potter ed., 1989) (originally presented at the Association for Literary and Linguistic Computing Fifty International Meeting, Friday, December 15, 1978, King's

Washington University Open Scholarship

facilitated by basic statistics packages geared toward humanities scholars, such as LitStats.[60] Computational analysis of style can be deployed on the level of the individual author or work or across wide swaths of literature, which is effectively the computational study of genre.[61]

In the legal literature, there is a nascent movement to replicate this trend.[62] An early important application of computational stylistic analysis by Mosteller and Wallace was directed at the attribution of the Federalist Papers.[63] Little uses a coding procedure to identify "linguistic devices that obscure" meaning and analyzes Supreme Court cases on federal jurisdiction.[64] Owens and Wedeking examine "cognitive clarity" in recent Supreme Court cases using the "linguistic inquiry and word court" ("LIWC") software package.[65] Long and Christensen examine the use of "intensifiers" and readability to test their theory that Justices broadcast weak legal position through use of language.[66] Johnson examines readability over time in the Court, comparing Flesch-Kincaid scores in the 1931–1933 and 2009–2011 terms,[67] while Black and Spriggs examine

College, University of London) (discussing "[a]pplication of the computer to problems of style" in literary analysis). The other essays in the same volume apply computational text analysis to the works of Milton, Chaucer, and Henry James, *inter alia*. *Id.*; *see also* Estelle Irizarry, *Literary Analysis and the Microcomputer*, 71 HISPANIA 984, 984 (1988) (discussing efforts in the 1950s and 1960s to computationally analyze literary texts on "huge mainframe computers" that processed "data on punched cards"; also noting that Jonathan Swift imagined computational text analysis in *Gulliver's Travels*).

60. *See* Estelle Irizarry, *LitStats, Version 1.62*, 25 COMPUTERS & HUMAN. 455, 455 (1991) (lauding LitStats as "[u]ncomplicated, accurate, and easy to use").

61. *See generally* FRANCO MORETTI, DISTANT READING (2013) (discussing program of applying quantitative text analysis to questions concerning literary criticism); *see also* MATTHEW L. JOCKERS, MACROANALYSIS: DIGITAL METHODS & LITERARY HISTORY 17–20 (2013) (discussing potential for computational analysis of literature).

62. For an early example of quantitative analysis of legal writing style, see generally Lawrence M. Friedman, Robert A. Kagan, Bliss Cartwright & Stanton Wheeler, *State Supreme Courts: A Century of Style and Citation*, 33 STAN. L. REV. 773 (1981) (examing stylistic features such as citation frequency and opinion length).

63. Frederick Mosteller & David L. Wallace, *Inference in an Authorship Problem*, 58 J. AM. STAT. ASS'N 275, 275 (1963); *see also* FREDERICK MOSTELLER & DAVID L. WALLACE, APPLIED BAYESIAN AND CLASSICAL INFERENCE: THE CASE OF THE FEDERALIST PAPERS (2d ed. 1984).

64. Laura E. Little, *Hiding with Words: Obfuscation, Avoidance, and Federal Jurisdiction Opinions*, 46 UCLA L. REV. 75, 80 (1998). Incidental quantitative examination of writing style is sometimes included in more content-focused analyses. *See* Peter H. Schuck & E. Donald Elliott, *To the* Chevron *Station: An Empirical Study of Federal Administrative Law*, 1990 DUKE L.J. 984, 1003–04, 1070–74 (includes information on opinion length in administrative law cases).

65. Ryan J. Owens & Justin P. Wedeking, *Justices and Legal Clarity: Analyzing the Complexity of U.S. Supreme Court Opinions*, 45 LAW & SOC'Y REV. 1027, 1038–39 (2011).

66. Lance N. Long & William F. Christensen, *When Justices (Subconsciously) Attack: The Theory of Argumentative Threat and the Supreme Court*, 91 OR. L. REV. 933, 935–36 (2013).

67. Stephen M. Johnson, *The Changing Discourse of the Supreme Court*, 12 U. N.H. L. REV. 29, 31 (2014).

opinion length over the entire period of the Court's existence.[68] Osenga analyzes patent claims to determine how they have changed in length and complexity over time, finding that length has remained relatively stable despite increasing complexity.[69] As discussed in more detail below, there have been several attempts to study the influence of clerks over opinion drafting through computational analysis of writing style.[70] Computational analysis of judicial writing style has even found its way into pop culture: Chilton, Jiang, and Posner use token analysis—a measure of sophistication in language use—to compare the vocabulary of several Justices to famous rappers and Shakespeare in a blog post on *Slate*.[71] They find that Jay Z and most of the Justices have similar vocabulary use, while rapper Aesop Rock and Justice Holmes have exceptionally large vocabulary use, and DMX and Justice Kennedy are on the low end.[72]

This Article adds to the existing literature in several ways. Our analysis uses more extensive data than prior research: all opinions of the Supreme Court from 1791 to 2008. With this larger data set, we are able to test and extend the earlier quantitative analyses. We also use a more general stylistic measure rather than analyzing specific elements of style. This measure allows us to examine style at a high level of abstraction, rather than restrict analysis to a specific stylistic feature in judicial writings. Finally, we are able to pair our data with existing content coding for a portion of our study period, allowing analysis of the interaction of judicial writing style and legal substance.

II. MEASURING STYLE

This Part describes our data and methodology and re-examines some of the prior work in computational style analysis based on our newly created dataset.

---

68. *See* Black & Spriggs, *supra* note 4, at 630.

69. Kristen Osenga, *The Shape of Things to Come: What We Can Learn from Patent Claim Length*, 28 SANTA CLARA COMPUTER & HIGH TECH. L.J. 617, 634 (2012).

70. *See infra* Part V.A.

71. Adam Chilton, Kevin Jiang & Eric Posner, *Rappers v. SCOTUS: Who Uses a Bigger Vocabulary, Jay Z or Scalia?*, SLATE (June 12, 2014), http://www.slate.com/articles/news_and_politics/view_from_chicago/2014/06/supreme_court_and_rappers_who_uses_a_bigger_vocabulary_jay_z_or_scalia.html (taking the number of unique words over the total vocabulary as the measure of sophistication).

72. *Id.*

Washington University Open Scholarship

A. Data

Computational analysis of legal texts is hampered by difficulties accessing the relevant data.[73] While judicial opinions are not copyrightable, the commercial databases that provide ready digital access to these opinions are protected by terms of use agreements.[74] Limits on machine reading may be necessary to protect the proprietary content that has been produced by these publishers, but they can also inhibit academic research and access to the non-copyrighted government documentary information included within these resources.

Public.Resource.Org, a private nonprofit corporation, has created a digital version of the Supreme Court and federal appellate court corpus based on the non-copyrightable information within the Westlaw database, and it published that information online at "bulk.resource.org."[75] The bulk resource data has been used in prior n-gram studies of text usage in the federal courts and Supreme Court,[76] and it provides the public with access to a digital version of the nation's judicial opinions. However, the bulk resource data has some important limitations, including a lack of readily identifiable author and date information.

Because our analysis was limited to the Supreme Court, which has a relatively manageable universe of approximately 25,000 decisions, we were able to augment the Public.Resource.Org data to generate a new dataset. Human researchers conducted a series of "by year" searches on a commercial database to download digitized versions of all Supreme Court cases. All proprietary information was stripped out. Next, a series of iterative human and Python-based analyses were carried out to separate majority, dissenting, and concurring opinions and to assign an authoring Justice and year to each opinion.[77] Per curiam decisions were removed from the dataset, as were opinions with a file size smaller than one

---

73. *See* Olufunmilayo B. Arewa, *Open Access in a Closed Universe: Lexis, Westlaw, Law Schools, and the Legal Information Market*, 10 LEWIS & CLARK L. REV. 797, 797 (2006) (discussing consequences of the "competitive duopoly" enjoyed by commercial legal publishers).

74. *See, e.g.*, *Terms of Use*, THOMSON REUTERS, http://legalsolutions.thomsonreuters.com/law-products/about/legal-notices/terms-of-use (last visited Mar. 31, 2016).

75. BULK.RESOURCE.ORG, https://bulk.resource.org/ (last visited Mar. 31, 2016).

76. *See* Katz et al., *supra* note 28, at 2–3.

77. Python is an open source, general-purpose programming language that can be used for a variety of programming tasks. *See generally What is Python? Executive Summary*, PYTHON, https://www.python.org/doc/essays/blurb/ (last visited Mar. 31, 2016).

kilobyte. Data concerning the number of clerks employed in chambers was provided by the Supreme Court Library.[78]

The resulting data covers all opinions for the years 1792 to 2008. Our data includes 25,407 decisions.[79] We exclude footnotes from our analysis. There are roughly 8,000 dissents and 4,600 concurrences. We have data for 110 Justices: Justices Sotomayor and Kagan were appointed after the end of our study period. We have partial data for Justices who began their terms prior to 2008 but either retired after our study period or remain on the Court.[80] Our analysis was conducted when Justice Scalia was an active member of the Court.

B. Preliminary Analyses

Before introducing the primary stylistic metric that is used for the balance of our analysis, we report the results from three preliminary analyses.

1. Productivity

Our first analysis examines the "productivity" of each Justice, as measured by the total number of words authored by that Justice in all of their opinions. Figure 1 presents the number of words produced by each Justice, with each Justice located on the horizontal axis according to his or her median year on the Court. An ordinary least squares ("OLS") analysis comparing Justices' production and their median year of service produced highly significant results.[81] More recent Justices tend to produce more

78. The Supreme Court of the United States Library provided an unofficial list of clerks that was used as the basis for our analysis.

79. We define a "decision" as the set of opinions that relate to a case, identifiable through a citation in the United States Reporter, for example, "347 U.S. 483 (1954)." A decision can include multiple opinions, including a majority opinion, plurality opinions, and one or more dissents or concurrences. In our data, we do not distinguish majority from plurality opinions.

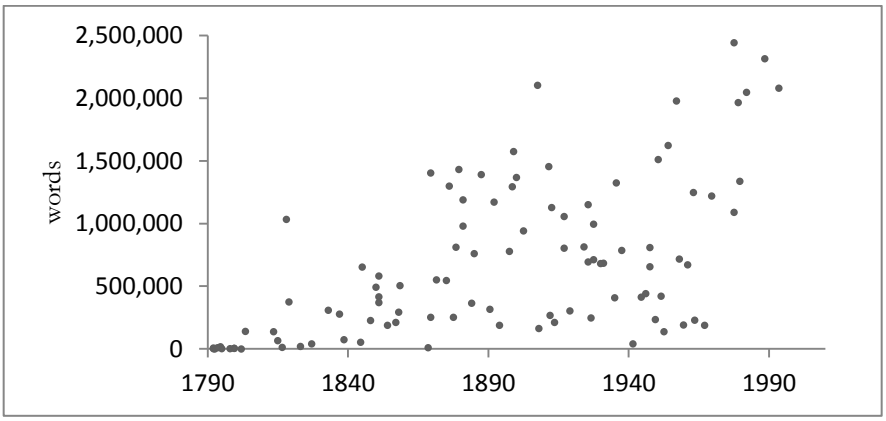
80. These Justices are Samuel Alito, Stephen Breyer, Ruth Bader Ginsburg, Anthony Kennedy, John Roberts, Antonin Scalia, David Souter, John Paul Stevens, and Clarence Thomas.

81. The coefficient is 6,831, the R-squared value is 0.32, and the p-value is less than 0.01%. Throughout this Article, we will conduct a series of very basic statistical tests on our data, including a number of ordinary least squares regression analyses. For a general introduction to regression analysis, see Alan O. Sykes, *An Introduction to Regression Analysis*, *in* CHICAGO LECTURES IN LAW AND ECONOMICS (Eric A. Posner ed., 2000), *available at* http://www.law.uchicago.edu/files/files/20.Sykes_.Regression.pdf. These statistical tests should be understood as a first cut analysis, meant to identify general relationships in writing style that are illuminated by the data and stylistic measure that we use.

Washington University Open Scholarship

total words than Justices that served in earlier periods.[82] The analysis of productivity excludes Justices that are currently sitting and the two non-sitting Justices who left after the end of the study period (Justices Souter and Stevens), leaving a total of 101 observations.

FIGURE 1: PRODUCTIVITY OVER TIME
(EXCLUDING SITTING JUSTICES)



There are many factors that could account for the growth in productivity over time, including longer average opinions, more opinions produced per year, and longer Justice length-of-service. Black and Spriggs provide a detailed treatment of time trends associated with opinion length.[83] They find that while the number of decisions has declined since peaking around the turn of the century, concurrences and dissents have become much more prevalent.[84] They also find that average opinion length tends to go through cyclical patterns.[85] The cyclical pattern identified by Black and Spriggs was growth from 1790 with trend reversals in 1830, 1870, 1900, and 1940, and a final period of growth thereafter.[86]

82. Earlier working paper versions of this Article (posted online on February 23, 2015, and March 11, 2015) carried out analysis of output examining only majority opinions, finding a coefficient of roughly 4,000.

83. Black & Spriggs, *supra* note 4, at 632–38.

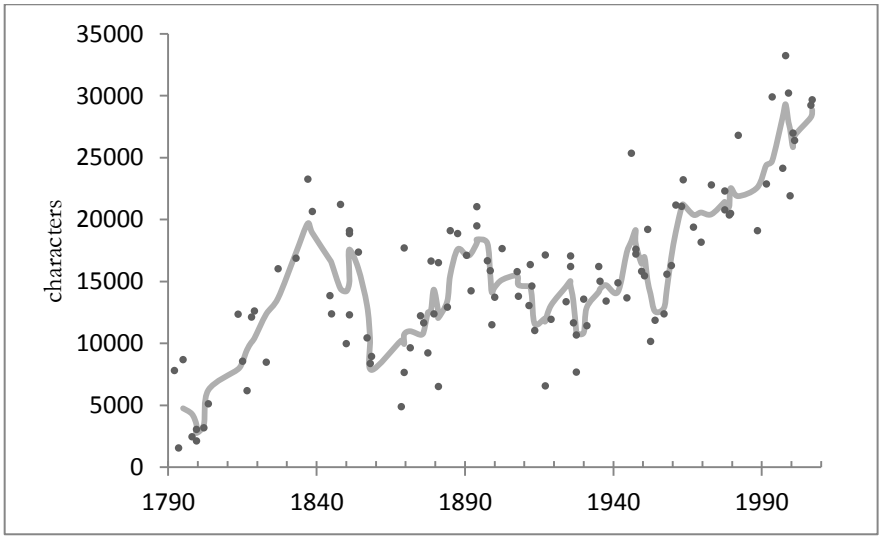
84. *Id.* at 633 & fig.1.

85. *Id*. at 635 & fig.2.

86. *Id*.

Figure 2 presents an analysis of average opinion length by Justice, ordered by their median year on the Court.[87] There is a general time trend in average length.[88] The cyclical pattern identified by Black and Spriggs is roughly present (presented in Figure 2 as a four-year moving average) around a general trend of growth. It should be noted that while the time trend noted by Black and Spriggs was for opinion length by year (in words), the analysis presented here is the average opinion length by Justice (in characters), with the median year of service of the Justice as the explanatory variable.

FIGURE 2: TIME TRENDS IN OPINION LENGTH



87. This analysis examines majority opinions only, excluding three Justices who authored only dissents or concurrences (Blair, Iredell, and Thomas Johnson). Justices with partial data are included, for a total of 107 observations.

88. The p-value for a simple linear time trend is less than 0.01%; the R-squared is 0.49. We examined variability in opinion length as well, finding no statistically significant time trend in the standard deviation of a Justice's opinion length.

Washington University Open Scholarship

2. *Friendliness*

Our next analysis examines the "friendliness" of each Justice, as measured by his or her use of positive and negative words. This analysis is based on a list of words constructed to examine the "sentiment" of written texts.[89] Positive and negative words have been used to evaluate online reviews, among other texts, and analyzing their use has generally been found to be a useful means of engaging in computational analysis of large text corpora to "determine[] whether a document or sentence is opinionated, and if so whether it carries a positive or negative opinion."[90] Some examples of negative words are "admonish" and "problematic." Positive words include "adventurous" and "preeminent." Together, there are around 7,000 words characterized as either positive or negative.[91]

A Python script was programmed to determine for each Justice the total number of negative words and the total number of positive words in opinions authored. The numbers of negative and positive words were then each expressed as percentages of the total number of words authored by a Justice. The percentage of negative words was subtracted from the percentage of positive words to generate what we call a "friendliness score."

This analysis—while based on measures of sentiment that have been used in a variety of other contexts—should be approached with a healthy dose of skepticism. Comparing texts over a long time horizon may be problematic for a variety of reasons, including that a text that reads as

89. Bing Liu & Minqing Hu, *Opinion Mining, Sentiment Analysis, and Opinion Spam Detection*, UIC, http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html (last visited Mar. 31, 2016); *see also* Minqing Hu & Bing Liu, *Mining and Summarizing Customer Reviews*, 10 PROC. ACM SIGKDD INT'L CONF. ON KNOWLEDGE DISCOVERY & DATA MINING 168, 169–70 (2004), *available at* https://www.cs.uic.edu/~liub/publications/kdd04-revSummary.pdf.

90. Bing Liu, *Sentiment Analysis and Subjectivity*, *in* HANDBOOK OF NATURAL LANGUAGE PROCESSING 1, 32 (N. Indurkhya & F. J. Damerau eds., 2d ed. 2010), *available at* http://www.cs.uic.edu/~liub/FBS/NLP-handbook-sentiment-analysis.pdf.

91. For detail on the methods used to construct these lists, see *id.* In short, the list was constructed by using a small set of clearly positive and negative words as seed words. Positive seed words may include "good," "amazing," and "wonderful," while the negative seed list might have "bad," "terrible," and "horrible." These seed words are then analyzed in conjunction with an online dictionary that contains synonyms and antonyms. The computational analysis of the online dictionary adds the synonyms and antonyms of the seed words to the appropriate set and then iterates again, thus building up a dictionary of positive and negative words. The final lists were inspected by the researchers to remove any clearly misclassified words. A few more negative words that made the final list include "flawed," "insulting," "moronic," and "unfounded." Positive words include "convincing," "fabulous," "heroically," and "sincerely." This is a general description of how the list was constructed and glosses over some of the machine learning elements incorporated in the methodology. Readers with an interest should refer to the Liu piece, *id.*, which provides a reasonably accessible introduction.

relatively friendly in one time period may read as downright nasty in another (or vice versa).

With these caveats in place, our analysis finds a clear time trend toward lower friendliness scores. Table 1 includes the twenty Justices with the highest and lowest friendliness scores, ordered alphabetically, with their median year of service in parentheses. For this analysis, we exclude Justices with low total production.[92] The analysis of the same data is shown graphically in Figure 3, with the median year of the Justice's term on the horizontal axis and the friendliness score on the vertical axis.

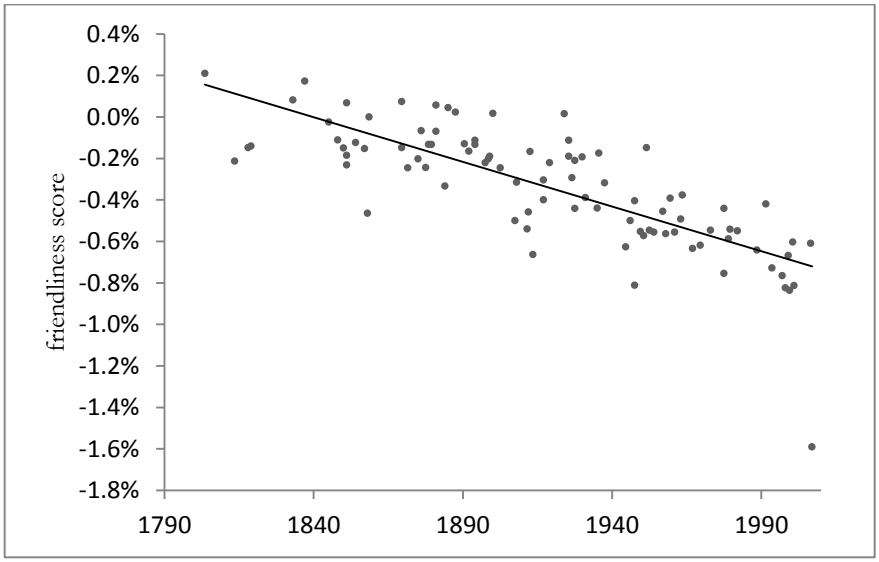
TABLE 1: TOP TEN HIGHEST AND LOWEST FRIENDLINESS SCORES

Highest	Lowest
Henry Baldwin (1837)	Samuel Alito (2007)
Samuel Blatchford (1888)	Stephen Breyer (2001)
David Josiah Brewer (1900)	Robert Jackson (1948)
Samuel Chase (1803)	Anthony Kennedy (1998)
David Davis (1870)	Joseph Rucker Lamar (1914)
Stanley Matthews (1885)	Sandra Day O'Connor (1994)
Smith Thompson (1833)	Antonin Scalia (1997)
Willis Van Devanter (1924)	David Souter (1999)
Morrison Waite (1881)	Clarence Thomas (2000)
James Moore Wayne (1851)	Byron White (1998)

92. We exclude the Justices who produced less than 100,000 words based on all of their writings (majority, concurring, and dissenting opinions). This leaves ninety-two Justices in our sample. We exclude the "low production Justices" because small total production makes it difficult to draw useful inferences; some of these Justices authored as little as a few hundred total words, leaving less than a dozen positive or negative words in their entire corpus.

Washington University Open Scholarship

FIGURE 3: FRIENDLINESS SCORE BY MEDIAN YEAR



The results are evocative: there is a highly significant negative correlation between time and friendliness scores.[93] There are a variety of potential avenues that future researchers could explore to untangle the causes of this interesting correlation. Some of this effect may be due to an increasing number of dissenting opinions, or the use of less formal language on the part of the Justices, and may be skewed by the use, or non-use, of particular negative or positive words. The changing sentiment on the Court may also reflect broader changes in language usage in political institutions (such as Congress) or within the broader culture. The time trend in friendliness scores, and its causes and potential consequences, may be worthy of future analysis.

93. An OLS regression on this data showed an R-squared of 0.61 and a p-value of less than 0.01%. Using data based on all 110 Justices adds some noise to the analysis but the results do not substantially change: the R-squared falls to 0.34 and the p-value remains below 0.01%. We also conduct the analysis on all only majority opinions, dropping the three Justices who only authored dissenting or concurring opinions, and arrive at similar results (R-squared falling to 0.27 and similarly low p-values). Dropping the lower production Justices from the majority only analysis reduces the noise considerably. For majority only opinions, dropping Justices with less than 100,000 words of production, the R-squared is 0.6.

3. *Defensiveness*

Our final preliminary analysis reexamines prior research on "defensiveness" that was conducted by Long and Christensen in 2013.[94] The basic theory underlying Long and Christensen's analysis is that people broadcast weakness through their use of language, and in particular through specific "defensive" forms of speech, including the use of intensifiers, such as the word "clearly," and more complex semantic structure.[95]

To test whether this theory describes behavior on the Supreme Court, Long and Christensen hypothesize that dissents will demonstrate these stylistic characteristics more than majority opinions. For their analysis, Long and Christensen examined 526 Supreme Court opinions in the years 2006 and 2007.[96] They counted the intensifiers in majority and dissenting opinions as a percentage of total words.[97] For their measure of complexity of semantic structure, they relied on the familiar Flesch-Kincaid reading "grade level" score.[98] Flesch-Kincaid reading grade levels are based on the average number of words per sentence and average number of syllables per word—increasing numbers of either raises the grade level.[99] Long and Christensen found a significant increase in the use of intensifiers in dissenting opinions, but found that the grade level scores were actually higher in majority opinions, although that finding was not statistically significant.[100]

We re-ran the analysis from Long and Christensen on our larger dataset to see how well their findings held up. For every Justice who filed at least one majority opinion and one dissent, an average grade level and intensifier percentage was developed for that Justice's majority opinions and dissenting opinions.[101] Tracking Long and Christensen, we found that majority opinions had somewhat higher grade levels, but that difference

---

94. *See* Long & Christensen, *supra* note 66.

95. *Id*. at 935–36.

96. *Id*. at 948 & n.70 (the majority, dissenting, and concurring opinions were drawn from 266 decisions issued between February 21, 2006, and June 28, 2007).

97. *Id*. at 949 fig.1 (illustrating "intensifier rates" per 1,000 words).

98. *Id*. at 948–49.

99. *See id*. at 943 n.40.

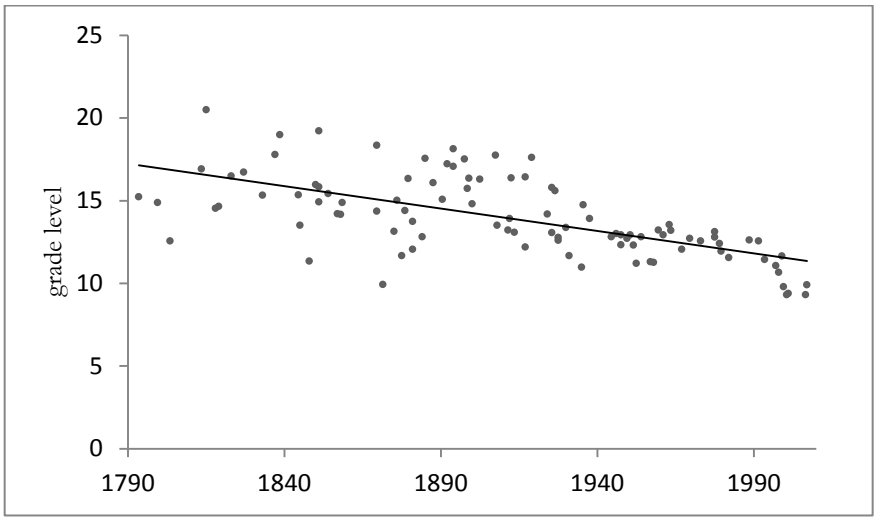
100. *Id*. at 950.

101. There were a total of ninety-nine observations. As before, Justices Blair, Iredell, and Thomas Johnson were excluded because they authored no majority opinions. There were seven Justices who did not author any dissenting opinions (Chief Justices John Jay, John Rutledge, Oliver Ellsworth, and Salmon Chase, and Justices William Cushing, Thomas Todd, and James Byrnes).

Washington University Open Scholarship

was not statistically significant.[102] More interestingly, there was a marked time trend in the sophistication of writing (as measured by grade level), with more recent Justices writing at lower grade level.[103] The time trend analysis is presented in Figure 4.

FIGURE 4: GRADE LEVEL BY MEDIAN YEAR



From this analysis, it appears that the Court has generally reduced the complexity of its language (as measured by Flesch-Kincaid Grade Level) over time.[104] This finding runs contrary to the findings of Johnson's examination of grade level trends using a smaller sample: cases written during the 1931–1933 and 2009–2011 terms.[105] Johnson found that writing complexity, and Flesch-Kincaid Grade Level specifically, had increased over time. A glance at Figure 4 reveals that there is variability around the general time trend toward lower grade level scores, making any inference

102. A Student's t-test was used to determine whether there was a statistically significant difference in means for either grade level or intensifier use. *See* CHARLES HENRY BRASE & CORRINNE PELLILLO BRASE, UNDERSTANDABLE STATISTICS 479 (11th ed. 2014).

103. To develop a single grade level for each Justice's writings, we averaged the grade level for their dissents and majority opinions.

104. An OLS regression returned an R-squared value of 0.4 and a p-value of less than 0.01%. The coefficient was -0.03.

105. *See* Johnson, *supra* note 67, at 57–58 (finding that the Court's 2009–2011 opinions were written at about a grade level higher on the Flesch-Kincaid Grade Level scale than the Court's 1931–1933 opinions).

of a general time trend from limited data difficult. Of course, the actual grade level comparison between the two temporal sets made by Johnson remains valid, even if it does not appear to be representative of a longer and more general time trend.

It should be noted that Flesch-Kincaid scores have been criticized as a measure of sophistication and complexity.[106] A general time trend toward lower grade level does not necessarily mean that the Court's reasoning is less sophisticated, or that its writing is of lower quality. Good writing does not necessarily involve long words or long sentences.[107] An interesting question for future research would be whether the trend toward lower complexity in the Court's writings is mirrored in broader social trends, or marks a trend toward more vernacular writing that is more closely in line with non-judicial writing styles.

We also also confirmed Long and Christensen's findings on intensifier use.[108] There was a markedly higher use of intensifiers in dissents, with means of 0.12% of words for majority opinions and 0.18% for dissents, a statistically significant difference.[109] Unlike friendliness and grade level, there was no obvious time trend in intensifier use—it appears that intensifiers have been used at roughly similar rates across the data.

C. The Stylistic Fingerprint and Similarity Scores

Our stylistic analysis moves beyond attempts to measure specific stylistic features of writing, and instead relies on a measure that is meant to serve as a broad proxy for a range of stylistic characteristics: the use of function words.

Function words play a special role in language and remain stable over time.[110] Content words, on the other hand, are constantly added. For example, the 2014 update to Merriam-Webster's Collegiate Dictionary includes "hashtag," "selfie," and "crowdfunding"—all very much content words.[111] Function words can often be very short, such as "I, the, a, of,"

---

106. *Id*. at 50–51.

107. *See generally* George Orwell, *Politics and the English Language*, 13 HORIZON 252 (1946) (criticizing use of overly stylized and dehumanizing language in political discourse).

108. *See* Long & Christensen, *supra* note 66, at 950.

109. The p-value was less than .01%.

110. ADRIAN AKMAJIAN, RICHARD A. DEMERS, ANN K. FARMER & ROBERT M. HARNISH, LINGUISTICS: AN INTRODUCTION TO LANGUAGE AND COMMUNICATION 21 (4th ed. 1995) (discussing function words).

111. Katy Steinmetz, *# Selfie, Steampunk, Catfish: See This Year's New Dictionary Words*, TIME (May 19, 2014), http://time.com/103503/merriam-webster-dictionary-selfie-catfish/, *archived at* https://perma.cc/QCC2-6DYF.

Washington University Open Scholarship

while content words are rarely as short. There also appear to be neurological differences between function and content words. Function words are acquired by children later than content words, and specific types of neurological injuries can lead to a loss of use of function words, while content words remain accessible.[112] Various neurological studies have found that function and content words are stored and processed in different brain regions.[113] For purposes of the following analysis, the most important characteristic of function words is that they have been found to provide a source for the development of a useful stylistic "fingerprint" that can be used for author attribution, and we therefore use it as a proxy for writing style more generally.[114]

Our study relies on 307 standard function words (or "content-free words" or "CFWs") listed in Table 2. The individual occurrences of each CFW for each author are aggregated and normalized so that the components sum to one.[115] These normalized vectors are the *feature vectors* for each author.[116] In their normalized form, they also represent probability distributions (their components are nonnegative and they sum to one). To construct feature vectors, the relevant corpus was identified, such as all writings associated with a Justice, the writings of a Justice in a given year, or all of the writings of all of the Justices in a given year. A Python script was used to count each of the CFWs and output the feature vectors into a simple text format.

---

112. *See* Lise Menn & Loraine K. Obler, *Cross-Language Data and Theories of Agrammatism*, *in* 3 AGRAMMATIC APHASIA: A CROSS-LANGUAGE NARRATIVE SOURCEBOOK 1369, 1370, 1388 n.5 (Lise Menn & Loraine K. Obler eds., 1990).

113. *See, e.g.*, Jonathan W. King & Marta Kutas, *A Brain Potential Whose Latency Indexes the Length and Frequency of Words*, CRL NEWSLETTER (Ctr. for Res. in Language, UCSD, San Diego, CA), Nov. 1995, at 1, 1.

114. *See* HFKR (2012), *supra* note 7, at 7682. In their study of clerk influence on the Supreme Court, Rosenthal and Yoon use sixty-three function words. Jeffrey S. Rosenthal & Albert H. Yoon, *Detecting Multiple Authorship of United States Supreme Court Legal Decisions Using Function Words*, 5 ANNALS APPLIED STAT. 283, 287 (2011).

115. We normalized the values by dividing by the L^1-norm, which is the sum of the absolute values of the components of the vector. HFKR (2012), *supra* note 7, at 7683.

116. Wikipedia provides the following useful definition of a feature vector: "In pattern recognition and machine learning, a feature vector is an n-dimensional vector of numerical features that represent some object." *Feature Vector*, WIKIPEDIA, https://en.wikipedia.org/wiki/Feature_vector (last visited Mar. 31, 2016), *archived at* https://perma.cc/V5R6-JK89. In essence, our stylistic feature vectors are a string of 307 numbers that are the relative frequency of the 307 content free words within the relevant group of writings. As relative frequencies, these vectors also represent probability distributions.

TABLE 2: CONTENT-FREE WORDS

a about above across after afterwards again against all almost alone along already also although always am among amongst amoungst amount an and another any anyhow anyone anything anyway anywhere are around as at back be became because become becomes becoming been before beforehand behind being below beside besides between beyond both bottom but by call can cannot cant con could couldnt cry describe detail do done down due during each eight either eleven else elsewhere empty enough etc even ever every everyone everything everywhere except few fifteen fifty fill find fire first five for former formerly forty found four from front full further get give go had has hasnt have he hence her here hereafter hereby herein hereupon hers herself him himself his how however hundred ie if in inc indeed into is it its itself keep last latter latterly least less ltd made many may me meanwhile might mine more moreover most mostly move much must my myself name namely neither never nevertheless next nine no nobody none noone nor not nothing now nowhere of off often on once one only onto or other others otherwise our ours ourselves out over own part per perhaps please put rather re same see seem seemed seeming seems serious several she should show side since six sixty so some somehow someone something sometime sometimes somewhere still such take ten than that the their them themselves then thence there thereafter thereby therefore therein thereupon these they thin third this those though three through throughout thru thus to together too top toward towards twelve twenty two under until up upon us very via was we well were what whatever when whence whenever where whereafter whereas whereby wherein whereupon wherever whether which while whither who whoever whole whom whose why will with within without would yet you your yours yourself yourselves

The degree of difference between two probability distributions can be measured as the Kullback-Leibler ("KL") divergence. KL divergence is a standard measure for comparing distributions, and it has been used in prior studies of the evolution of writing style.[117] Following convention, we use a symmetrized version of KL divergence that is the average of the KL divergences of A with respect to B and B with respect to A.[118] The KL divergence is then scaled to generate a similarity score between zero and one.

It is important to reiterate that our stylistic fingerprint is not meant to capture the totality of judicial writing style. It would be strange indeed to claim that the frequency with which Justice Holmes used the word "it" accounts for the claim by Judge Posner—nearly a century later—that the dissenting opinion in *Lochner* is a "rhetorical masterpiece."[119] Instead, the

117. HFKR (2012), *supra* note 7, at 7683. For an introduction to KL divergence in the context of natural language processing, see CHRISTOPHER D. MANNING & HINRICH SCHUTZE, FOUNDATIONS OF STATISTICAL NATURAL LANGUAGE PROCESSING 39–80 (1999). Note that to avoid undefined division by zero, we add .0001 to all of the components in all of the feature vectors, regardless of whether a word was used. This technique is called "smoothing" and will always result in an increase in our similarity score when compared to the unsmoothed version.

118. Hereinafter, when we refer to using the KL divergence we implicitly mean in this symmetrized form.

119. Richard A. Posner, *Law and Literature: A Relation Reargued*, 72 VA. L. REV. 1351, 1383 (1986).

Washington University Open Scholarship

feature vector is meant to serve as a proxy for the larger set of stylistic characteristics that distinguish one writer from another.

There are other potential measures of style. For example, the LitStats software mentioned above reports statistics on eight specific stylistic factors: average footnote length, average sentence length, average word length, word length diversity, sentence length diversity, footnote frequency, type-token ratio,[120] and the once-word rate.[121] These factors have been used in analysis of juridical writings.[122] Alternatively, scholars have looked to compression software to generate a measure for similarity between writings.[123] All of these methods are plausible, and there is no consensus on a dominant quantitative methodology for quantitative measurement of style.[124] The stylistic measure used in this study has the advantage of simplicity, and it is commonly used in both forensic and literary attribution work.[125]

III. TIME TRENDS IN JUDICIAL STYLE

This Part applies the methodology just described to examine how writing style on the Court changes over time. Specifically, we ask whether there is a "style of the time," in the sense that contemporaneous Justices tend to write more similarly than Justices who are temporally remote from one another. As will be clear from the analysis below, the answer to that question is "yes."

To undertake our analysis of the relationship between temporal distance and writing style similarity, we first calculated feature vectors for all Justices and created similarity scores for every Justice-pair within the study period. Each Justice was also assigned a place in time, based on the mid-point of their term on the Court.[126]

---

120. The type-token ratio is "the number of different words in an opinion (types) as a percentage of the total number of words in the opinion (tokens)." Paul J. Wahlbeck, James F. Spriggs II & Lee Sigelman, *Ghostwriters on the Court? A Stylistic Analysis of U.S. Supreme Court Opinion Drafts*, 30 AM. POL. RES. 166, 176 (2002).

121. The once-word rate is "the relative frequency of words that appear exactly once in an opinion." *Id*. (listing and explaining the eight stylistic factors).

122. *See id*.

123. Stephen J. Choi & G. Mitu Gulati, *Which Judges Write Their Opinions (And Should We Care?)*, 32 FLA. ST. U. L. REV. 1077, 1105 (2005). The greater the rate of compression, the greater the assumed stylistic consistency. *Id*.

124. *See id*. at 1105 nn.82–83 (noting different statistical approaches to attributing authorship and citing relevant scholarship).

125. *See* HFKR (2012), *supra* note 7, at 7682.

126. For Justices serving at the end of the study period, we used the last year as the end of their term.

Our first analysis is a representation of similarity scores as a style "network" with Justices "linked" to each other based on stylistic similarity. In the terminology of network analysis, the Justices are "nodes" and a thresholding technique on the stylistic similarity is used to determine when "edges" (or links) are placed between the nodes.[127] Each Justice is a node in the network, and an edge was created between that Justice and the 5% of other Justices with the highest similarity scores in their set.[128] We then undertake a quantitative estimate of groups within the style network, using the methodology of spectral clustering analysis.[129] Boyd, Hoffman, Obradovic, and Ristovski describe a use of the spectral clustering methodology, which is a technique used to "classify and group" items within a dataset.[130] In essence, spectral clustering "cuts" a network into some defined number of groups (i.e., accomplishes a "clustering"), relative to the condition that similarity between members of the groups should be relatively high and the similarity between members of different groups relatively low.[131] A related (and often thorny) problem in spectral clustering is the determination of the number of clusters as based on the data.[132] We did not address that second problem—which is not necessary to our analysis—and instead set the number of clusters to be identified at seventeen, which is the number of Chief Justices that have served on the Court. That number is admittedly somewhat arbitrary, but it is sufficient for our purposes, which is generally to examine whether Justices' writing styles appear to cluster together on a temporal basis. The groups generated by the spectral clustering analysis are ordered by the median year of the

127. Nicholas J. Foti, James M. Hughes & Daniel N. Rockmore, *Nonparametric Sparsification of Complex Multiscale Networks*, 6 PLOS ONE 1, 1 (2011), *available at* http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0016431.

128. If the edge already exists (because it was added when a previous Justice was considered), it was not added again, but a new edge was not moved into the top 5% to replace it. Three Justices are excluded from the graph based on a lack of similarity to any other Justice: Chief Justice John Rutledge, Justice Moore, and Justice Thomas Johnson.

129. *See generally* Ulrike Von Luxburg, *A Tutorial on Spectral Clustering*, 17 STAT. & COMPUTING 395 (2007), *available at* http://link.springer.com/article/10.1007%2Fs11222-007-9033-z.

130. *See* Christina L. Boyd, David A. Hoffman, Zoran Obradovic & Kosta Ristovski, *Building a Taxonomy of Litigation: Clusters of Causes of Action in Federal Complaints*, 10 J. EMPIRICAL LEGAL STUD. 253, 262–64 (2013). See Frank B. Cross & Stefanie Lindquist, *Judging the Judges*, 58 DUKE L.J. 1383 (2009), and Corey Rayburn Yung, *A Typology of Judging Styles*, 107 NW. U. L. REV. 1757 (2013), as other examples of cluster analysis in empirical legal studies.

131. Boyd, Hoffman, Obradovic & Ristovski, *supra* note 130, at 263. In Appendix B, Boyd et al. provide an example of a general spectral clustering algorithm, similar to the program used in our analysis. *Id.* at 285–87.

132. *See generally* Asa Ben-Hur, Andre Elisseeff & Isabelle Guyon, *A Stability Based Method for Discovering Structure in Clustered Data*, 17 PAC. SYMP. ON BIOCOMPUTING 6 (2002) (describing a method for determing the "optimal number of clusters").

Washington University Open Scholarship

Justices in the cluster, and the range of median years is presented alongside the group as well. Table 3 presents the results from the spectral clustering analysis.[133]

TABLE 3: SPECTRAL CLUSTERING ANALYSIS

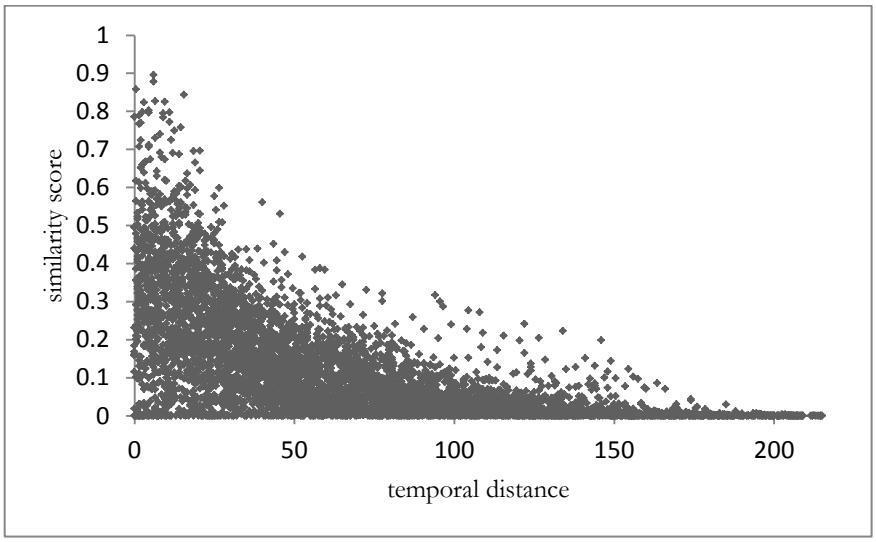
ID	Yr/Rg							
1	1812	Blair	Jay	Johnson_T	Wilson	Iredell	Rutledge_J	
	78	1792	1792	1793	17945	1795	1795	
		Ellsworth	Cushing	Paterson	Moore	Todd	Duvall	
		1798	1800	1800	1802	1816.5	1823	
		Woodbury	Chase_Salmon	Clifford				
		1848	1869	1870				
2	1837	Washington	Johnson_W	Trimble	Thompson	Barbour	Daniel	Story
	65	1813.5	1819	1827	1833	1838.5	1851	1879
3	1842	Livingston	Marshall_J	Baldwin	Curtis	Grier	Davis	
	55	1815	1818	1837	1854	1858	1870	
4	1848	Mclean	Catron					
	6	1845	1851					
5	1853	Mckinley	Taney	Campbell	Nelson			
	14	1844.5	1850	1857	1858.5			
6	1875	Wayne	Miller	Bradley	Waite	Woods		
	33	1851	1876	1881	1881	1884		
7	1890	Field	Matthews	Lamar_L	Gray	Harlan_I	Jackson_H	Shiras
	18	1880	1885	1891	1892	1894	1894	1898
8	1891	Swayne	Strong	Hunt	Brown	Brewer	Peckham	Mckenna
	40	1872	1875	1878	1899	1900	1903	1912
9	1892	Chase_Samuel	Lamar_J	Holmes	Cardozo			
	132	1804	1914	1917	1935			
10	1893	Blatchford	Fuller					
	12	1888	1899					
11	1919	White_E	Moody	Lurton	Day	Pitney	Mcreynolds	
	30	1908	1908	1912	1913	1917	1928	
		Sutherland	Roberts_O					
		1930	1938					
12	1926	Clarke	Vandevanter	Hughes	Taft	Sanford	Brandeis	Butler
	12	1919	1924	1925.5	1926	1927	1928	1931
13	1948	Stone	Byrnes	Murphy	Reed	Vinson	Burton	
	24	1936	1942	1945	1947.5	1950	1952	
		Minton	Whittaker					
		1953	1960					
14	1954	Rutledge_W	Jackson_R	Frankfurter	Black	Douglas	Fortas	
	21	1946	1948	1951	1954	1957	1967	
15	1961	Clark	Warren	Harlan_Ii	Goldberg			
	6	1958	1961	1963	1964			
16	1980	Stewart	Brennan	Burger	White_B	Marshall_T	Powell	
	22	1970	1973	1978	1978	1979	1980	
		Blackmun	Rehnquist	Stevens				
		1982	1989	1992				
17	2000	O'Connor	Scalia	Kennedy	Souter	Thomas	Ginsburg	
	14	1994	1997	1998	1999	2000	2001	
		Breyer	Roberts_J	Alito				
		2001	2007	2007				

133. A representation of the network is available as a supplemental online figure at https://math.dartmouth.edu/images/screenshot_103919.png.

The most striking observation from these analyses is the degree to which Justices are more stylistically similar to their contemporaries than to temporally distant Justices. This is especially the case in the modern era, with the Justices on the current Court quite isolated stylistically from Justices in earlier years, which can be seen quite clearly in the supplemental online figure. In general, the spectral clustering analysis displayed in Table 3 created groups that were time-based, with temporal ranges of a few decades, and some closer to a single decade.[134]

To analyze more closely the relationship between time and stylistic similarity, we characterized every Justice by the median year of his or her term of service on the Court. For sitting Justices (including Justice Scalia), 2008 was used as the end of their tenure. We then calculated the distance in time for every pair of Justices, and related those distances to the similarity score for those Justices. The results are presented in Figure 5.

FIGURE 5: SIMILARITY SCORES BETWEEN JUSTICES, AS A FUNCTION OF TIME



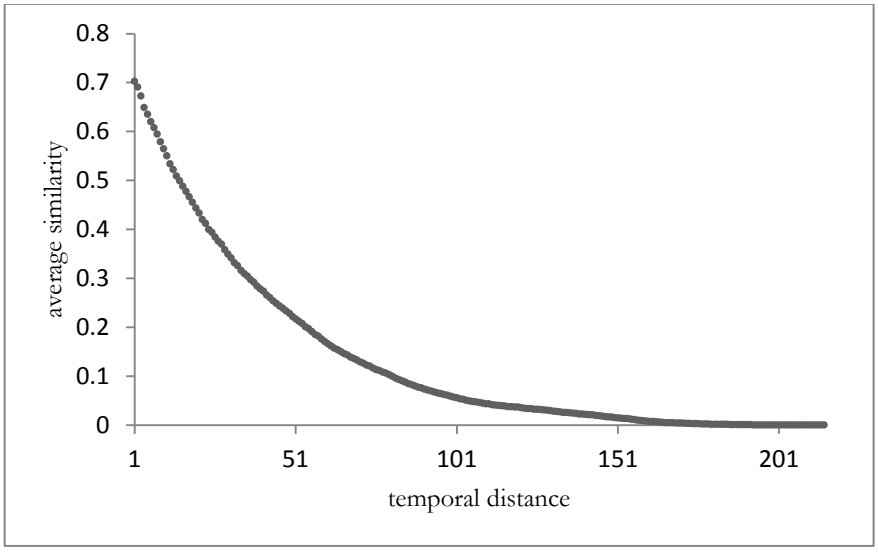
134. The outlier groups are 9 and 1, which have somewhat larger ranges.

Washington University Open Scholarship

An OLS regression generated an R-squared of 0.18 and a p-value of less than 0.01%, and a coefficient for temporal distance of 0.047.[135] These findings can be interpreted as indicating that, while there are sources of variation in the data other than time, there is also a strong trend of declining similarity in time, with a rate of decay in similarity score of roughly 4–5%. As Justices move farther apart in time, they become increasingly distinct in their writing styles.

To examine the influence of time from a somewhat different angle, we next calculated feature vectors for all years and created similarity scores for every year pair within our study period. We then calculated average similarity scores based on temporal distance: one-year distant pairs were averaged into a single similarity score; two-year distant pairs were averaged into a second; and so on. The average similarity score for temporally matched pairs is represented in Figure 6.

FIGURE 6: AVERAGE SIMILARITY AND TEMPORAL DISTANCE



135. Similarity scores were normalized with a natural log transformation.

Overall, these results indicate a decline in the similarity of year feature vectors as they move farther apart in time, with a rate of decay in similarity of around 4–5%.[136] This analysis again provides strong evidence that style on the Court is not time independent, but instead changes over time. The writings of Justices working together in a given decade are far more stylistically similar to each other than they are to writings of Justices on a temporally remote Court.

IV. POTENTIAL MECHANISMS

The foregoing analysis raises an interesting question as to why stylistic similarity within judicial writing declines with temporal distance between Justices. There are a variety of potential mechanisms that could cause writing style to change with time. When examining the Court, it is perhaps most natural to look to external factors, including broader societal trends in writing style. Questions surrounding change of writing style on the Court, then, would necessarily implicate a larger set of questions concerning change of writing style outside the Court in various media, including literature, popular culture, and personal communication.[137] Those questions, while no doubt of interest, are outside the scope of this project.

We restrict our analysis to internal factors that could help explain a change in style. In this Part, we examine three potential causal mechanisms. The first is influence by highly respected prior decisions. We do not find convincing evidence that more frequently cited prior decisions exert any particularly great influence on later style. We then examine changes in the Court's composition and do not find evidence that the partisan affiliation of Justices has an effect on style. Finally, we examine the potential influence of substance by comparing dissenting and majority opinions and find some evidence that writing style bears some relationship to opinion type.

136. The coefficient for temporal distance is -0.04 in an OLS regression of the natural log; the p-value is less than .01%. The R-squared for this analysis is 0.92, reflecting the reduction in noise due to the averaging procedure.

137. *Cf.* HFKR (2012), *supra* note 7, at 7684–85 (examining writing style in literary texts). It is worth noting that a "style of the time" appears to exist across textual domains; both in literature, where novelty and innovation are prized, as well as in judicial writing, which is, at least arguably, more formal and formulaic.

Washington University Open Scholarship

A. Prior Decisions

The first mechanism that we examine is the possibility of a causal role played by influential past decisions. Just as past decisions generate legal standards and norms of judicial reasoning, they serve as the backdrop against which a Justice's writing style is perceived. While some innovation in writing style may be rewarded, Justices are likely to express some degree of conformity to prevailing conventions. Justices may also consciously model their writing style on prior Justices who they find to be particularly worthy of emulation, or may be subconsciously influenced by the decisions that they read.

To test for the influence of prior Justices, we rely on the current Court as our baseline. To create the baseline, we used the writings of each of the currently sitting Justices in our dataset (Alito, Breyer, Ginsburg, Kennedy, Roberts, Scalia, and Thomas) to generate a single stylistic feature vector.[138] We then excluded from our analysis all of the Justices who served at the same time with any sitting Justice as a way to separate out any cross-influence between Justices and ensure that the causal relationship runs in the anticipated direction. For each remaining Justice, we constructed a feature vector for their writing, and calculated the KL divergence between that feature vector and the current Court baseline.

For each Justice in the analysis, we then constructed a "ghost" vector made up of the texts produced by the Court in each of their years on the bench, excluding that Justice's writings.[139] We calculated the KL divergence between the ghost vectors and the current Court baseline vector. Finally, we subtracted the KL divergence for each Justice's ghost vector from their KL divergence to generate what we call "prediction scores": a difference less than zero indicated that a Justice's writing was more similar to the current Court's style than to the other writings of the Court in the years when that Justice was on the bench. Justices who perform well tend to "predict" the current style of the Court better than Justices who perform poorly (with lower numbers associated with better prediction). There were few Justices with prediction scores of less than zero, because each Justice typically authors only a small fraction of cases

138. At the time of our analysis, Justice Scalia was an active member of the Court.

139. For this analysis, we use a smaller list of seventy-five non-content words:

"first between also where who those part than him will could without whether must after before within should these only them when against same so one would their there has they other all made may if we us he under but been had his were no have are any its upon such at an with from on which this not or as for be it was by is a that in and to of the"

in a given year, meaning that random sources of variation are much less likely to substantially influence the ghost vectors than an individual Justice's vector.

We then compared the resulting prediction scores to a measure of "historical value" for each Justice, which was generated by Kosma in 1998 based on citation counts.[140] This variable is meant to capture the possibility that Justices who are widely cited exert greater stylistic pressure on subsequent Justices. We controlled for the relationship between a Justice's total production, in words, and our prediction scores. There are two potential mechanisms for this variable to affect the prediction scores. First, Justices that produce a great deal of text contribute more to the total body of the law that later Justices read. For that reason, perhaps they exert greater stylistic influence. Higher levels of production also imply less opportunity for random sources of variability in the use of function words to affect a Justice's feature vector. Finally, we controlled for time, to account for temporal effects that are not captured in the ghost vector-based normalization, and examine the interaction between production and Kosma's historical value.[141]

---

140. *See* Kosma, *supra* note 10, at 352 tbl.2. In Kosma's analysis, Chief Justice Hughes is given two different scores, corresponding to the two different stints that he spent on the Court. *Id.* at 350 n.39. Because of the lack of correspondence to our single entry for Hughes, we dropped him from both sides of the analysis.

141. For this analysis, we use the data described above to construct three variables. *Hist* is based on Kosma's historical value scores, normalized through a cube root function, centering at zero and scaled by the standard deviation. *Prod* is based on total word production, again normalized through a cube root function centering at zero and scaled by the standard deviation. *Predict* is the natural log of the prediction scores. We created a fourth variable, *Predict1*, which is a cube root transformation of the prediction scores: the log transformation better normalizes the data but creates difficulties around the negative prediction scores. For *Predict* we dropped the negative observations, which are retained in *Predict1*.

Washington University Open Scholarship

TABLE 4: HISTORICAL INFLUENCE

	Predict	*Predict*	*Predict*	*Predict1*
Hist	-0.488	-0.098	-0.150	-0.033
	(6.94)**	(0.53)	(0.83)	(1.22)
Prod		-0.414	-0.293	-0.020
		(2.27)*	(1.48)	(0.66)
Year			0.000	0.000
			(0.10)	(0.78)
*Hist*Prod*			0.242	0.036
			(3.55)**	(3.44)**
_cons	-4.074	-4.071	-4.609	-0.171
	(58.46)**	(59.82)**	(1.41)	(0.34)
Adj.-R^2	0.36	0.40	0.49	0.34
N	87	87	87	91

* $p<0.05$; ** $p<0.01$, *t* statistic in parentheses

Historical value is significant in the first specification as a standalone variable. It drops out of significance in the second model, which includes production. In the full model, the interaction term is significant.[142] Recall that lower prediction scores imply greater similarity. To interpret the interaction, we examined the effects of historical value at different levels of production, finding that historical value is significant at production levels less than one standard deviation below the mean, where it improves (i.e., lowers) prediction scores.[143]

What to make of this analysis? Justices at the lowest level of productivity have poor prediction scores, and their prediction scores improve as they become more productive. Our analysis cannot determine whether that effect is from a reduction in statistical noise or a greater likelihood that a future Justice read and internalized their writing style. Among the lower productivity Justices, authoring more highly cited opinions appears to improve their prediction scores. For Justices at higher levels of productivity, we do not find that additional citation contributes to greater stylistic similarity with future Justices.

142. Note that this is true for both the logarithmic and the cube root transformations of prediction scores. The lack of significance for *Hist* and *Prod* in the full model indicates that, at the mean value of *Prod*, *Hist* is not significant, and vice versa.

143. This is the case for both the logarithmic and the cube root transformations of prediction scores. *See generally* A. COLIN CAMERON & PRAVIN K. TRIVEDI, MICROECONOMETRICS USING STATA 343–57 (2010) (discussing calculation of marginal effects).

B. Partisan Affiliation

We next examine the possibility that writing style is associated with some other set of cognitive, ideological, value-based, or perceptive characteristics, and that the change in writing style over time on the Court reflects a broader shift in Weltanschauung.[144] This type of relationship is hard to test, for obvious reasons, but we conduct a very general analysis by examining whether there is any systematic stylistic difference between Justices appointed by Presidents of different parties.

We only test differences between the contemporary Democratic and Republican parties, and so restrict our analysis to the latter half of the twentieth century.[145] Looking over the entire study period, there is somewhat less similarity between Democratic-appointed Justices and Republican-appointed Justices than within the party groupings. However, there is an obvious temporal problem, because the relative representation of the two parties on the Court has shifted markedly over time.

To account for this feature in the data, we compare the inter- and intra-party difference for each year, starting in 1955. We generate a feature vector for the texts authored by the Justices appointed by Republican and Democratic Presidents and calculate a similarity score between them. This analysis is done for each year starting in 1955. For each party, we then subdivide the opinions, randomly, into two test groups and generate feature vectors for the test groups. Finally, we calculate similarity scores for the feature vectors for the same-party test groups, and then average the two scores (Democratic and Republican) to generate a measure of intra-party distance. The hypothesis is that the inter-party similarity scores will be lower than the intra-party scores.

On average, the similarity scores were a shade *higher* for the inter-party group (contrary to the hypothesis). We also conducted a very simple additional test by computing similarity scores for "parties" generated by

144. On the relationship of writing style with other cognitive or ideological factors on the Court, see Deborah H. Gruenfeld, *Status, Ideology, and Integrative Complexity on the U.S. Supreme Court: Rethinking the Politics of Political Decision Making*, 68 J. PERSONALITY & SOC. PSYCHOL. 5 (1995); Tetlock, Bernzweig & Gallant, *supra* note 48.

145. We use the dawn of the FDR coalition as the point at which the contemporary meaning of "Democrat" and "Republican" take shape. Although there is some controversy over the meaning of "realignment" elections and their relationship to party systems. *See generally* DAVID R. MAYHEW, ELECTORAL REALIGNMENTS (2002) (arguing that realignment elections are overemphasized by political scientists). There is enough data to generate inter- and intra-party similarities for post-1932 appointees starting in the mid-1950s.

Washington University Open Scholarship

random assignment of opinions for each year.[146] The actual similarity scores were somewhat lower—both inter- and intra-party—compared to randomly generated groupings. There is no clear interpretation for this feature of the data, but it does not provide particular evidence that partisan affiliation is associated with stylistic difference.

To account for the possibility that writing style has become more polarized over the course of our dataset (as the parties have polarized) we examined whether there was any time trend toward increasing dissimilarity between inter- and intra-party similarity scores. We found a very mild time trend toward greater dissimilarity, but time accounted for very little of the variation and the trend was not statistically significant.[147] Overall, for this relatively small portion of the dataset (53 years), we did not find evidence that differences between the parties account for changes in judicial writing style over time.

C. Substantive Factors

We also examined the degree of difference between opinion types (dissenting and majority opinions) compared to the degree of divergence within opinion type. For this analysis, we eliminated pre-1950 texts, when dissenting opinions were relatively rare. We then randomly separated majority opinions into two groups and dissenting opinions into two groups, and calculated the KL divergence between the feature vectors constructed between those two groups. As expected, given the large number of texts in each group, the KL divergence was quite small.[148] We then examined the differences between majority and dissenting opinions and found that the KL divergence for these groupings was two orders of magnitude higher.[149] This is a statistically significant result.[150]

To account for the possibility that the growing number of dissents combined with general stylistic trends caused these differences, we constructed corpora of dissents and majority opinions for each Justice, and conducted the same within-group and between-group analyses. Because their groups were smaller, there was greater opportunity for random

146. *See* MAX KUHN & KJELL JOHNSON, APPLIED PREDICTIVE MODELING 69–73 (2013) (discussing techniques used to test model efficacy).

147. The R-squared value was 0.02, and the p-value was 38%.

148. The KL divergence was 0.00026 for dissenting opinions, and 0.00011 for majority opinions.

149. Majority1-Dissent1, 0.011; Majority1-Dissent2, 0.010; Majority2-Dissent1, 0.011; Majority2-Dissent2, 0.010.

150. An analysis on simulated groupings showed that the likelihood of randomly generating such a difference between similarly sized groups was well below 0.01%.

variation to affect the feature vectors, and the KL divergences were greater in general.[151] We also examined the KL divergence between majority and dissenting opinions, finding that there was not the same order of magnitude difference, but that there were statistically significant differences between the mean KL divergences.[152]

The bottom line of this analysis is that there does appear to be a difference in writing style between dissents and majority opinions, even for the same Justice. One potential source of temporal variation in writing style, then, may be the growing prevalence of dissents on the Court.[153] Given that this particular form of judicial writing appears to be stylistically distinct, its growth in popularity may account for some of the temporal drift in writing style on the Court.

V. CLERK INFLUENCE

As judicial clerks have become an enduring feature of the operation of the federal courts, the role of these recent law graduates has been the subject of both scholarly and public debate.[154] An important empirical predicate to this debate is the belief that clerks play a substantial role in authoring opinions.[155] At least for the Supreme Court, there is a long

151. The average KL divergence between majority opinions was 0.011; the average KL divergence between dissenting opinions was 0.05.

152. The average KL divergence between opinion types, for all groups (majority1-dissent1; majority2-dissent1; majority1-dissent2; majority2-dissent2), for all Justices was .04. We conducted a t-test on the difference in means between the KL divergence within majority opinions and between opinion types, and the difference in means between the KL divergence within dissenting opinions and between opinions types. Both were significant ($p<0.01\%$). It is interesting to note that, for dissents, the average KL divergence within type was *greater* than the average KL divergence between majority and dissenting opinions. This seems to indicate that there is quite substantial stylistic variation within dissenting opinions of the same Justice.

153. Our tests of concurrences found that they had half the KL divergence from majority decisions as dissents; also, it is possible the majority writings have begun to take on the tone of earlier dissents, such as being more argumentative.

154. *See, e.g.*, TODD C. PEPPERS, COURTIERS OF THE MARBLE PALACE: THE RISE AND INFLUENCE OF THE SUPREME COURT LAW CLERK (2006) (discussing role of clerks on the Court); ARTEMUS WARD & DAVID L. WEIDEN, SORCERERS' APPRENTICES: 100 YEARS OF LAW CLERKS AT THE UNITED STATES SUPREME COURT (2006) (same). The debate is longstanding. *See, e.g.*, Chester Newland, *Personal Assistants to Supreme Court Justices: The Law Clerks*, 40 OR. L. REV. 299 (1961) (early article discussing role of law clerks). For discussion of the effects of clerks on judicial writing style, see ANTHONY T. KRONMAN, THE LOST LAWYER: FAILING IDEALS OF THE LEGAL PROFESSION 347–51 (1993) (criticizing "the rise, to a position of previously unimaginable influence, of law clerks"); RICHARD A. POSNER, THE FEDERAL COURTS: CHALLENGE AND REFORM 146 (1996) (speculating that the reliance of judges on clerks may lead to longer opinions).

155. The effect of clerks on the selection of cases is also important. Research on this element of clerk influence has typically assumed that clerks act as more or less faithful agents for their hiring

Washington University Open Scholarship

history of anecdotal evidence supporting the claim that law clerks exert some influence over judicial decision-making.[156] There is also a nascent literature that uses quantitative techniques to address the question of clerk influence over both substance and style. This Part investigates whether the stylistic measures discussed above can provide insights into whether clerks have had a measurable effect on writing style on the Court.[157]

A. Previous Studies

The question of clerk influence over opinion drafting has been the subject of several attempts at computational content analysis. Before moving to our analysis, we briefly review that literature and its findings.

A paper by Wahlbeck, Spriggs, and Sigelman in 2002 is the first attempt (to our knowledge) to use computational text analysis to address the question of clerk authorship.[158] That analysis relied on seventy-one opinions during the 1985 term that were authored by Justices Powell and Marshall.[159] The authors identify the clerks most likely to have assisted drafting each opinion, based on Justice Powell's records (for Powell's clerks) and the identity of the clerk who authored the bench memo on the case (for Marshall's clerks).[160] There were nine clerks in the study period.[161] The authors analyzed the seventy-one opinions using LitStats 1.62, a software package that conducts basic analysis of digital textual information.[162] The authors analyzed eight stylistic features of the texts, including average word length, average sentence length, footnote

---

Justice. *See* Ryan C. Black & Christina L. Boyd, *The Role of Law Clerks in the U.S. Supreme Court's Agenda-Setting Process*, 40 AM. POL. RES. 147, 150–52 (2012).

156. For an early popular exchange on the subject, *compare* William H. Rehnquist, *Who Writes Decisions of the Supreme Court?*, U.S. NEWS & WORLD REP., Dec. 13, 1957, at 74–75 (raising concern about the role of clerks, especially with respect to their role in certiorari decisions), *with* William D. Rogers, *Clerks' Work is "Not Decisive of Ultimate Result,"* U.S. NEWS & WORLD REP., Feb. 21, 1958, at 114–16 (disputing claim of strong clerk influence), *and* William H. Rehnquist, *Another View: Clerks Might "Influence" Some Actions*, U.S. NEWS & WORLD REP., Feb. 21, 1958, at 116 (reasserting potential influence of clerks).

157. Of course, even if clerks authored opinions, they are not necessarily able to influence substantive outcomes. Nevertheless, the ability to influence style may be indicative of substantive influence, and the question of stylistic influence is interesting in its own right.

158. *See* Wahlbeck, Spriggs & Sigelman, *supra* note 120.

159. *See id.* at 175. In analysis of clerkship influence over judicial opinions, the question of "authorship" is, to some extent, the object of study. By convention, we refer to the Justice under whose name an opinion appears as the "author" throughout this Article.

160. *Id*. at 174.

161. *Id*. at 175.

162. *Id*. at 176.

frequency, and footnote length.[163] The authors then calculated a simple measure of similarity for the stylistic features between two texts, and estimated the predicted similarity for each of the nine clerks.[164] The authors concluded that "Powell's clerks displayed less autonomy than Marshall's," a finding that they found was consistent with anecdotal accounts of the two Justices' work habits.[165]

Choi and Gulati, writing in 2005, rely on the GZip compression software, which "compresses documents based on the similarities in the basic linguistic building blocks of . . . two files."[166] Greater compression for a corpus implies that it comprises more similar texts.[167] The authors examine the writings of ninety-eight judges in the federal appellate courts that were active in their study period of 1998–2000, drawing four text samples of 8,000 characters each for each judge.[168] The authors rely on compressions scores for each two-case pair and construct two different measures of consistency to rank judges in their sample.[169] On neither consistency measure, however, do their rankings conform to their ex-ante hypothesis that Judges Frank Easterbrook, Richard Posner, and Michael

---

163. The authors provide the following description of the features in their study:

1. *The type-token ratio*. This is the number of different words in an opinion (types) as a percentage of the total number of words in the opinion (tokens). For example, the five-word sentence "The boy threw the ball" contains four different words (*the* appears twice), so the type-token ratio is 0.80.

2. *Once-words*. This is the relative frequency of words that appear exactly once in an opinion. For example, in the sentence "The boy threw the big ball and the big man hit it," the once-words ratio equals 0.778, because seven of the nine types in the sentence (*boy*, *threw*, *ball*, *and*, *man*, *hit*, and *it*) appear once; the type-token ratio for the same 12-word sentence is 0.75, because it contains 9 different words.

3. *Average word length*, expressed as the mean number of letters per word in an opinion.

4. *Word length diversity*, expressed as the standard deviation of the number of letters per word in an opinion.

5. *Average sentence length*, expressed as the mean number of words per sentence in an opinion.

6. *Sentence length diversity*, expressed as the standard deviation of the number of words per sentence in an opinion.

7. *Footnote frequency*, expressed as the number of footnotes in an opinion per 1,000 words in the text of the opinion.

8. *Footnote length*, expressed as the total number of words in footnotes as a percentage of the total number of words in the text of the opinion.

Id. at 176–77 (footnote omitted).

164. The measure of similarity was simply the differences for each factor, squared and summed. *See id.* at 177–78.

165. *Id*. at 182.

166. *See* Choi & Gulati, *supra* note 123, at 1105.

167. *Id.*

168. *Id*. at 1106.

169. *Id*. at 1107–10.

Washington University Open Scholarship

Boudin would have relatively higher consistency scores compared to their peers.[170] The authors also conducted more straightforward analysis on rates of self-citation, variance in opinion length, and total opinion length.[171] The results of these analyses more closely accorded with their hypotheses that these three judges played a larger role in opinion drafting.[172]

In the course of an analysis of trends in opinion length on the Supreme Court, Black and Spriggs examine whether the increased role of clerks has played a causal role in a broader trend toward longer opinions.[173] For their study, the authors access "every orally-argued, signed, or per curiam majority opinion decided from 1791 to 2005, for a total of 26,715 opinions," and then, using a simple program, count the words in each opinion.[174] They divided the study period into four periods: the first, in which there were no clerks in the modern sense; the second, which lasted from 1886 to 1919, in which clerks acted as "stenographers"; the third, from 1920 to 1952, in which clerks "took on the role of an assistant"; and the fourth, post-1953, in which "law clerks went from being assistants to being something akin to law firm associates."[175] The period typology is drawn from Peppers, and we therefore will refer to these time segments as "Peppers groups."[176] While the authors do find that "opinions under the associate regime . . . are significantly longer than those in any of the earlier regimes,"[177] the relationship between clerkship regime and opinion length disappears when the time-series nature of the data is taken into account—the authors find no meaningful effect from clerkship regime that is not accounted for in the general increase of opinion length over time.[178] The authors also examine opinion length against anecdotal evidence concerning how many Justices on the Court relied on clerks for drafting in the years 1953–1990. They found that the total number of Justices that relied on clerks was not correlated with increased opinion length, once the time-series nature of the data was taken into account.[179]

170. *Id*.
171. *Id*. at 1111–20.
172. *Id*.
173. *See* Black & Spriggs, *supra* note 4, at 638–45.
174. *Id*. at 630.
175. *Id*. at 638–39.
176. *Id*. at 638 n.53 (citing PEPPERS, *supra* note 154).
177. *Id*. at 639.
178. *See id*. at 640–42 & nn.64–70.
179. *See id*. at 640 fig.4, 642–43.

Peppers and Zorn take an alternative quantitative approach to measuring clerk influence based on a survey of recent clerks' partisan affiliation.[180] The authors identified the clerks that served between 1940 and 2000 and mailed a survey to approximately 1,000 individuals for whom addresses could be identified.[181] The authors received 639 replies, and received information on partisan affiliation from 532.[182] Unsurprisingly, the authors find a high degree of correlation between the party affiliation of a Justice and the clerks that the Justice hires.[183] Even accounting for that high correlation, the authors find that the party affiliation of clerks was associated with case outcomes, with Democratic-affiliated clerks correlated with more liberal case outcomes.[184]

Rosenthal and Yoon, writing more recently, examine clerk influence on opinions in the Supreme Court by testing variability of writing style.[185] As we do in the analysis above, Rosenthal and Yoon use a function word approach to identify author style. The authors use a list of sixty-three function words for their analysis, and they draw texts from the majority opinions in the period 1991–2009.[186] The authors then develop several statistical tests for variability in an author's use of the content-free words and rank the Justices in their study period. They find that their hypothesis that Justice Kennedy has a greater stylistic variability than Justice Scalia was confirmed beyond a significance threshold of $p<.05$.[187] Rosenthal and Yoon also use the content-free word approach to address the problem of author attribution, finding that machine learning approaches achieved high levels of correct attribution.[188] Li et al. build on the author attribution methodology in a study of unsigned per curiam opinions in the Supreme Court.[189] That analysis was based on a set of content words and word phrases that were generated from the data based on predictive power.[190]

180. Todd C. Peppers & Christopher Zorn, *Law Clerk Influence on Supreme Court Decision Making: An Empirical Assessment*, 58 DEPAUL L. REV. 51, 53 (2008).

181. *Id.* at 60.

182. *Id.* at 60, 62.

183. *See id.* at 66 tbl.2.

184. *Id.* at 73 tbl.4. Of course, as the authors note, their findings are correlational, and "offer no support for any particular causal model." *Id.* at 75.

185. Rosenthal & Yoon, *supra* note 114. The authors also use similar statistical tools for purposes of author attribution.

186. *Id.* at 287–88. The text for their analysis was drawn from the Cornell Law School website. *Id.*

187. *Id.* at 293–94. The authors suspect that an author's increased variability is due to greater reliance on clerks.

188. *Id.* at 301.

189. Li et al., *supra* note 47, at 505.

190. *Id.* at 516.

Washington University Open Scholarship

Using two separate models, the authors achieve accuracy rates greater than 75% for known-author cases.[191]

Sulam uses attribution techniques to measure clerk influence.[192] The author uses the stylistic measures from Wahlbeck, Spriggs, and Sigelman and generates style profiles for clerks in the period 1986–1993 based on the certiorari pool memos authored by those clerks.[193] The author then uses non-opinion writings (such as articles or books) as well as opinions in the term prior to the tested term to develop stylistic measures for each Justice.[194] Based on these profiles, Sulam uses an attribution model to determine whether opinions in the tested term are more consistent with the clerk's or the Justice's stylistic profile.[195] In nearly all cases, the models predict the Justice as the author.[196] This study also used plagiarism software to determine the extent of borrowing from the certiorari pool memos, finding that there was relatively little borrowing compared to other sources, such as the parties' briefs and lower court opinions.[197]

B. Comparing Inter-Year Variability

Our analysis expands this prior research. Wahlbeck, Spriggs, and Sigelman as well as Rosenthal and Yoon test hypotheses about individual Justice pairs based on anecdotal evidence concerning reliance on clerks. While both studies reject the null hypotheses about a single Justice-Justice pair with a high degree of confidence, it is difficult to extrapolate their findings to a more general conclusion about clerk influence. Rosenthal and Yoon also find a time trend of increasing variability, which is consistent with clerk influence, but they did not test whether clerks, or some other time-dependent variable, accounted for the change. Choi and Gulati's computationally intensive test finds no greater stylistic consistency for reputed likely author judges; their more straightforward measures are loosely commensurate with their anecdotal hypothesis, but the evidence is

191. An incidental finding in their analysis was that Justices have greater intra-year than inter-year consistency, a finding that is consistent with clerks having some stylistic influence. *Id*. at 526. We leverage that finding in our analysis of the role of clerks.

192. Ian Sulam, Editor in Chief: Opinion Authorship and Clerk Influence on the Supreme Court 11 (May 9, 2014) (unpublished manuscript), *available at* http://icsulam.github.io/pdf/EditorInChief.pdf.

193. *Id.* at 13–14.

194. *Id.* at 13.

195. *Id*. at 16.

196. *Id*.

197. *Id*. at 18 (explaining that the rates of borrowing were: cert pool memo (3.72%); briefs (6.6%); lower court opinion (7.24%)).

not overwhelming. Black and Spriggs find no relationship between clerks and opinion length (the only variable that they studied). Peppers and Zorn find that clerk ideology has additional predictive value for case outcome, but do not establish that clerks causally influence outcomes.[198] Finally, Sulam finds little evidence of clerk influence. The most substantial hint of clerk influence over writing style comes from Li et al., but that finding was incidental to the authors' project (which concerned attribution of per curiam decisions) and was not explored in detail.[199]

We focus on variation in writing style and focus specifically on inter-year stylistic variability. Our model of clerkship influence is different than that used by Rosenthal and Yoon and Wahlbeck, Spriggs, and Sigelman. For those authors, variation is hypothesized to be a consequence of different clerks drafting different opinions in a given year. While this may well be a source of variation, it is extremely difficult to identify the roles of individual clerks, and there are reasons to believe that multiple clerks may be involved at some point in the drafting and editing process.[200] Our model differs in its focus on *clerk turnover* as the source of variability, rather than simply the existence of clerks within chambers. One of the peculiar features of the contemporary clerkship is that it is so short, typically lasting a mere year. We exploit this fact in our inter-year measure of variability. In addition, we construct a new measure of the total consistency of the Court. This measure examines the writing style consistency of the Court as an institution, rather than individual Justices. We then compare both of our new measures of consistency to the time periods used by Black and Spriggs (the Peppers groups) to determine if the changing nature of the clerkship institution has affected either intra-year consistency of the Court or inter-year consistency of individual Justices.

The first measure of variability that we introduce is *centroid distance*. This measure is based on the writings of the entire Court in each year. The distance between the feature vector for each text in a given year and the remainder of the Court's writings in that year are computed, and those

198. It is possible, for example, that accounting for the party affiliation of the clerks chosen by Supreme Court Justices to work in their chambers adds additional information about the Justice, essentially creating a more refined proxy for judicial values and ideology. If that is the case, the clerk does not necessarily exert any independent influence.

199. The authors of that study found that inter-year variation is greater than intra-year variation in writing style, which is consistent with our turnover model of clerk influence. *See* Li et al., *supra* note 47, at 525–26.

200. *See* Wahlbeck et al., *supra* note 120, at 170–72 (discussing Justice Powell's process).

Washington University Open Scholarship

distances are summed for each year.[201] This provides a measure of how tightly clustered the Court's style is in a given year: the greater the centroid distance, the bigger the stylistic "spread."

An OLS regression on this data examining the relationship between year and centroid distance found that there is a statistically significant relationship over the entire period.[202] Over time, the intra-year consistency on the Court has measurably increased.

To examine whether the overall trend toward greater consistency differed as the institution of the modern clerk developed, we conducted a structural break test on the data. A structural break is a concept from econometrics that is primarily used in time series analysis of macroeconomic data.[203] The point of a structural break analysis is to determine whether there has been an underlying shift in the data generating mechanisms, such that the distribution of data from the period after the "break" is systematically different than the distribution prior to the break. For example, the United States economy generates data on productivity, employment, and other economic variables. In our analysis, the data generating mechanism is the US Supreme Court. The structural break analysis is meant to examine whether the variable of interest—intra-year consistency—exhibited a different relationship to time during the periods when clerks played very different roles in chambers.

We first ran a Chow structural break test, which is a standard tool to determine whether there are changes in the relationships between time and another variable over different time periods.[204] For the potential break points, we used the Peppers groups. The results of the Chow tests are presented in Table 6.

---

201. For the centroid distance estimate, we use cosine similarity, which, like KL divergence, is a representation of distance in a multi-dimensional vector space. *See generally* Mihalcea, Corley & Strapparava, *supra* note 23 (using cosine similarity as baseline measure to evaluate alternative similarity measures in text analysis). To avoid confusion, we take [1 - cosine similarity] as the measure of "distance" so that larger distances are associated with greater difference.

202. The p-value is less than 0.01% and the R-squared value is 0.5.

203. *See generally* Bruce E. Hansen, *The New Econometrics of Structural Change: Dating Breaks in U.S. Labor Productivity*, 15 J. ECON. PERSP. 117 (2001).

204. *See id.* at 118; *see also* Gregory C. Chow, *Tests of Equality Between Sets of Coefficients in Two Linear Regressions*, 28 ECONOMETRICA 591 (1960).

TABLE 6: CHOW TEST ON CENTROID DISTANCE

First Period	*Second Period*	*F-value*	*P-value*
(1791–1885)	(1886–2008)	5.7	<0.01
(1791–1885)	(1886–1919)	6.7	<0.01
(1886–1919)	(1920–1952)	5.5	<0.01
(1920–1952)	(1953–2008)	3.13	<0.05
(1886–1919)	(1920–2008)	16.1	<0.01

The Chow test rejects the null hypothesis that there are no structural breaks in the centroid distance data at the Peppers groups dates. We conducted two additional tests. The first examines whether there is some structural break in the data and estimates the break date.[205] For this test, we did not specify a hypothesized date. The analysis rejected the null hypothesis of no structural break.[206] The estimated break date that was returned was 1926, very close to the year that clerks took on a greater substantive role, as indicated by the Peppers group transition from "stenographers" to "assistants." We also conducted Wald and likelihood ratio-based structural break tests for the three hypothesized break dates of 1885, 1919, and 1952.[207] Both tests confirmed breaks at those dates with a p-value of less than 0.01.

To attempt to better estimate the effects of clerks specifically, we develop a new measure of writing consistency, focused exclusively on inter-year variability in writing style. For purposes of our analysis, a chamber in a given year can be thought of as a "team" made up of a Justice and several clerks. A team co-produces the opinions in a given year. When clerks turn over, it changes the composition of the team. In chambers with a larger number of clerks that turn over more frequently, there will be a higher percentage of team turnover from year to year. Although some inter-year stylistic variability can be expected even with a single author, we hypothesize that clerk turnover will decrease inter-year consistency.

The dependent variable in our analysis is an inter-year *consistency score*. To construct the consistency score, we rely on the feature vectors

205. For this analysis, we used the Supremum Wald test in Stata. For additional background on this test, see generally Pierre Perron, *Dealing with Structural Breaks*, *in* 1 PALGRAVE HANDBOOK OF ECONOMETRICS: ECONOMETRIC THEORY 278 (Terence C. Mills & Kerry Patterson eds., 2006).

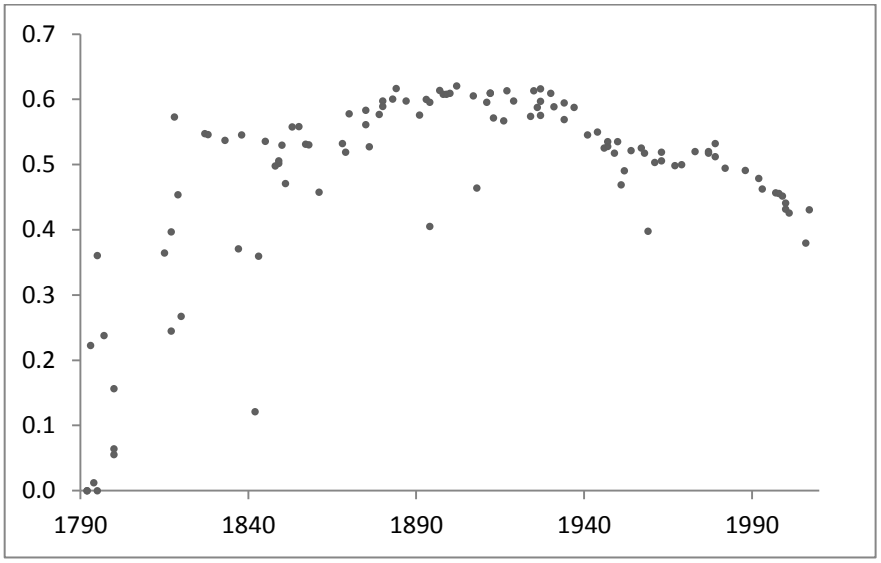
206. The p-value was less than 0.01.

207. *See generally* Perron, *supra* note 205.

Washington University Open Scholarship

based on the texts a Justice authored in each year of his or her tenure. So, for a Justice with a term from 1950 to 1959 (inclusive), there would be ten feature vectors for that Justice. To calculate the consistency scores, we calculate the KL divergence between each year's vector (interpreted as a probability distribution) and a feature vector based on the remainder of the Justice's writings. These values are scaled in a similar fashion as the similarity scores discussed above and are summed. Figure 7 presents each Justice's consistency scores ordered by median year of services.

FIGURE 7: CONSISTENCY SCORES BY MEDIAN YEAR



We examined the relationship between consistency score and the number of clerks that served in a Justice's chambers over the course of his or her tenure. The time trend was controlled for through a quadratic function. We controlled for each Justice's total production, under the theory that Justices who produce more may be more consistent, and there will be less statistical noise between years. We examined the interaction between clerks and time.[208]

208. For this analysis, we examine the period after 1885, with the introduction of clerks as "stenographers" under the Peppers grouping. We normalized the consistency scores using a cube function to construct *Consist*. The variable *Clerks* is the total number of clerks that served in a Justice's chamber, divided by that Justice's tenure on the Court, and normalized through a square root function. *Prod* is as described *supra* note 141. *Year* is year post-1885 (i.e., median year minus 1885).

TABLE 7: CLERK INFLUENCE ON CONSISTENCY

	Consist	*Consist*	*Consist*
Clerks	-0.044	-0.048	-0.079
	(10.32)**	(12.22)**	(4.19)**
Prod		0.020	0.017
		(4.19)**	(3.90)**
Year			0.002
			(2.55)*
*Clerks*Year*			0.001
			(3.34)**
$Year^2$			-0.000
			(3.35)**
_cons	0.184	0.177	0.160
	(39.10)**	(39.21)**	(10.61)**
Adj. R^2	0.63	0.71	0.77
N	65	65	65

* $p<0.05$; ** $p<0.01$, *t* statistic in parentheses

Table 7 reports the results of an OLS regression with median year and clerks per year as explanatory variables of consistency scores. In the first model, additional clerks are associated with a reduction in inter-year consistency. This relationship holds when a Justice's production is taken into account; that variable is significant and associated with increased consistency. Finally, in the full model, median year of service is significant, with a linear trend toward greater consistency and a squared trend toward lower consistency. The interaction between clerks and year indicates that clerks have had a less strong influence over time, perhaps indicating the declining marginal influence of an additional clerk as the Court has institutionalized a practice of each Justice having between four and five clerks.[209]

Because the institution of the modern clerkship was introduced gradually over time, it is difficult to fully disaggregate the effects of clerks

209. As with the historical value analysis above, *supra* Part IV.A, we examine the interaction variable by analyzing significance for *Clerks* as *Year* increase. *Clerks* are a significant predictor up to roughly 1940, which appears to indicate that the effect on writing style of additional clerks occurred during the period in which the clerks transitioned from stenographers to a more substantive role. Once they were largely integrated into chambers, effects from additional clerks appears to have diminished. Significance reappears at the 10% threshold, with the reverse sign on *Clerks*, in the years after 1995. This appears to indicate that adding an additional clerk to chambers for today's Justices would be unlikely to increase inconsistency from current levels, and might even have a positive effect (perhaps due to increased likelihood of inter-year holdovers).

Washington University Open Scholarship

from other time-dependent variables. We conduct one further structural break analysis based on the four Peppers groups, reported in Table 8.[210]

TABLE 8: CHOW TEST ON CONSISTENCY SCORE

First Period	*Second Period*	*F-value*	*P-value*
(1791–1885)	(1886–1919)	8.5946	<.01
(1886–1919)	(1920–1952)	5.5973	<.01
(1920–1952)	(1953–2008)	4.2538	<.01
(1886–1919)	(1920–2008)	5.8302	<.01

The first three tests identify the likelihood that the coefficients in the Peppers groups are the same, rejecting the null hypothesis in all cases. We also compare the period of clerks as stenographers to the latter two periods and reject the hypothesis that the time trends have the same coefficient.

It is worth remembering the difficulty of fully distinguishing the effects of unobserved time-related variables from the effects of clerks. Nevertheless, over the course of the twentieth century, the intra-year stylistic consistency of the Court as an institution has increased, while the inter-year consistency of writing style for individual Supreme Court Justices has declined. Over the same period of time, law clerks have become ever more integrated into the substantive work of the Court. Because the institution of the modern law clerk in the US Supreme Court evolved gradually over time, it is hard to know the degree to which clerks have contributed to changes in writing style, independent from some other set of time-related variables. But the information presented in this study is highly evocative. The turnover in each chamber every year of four clerks appears to reduce the writing style continuity that might otherwise exist, were Justices fully responsible for writing their own opinions. At the same time, these clerks appear to write similarly to their colleagues in different chambers, reducing the apparent stylistic differences between Justices. Thus, clerks can play a role in both increasing the consistency of the institutional voice of the Court and reducing the consistency of each Justice's individual voice.

210. Unlike in the case of centroid distance, we do not carry out the additional structural break analyses discussed above. For centroid distance, we had a single measure for every year except 1802 and 1811 and could construct an average for the two missing years without altering the data substantially. In the case of consistency scores, which are ordered according to the Justices' median year of service, there are many years missing, and a number of years with multiple entries.

CONCLUSION

Over the past several decades, there has been an explosion in quantitative analysis in legal scholarship. The lion's share of that research has focused on the statistical analysis of hand-coded cases, typically oriented toward legal content. This research has spurred a number of interesting debates about law, politics, and various influences (and non-influences) on judicial behavior. As computational text analysis has become more sophisticated and more accessible, scholars have begun to apply these tools to legal questions. The reduction of the costs of engaging in new forms of content analysis has allowed for new types of questions to be asked.

This Article follows an important thread of this larger overarching effort and examines the stylistic features of judicial writings in quantitative terms. We offer several important innovations. We construct a unique dataset of all Supreme Court cases in which dissents and concurrences are separated from main opinions; these texts are coded with identifying information for year of publication and authoring Justice. This substantial dataset, along with advances in computational power, allows us to conduct re-analysis of prior research to examine its validity in light of the new data. But in addition, we newly apply the "stylistic fingerprint" of frequency of function words (a known general proxy for writing style) to investigate trends as represented in the full decision corpus.

With this proxy variable, we test several hypotheses. The first hypothesis is that there is a style of the time in the Court, such that contemporaneous Justices write more similarly to their peers than to temporally remote Justices. Our analysis finds extremely strong support for this hypothesis. We also examine some potential causes, finding little support for the claim that highly cited Justices exert greater-than-average stylistic influence, but finding some non-conclusive support for the possibility that changes in legal content may account for some of the change. Finally, we examine the influence of judicial clerks on writing style. Specifically, we test two hypotheses concerning the modern institution of the rotating judicial clerk. First is the claim that this phenomenon has led to greater intra-year institutional writing consistency on the Court. Second is the claim that clerks have led to less inter-year individual writing consistency for the Justices. We find reasonably strong support for both propositions, which are consistent with greater influence by a pool of clerks whose writing styles are similar to others in their cohort, although it is impossible to exclude the effect of unobserved time-dependent variables on either.

Washington University Open Scholarship

Overall, we hope that the preceding analysis opens the door to new avenues of research of legal texts. In particular, we believe that there are a number of important and interesting research questions to be asked concerning the interaction of writing style on the Court with broader social and political trends. Specifically, Supreme Court opinions can be linked to other textual corpora, including appellate opinions and state opinions, as well as newspapers, published books, and Twitter feeds, among others. This textual data can be examined in light of more traditional social science sources, such as published economic (GDP, unemployment) or political (voting, electoral outcomes, campaign donations) data. Deploying the analytical tools described above, it may be possible to examine the interaction of writing style on the Court with courts more generally, with other forms of writing, and with broader social and political trends revealed in social science data. The possibilities of such analysis are exciting: human researchers can now find textual patterns that emerge at a macro-level, perceptible only recently with the digitization of vast textual corpora, the broad availability of massive computing power, and the continually evolving application of advanced concepts in mathematics and computer science to these "big" datasets. As these textual patterns become ever more perceptible, they offer the hope of new understandings in the use and evolution of language, from the staid chambers of the US Supreme Court to the unruly sprawl of the blogosphere.

http://openscholarship.wustl.edu/law_lawreview/vol93/iss6/6