

April 2018

The Power of Words: A Comment on Hamann and Vogel's Evidence-Based Jurisprudence Meets Legal Linguistics—Unlikely Blends Made in Germany

Mark C. Suchman

Follow this and additional works at: <https://digitalcommons.law.byu.edu/lawreview>

Recommended Citation

Mark C. Suchman, *The Power of Words: A Comment on Hamann and Vogel's Evidence-Based Jurisprudence Meets Legal Linguistics—Unlikely Blends Made in Germany*, 2017 BYU L. Rev. 1751 ().
Available at: <https://digitalcommons.law.byu.edu/lawreview/vol2017/iss6/15>

This Article is brought to you for free and open access by the Brigham Young University Law Review at BYU Law Digital Commons. It has been accepted for inclusion in BYU Law Review by an authorized editor of BYU Law Digital Commons. For more information, please contact hunterlawlibrary@byu.edu.

The Power of Words: A Comment on Hamann and
Vogel’s *Evidence-Based Jurisprudence Meets Legal
Linguistics—Unlikely Blends Made in Germany*

Mark C. Suchman*

By offering an international and interdisciplinary point of comparison, Hamann and Vogel demonstrate that current American forays into corpus-based legal scholarship reflect only a small sliver of the full range of possibilities for such research. This Comment considers several key branching points that may lie ahead, as the nascent literature begins to mature. In particular, the Comment examines two vexing ambiguities in the corpus-linguistic agenda: the first centers on the ambiguous meaning of legal “empiricism”; the second, on the ambiguous relationship between words and actions. To achieve its full potential, legal corpus linguistics will need to move beyond mere description, to identify patterned configurations, to interpret cultural meanings, and to trace causal processes. To do so effectively, researchers will need to look beyond legal corpora alone, to explore the varied and complex relationships between texts and acts, and between legal institutions and the surrounding society.

CONTENTS

I. INTRODUCTION	1752
II. THE AMBIGUITY OF LINGUISTIC EMPIRICISM.....	1753
A. Corpus Linguistics as Empirical Method.....	1754
B. Corpus Linguistics as Empirical Subject.....	1754
C. Corpus Linguistics as Empirical Purpose	1757
III. THE AMBIGUITY OF LINGUISTIC CAUSATION.....	1761

* Professor of Sociology, Brown University.

IV. THE AMBIGUOUS PROMISE OF LEGAL CORPUS LINGUISTICS	1765
V. CONCLUSION	1768

I. INTRODUCTION

Hanjo Hamann and Friedemann Vogel provide a useful reminder that legal corpus linguistics is a global product, not so much *made in America*¹ as blended and packaged for local tastes from ingredients sourced through a long international supply chain. Not being a corpus linguistics scholar, I cannot judge the provenance or quality of those ingredients. And not being a comparative law scholar, I cannot judge their suitability for local German tastes. However, as an empirical law and society researcher and a methodological omnivore, I can certainly appreciate an intriguing imported brew, even one unfamiliar to my palate and hard to find in my neighborhood.²

I offer the foregoing disclaimer and appreciation as a preface because my comments below center less on the Hamann and Vogel essay, per se, than on the broader enterprise of legal corpus linguistics—and the themes of this symposium—as refracted through the Hamann and Vogel article’s international and interdisciplinary lens. As a methodology, American legal corpus linguistics is still in its infancy, and like most infants, it is fascinated by the shiny objects closest at hand. But Hamann and Vogel’s German comparison foreshadows some of the challenges, anxieties, and identity crises that legal corpus linguistics will (or should) face as the field comes into its adolescence. My commentary focuses on two such challenges in particular, both involving ambiguities in the invocation of linguistic methods in the empirical study of law: the first centers on the ambiguous meaning of legal “empiricism”; the second, on the ambiguous relationship between legal words and legal actions. Corpus

1. Nor “[m]ade in Germany,” for that matter, although on this point Hamann and Vogel certainly deserve a bit of poetic license. See Hanjo Hamann & Friedemann Vogel, *Evidence-Based Jurisprudence Meets Legal Linguistics—Unlikely Blends Made in Germany*, 2017 BYU L. REV. 1473, 1473.

2. *Five Questions for Nina Bandelj*, ACCOUNTS (Am. Sociological Ass’n, Wash. D.C.), July 2013, at 2, 3 (“I am a methodological omnivore, driven more by empirical questions than by specific analytic methods.”).

linguistics offers a powerful new tool for the study of law, but the nature of its promise will depend in important ways on who embraces it and how.

II. THE AMBIGUITY OF LINGUISTIC EMPIRICISM

To describe a piece of scholarship as “empirical” is to highlight or assert the primacy of objective and systematic observation (or at least of *intendedly* objective and systematic observation) as a warrant for the scholar’s truth claims. Polemics, sermons, opinion pieces, and philosophical treatises are nonempirical because their primary warrants are passion, morality, authenticity, and logic, respectively, rather than systematic observation. Surveys, tabulations, experiments, focus groups, and ethnographies are all empirical because their primary warrants are objectivity and systematicity—even though their respective foci, methods, and epistemologies diverge quite radically. As a shorthand, one could say that scholarship is empirical if the scholar commits to a method of observation and a style of argumentation that allows external reality to discomfit the scholar’s prior expectations.

To understand the place of corpus linguistics in legal scholarship, one must recognize that the corpus linguist aspires to be empirical, whereas the traditional, doctrinal legal scholar does not. Even when a linguist and a doctrinalist work from the same corpus of legal text, the primary warrants that the two adduce will differ significantly. The linguist commits (albeit perhaps only tacitly) to the methodological premise that observation of a particular pattern in the text will warrant a particular conclusion; the doctrinalist looks to the text for inspiration, but ultimately warrants his or her argument on the basis of morality, logic, or untested assumptions about policy impact—not on the basis of objective and systematic observation. A linguistic argument is “correct” when it is borne out by observed patterns in the corpus; a doctrinal argument can be “correct” even if the corpus is silent, divided, or uniformly wrongheaded—indeed, those are precisely the conditions under which the doctrinalist’s skills may be most needed.

As sharp as this distinction between empiricism and doctrinalism may seem in principle, however, the dividing lines often become much hazier in practice—especially for legal scholars working from legal

texts. Doctrinal legal scholars often summarize and synthesize a corpus of prior texts, at least to contextualize other less empirical modes of argumentation; and empirical legal scholars often explore doctrinal implications, at least to motivate particular systematic observations. Legal corpus linguistics necessarily occupies this twilight zone, and to avoid misunderstandings, legal corpus linguists must give careful consideration to the ways in which their enterprise is (or is not) “empirical.”

A. Corpus Linguistics as Empirical Method

The first and perhaps easiest question is methodological: If empiricism requires objective and systematic observation, then *how* should legal corpus linguists observe? Here, Hamann and Vogel faithfully capture the spirit of the present symposium when they assert that “[m]ethodologically, [corpus research in law] relies on big data empiricism.”³ In the interest of brevity, I will not belabor this point, except to note that quantitative, computation-intensive “big data empiricism” is not the only empirical method for observing legal language. Some individual texts may have attributes that are qualitatively revealing despite being quantitatively rare. Some quantitative patterns become more interesting when juxtaposed with qualitative evidence on a particular historical period or social community. And some linguistic “big data” become interpretable only when linked, through demographic or geographic metadata, with other data sources, big and small. So, although academic politics may push legal linguists to embrace the cachet of big data empiricism, scholarly rigor cautions against fetishizing the computational analysis of large corpora to the exclusion of other complementary empirical methods.

B. Corpus Linguistics as Empirical Subject

A harder definitional question is substantive: If empiricism requires objective and systematic observation, then *what* should legal corpus linguists observe? The obvious answer, as Hamann and Vogel assert, is legal language—treated “not merely as the medium but as

3. Hamann & Vogel, *supra* note 1, at 1474.

the object of study.”⁴ But this answer can be no more than partial. Even if one accepts, arguendo, Hamann and Vogel’s claim that “[w]e cannot speak about legal norms without thinking about how they are constructed by and through speech and texts,”⁵ one must also take seriously the injunction from Legal Fact Research that “two sets of facts are to be taken into account: the life situations to be regulated, and the politics and valuations guiding the regulating activities”; and that, “[t]o know them both, the jurist must put himself *outside of the body of the existing rules*.”⁶ In other words, for many interesting and important empirical questions about law, the study of legal language, in itself, may not be enough.

To get a fuller sense of what else legal linguists might observe, consider the following simple typology of causes and effects:

Table 1: A typology of legal and extra-legal processes in language

Effect: Cause:	Law world	Social World ⁷
Law world	1: Autopoietic Law	2: Legal Impact
Social World	3: Legal Politics	4: Informal Norms

4. *Id.*

5. *Id.* at [114].

6. *Id.* at [104] (quoting Max Rheinstein, *Comparative Law and Conflict of Laws in Germany*, 2 U. CHI. L. REV. 232, 253 (1935) (emphasis added)).

7. I use the label “social world” as shorthand for those aspects of social life that occur largely outside formal legal institutions—whether in spontaneous daily interactions or in such extra-legal institutional spheres as business, medicine, or politics. This domain of extra-legal social activity is approximately equivalent to what German social philosophers would call “*lebenswelt*,” or the “lifeworld.” See, e.g., JURGEN HABERMAS, *THE THEORY OF COMMUNICATIVE ACTION: VOLUME TWO: LIFEWORLD AND SYSTEM: A CRITIQUE OF FUNCTIONALIST REASON* (Thomas McCarthy trans., Beacon Press 1987) (1981); EDMUND HUSSERL, *THE CRISIS OF EUROPEAN SCIENCES AND TRANSCENDENTAL PHENOMENOLOGY* (David Carr trans., Northwestern Univ. Press 1970) (1954). Elsewhere in the social sciences, it carries such labels as “civil society” and “everyday life.” See, e.g., JEAN L. COHEN & ANDREW ARATO, *CIVIL SOCIETY AND POLITICAL THEORY* (1994); ERVING GOFFMAN, *THE PRESENTATION OF SELF IN EVERYDAY LIFE* (1959). Of course, this distinction between a “law world” and a “social world” should not obscure the fact that formal legal institutions are themselves social. Nor should it obscure the fact that extra-legal social life is inevitably shaped and constituted by law. Nonetheless, the distinction is often analytically useful (and for most people intuitively obvious), because the law world is effectively bounded and demarcated from the extra-legal social world by linguistic barriers and credentialing requirements that can be bridged only via substantial investments in translation and re-presentation. Cf. *AUTOPOIETIC LAW: A NEW APPROACH TO LAW AND SOCIETY* (Gunther Teubner ed., 1988).

In Cell 1, we find the core domain of Computer Assisted Legal Linguistics (CAL²), as Hamann and Vogel have defined it: “[A]nalyz[ing] what judges and other jurists write down and how they thus exert power and authority”⁸ or, somewhat more broadly, “studying quantitatively how lawyers create and apply legal rules by ‘treating doctrine as a quantitative unit.’”⁹ Significantly, even within this “autopoietic”¹⁰ heartland, “[g]aps remain”¹¹—perhaps most significantly in studying legal language outside the publicly documented context of litigation, in the more occluded precincts of contractual drafting, statutory and regulatory rulemaking, corporate compliance activities, and alternative dispute resolution.

However, the other cells of the table suggest an even broader array of empirical objects that legal corpus linguistics might encompass: In Cell 3, we have Nussbaum’s “systematic search into the social, political and other fact conditions which give rise to the individual legal rules”¹²; and in Cell 2, his “examination of the social, political and other effects of those rules.”¹³ The former is the domain of political scientists in the Law and Society movement, as well as of their forebears in the Legal Realist tradition; the latter is the domain of policy scholars in the Empirical Legal Studies movement, and also of their forebears in the older tradition of Sociological Jurisprudence. Cell 4, the most remote from traditional legal scholarship, is nonetheless quite familiar to legal anthropologists and sociologists who study “private [normative] ordering,”¹⁴ “[lay] legal

8. Hamann & Vogel, *supra* note 1, at 1494.

9. *Id.* at 1482 (quoting Bernard Trujillo, *Patterns in a Complex System: An Empirical Study of Valuation in Business Bankruptcy Cases*, 53 UCLA L. REV. 357, 363 (2005)).

10. *Id.* at 1488; *see also* AUTOPOIETIC LAW, *supra* note 7.

11. Hamann & Vogel, *supra* note 1, at 1490.

12. *Id.* at 1477 (quoting Arthur Nussbaum, *Fact Research in Law*, 40 COLUM. L. REV. 189, 197 (1940)).

13. *Id.*

14. *E.g.*, Marc Galanter, *Justice in Many Rooms: Courts, Private Ordering, and Indigenous Law*, 19 J. LEGAL PLURALISM 1, 23 (1981); *see also, e.g.*, ROBERT C. ELLICKSON, *ORDER WITHOUT LAW: HOW NEIGHBORS SETTLE DISPUTES* (1991).

consciousness,”¹⁵ and “informal dispute resolution.”¹⁶ Although these three cells rarely generate well-archived corpora of explicitly legal text, all three are certainly relevant to the empirical study of legal language. Most importantly, Cells 2 and 3 foreground the process of translation between legal and extra-legal institutions, while Cell 4 (juxtaposed against Cell 1) highlights the sociolinguistic forces that can push the law world and the social world out of alignment.¹⁷ So, although practical logistics may tempt legal linguists to embrace the convenience of purely judicial corpora, scholarly rigor cautions against fetishizing the language of the courtroom to the exclusion of other complementary discourses.¹⁸

C. Corpus Linguistics as Empirical Purpose

A third definitional question for linguistic empiricism is motivational: in deciding how and what to observe, legal corpus linguists should also ask *why* (and to *whom*) such observations would be of interest. As Hamann and Vogel note, “Doctrinal analysis remains the mainstay of . . . legal academia, so empirical research strands are rarely perceived (let alone absorbed) by the field’s core authorities.”¹⁹ To alter this state of affairs, legal empiricists either must focus their

15. *E.g.*, PATRICIA EWICK & SUSAN S. SILBEY, *THE COMMON PLACE OF LAW: STORIES FROM EVERYDAY LIFE* (1998); SALLY ENGLE MERRY, *GETTING JUSTICE AND GETTING EVEN: LEGAL CONSCIOUSNESS AMONG WORKING-CLASS AMERICANS* (1990); Susan S. Silbey, *After Legal Consciousness*, 1 ANN. REV. L. & SOC. SCI. 323 (2005).

16. *E.g.*, Lauren B. Edelman & Mark C. Suchman, *When the “Haves” Hold Court: Speculations on the Organizational Internalization of Law*, 33 L. & SOC’Y REV. 941, 949–53 (1999); Henry S. Farber & Michelle J. White, *A Comparison of Formal and Informal Dispute Resolution in Medical Malpractice*, 23 J. LEGAL STUD. 777, 777–81 (1994).

17. These empirical problematics are important, but hardly new: The pioneering legal anthropologist Paul Bohannan would have recognized tensions along the 1–4 diagonal as instances of law being “out of phase with society,” and cycles along the 2–3 diagonal as processes of “double institutionalization.” Paul Bohannan, *The Differing Realms of the Law*, 67 AM. ANTHROPOLOGIST, Dec. 1965, at 33, 34–37.

18. Several pieces in this symposium tacitly embrace this prescription, usefully comparing legal corpora to corpora drawn from other domains. Indeed, the necessity of such comparison lies at the heart of the call for applying corpus linguistics to questions of “original public meaning” in constitutional and statutory interpretation. *See, e.g.*, James C. Phillips, Daniel M. Ortner & Thomas R. Lee, *Corpus Linguistics & Original Public Meaning: A New Tool to Make Originalism More Empirical*, 126 YALE L.J.F. 21, 29 (2016) (“[A] general corpus could confirm that a word or phrase is a legal term of art by showing that its ordinary meaning differs from its meaning in legal sources.”).

19. Hamann & Vogel, *supra* note 1, at 1491.

empirical methods on questions of doctrinal relevance or must enunciate new scholarly questions that are sufficiently compelling to carve a non-doctrinal enclave within or alongside the legal academy. The former has been the favored strategy of the Empirical Legal Studies movement; the latter, the strategy of Law and Society scholars in the social science disciplines.²⁰ Either approach, however, demands a clear-eyed recognition of the differences among several common empirical modes: the descriptive, the configurational, the interpretive, and the causal.

The simplest mode and motive for empirical research is raw *description*: *X* occurs ten percent of the time; *Y* occurs twenty-five percent of the time, etc. Unfortunately, however, raw description is a thin gruel, and those who peddle it are derided as “crass empiricist[s]”²¹ for good reason. At the dawn of a new methodology, such unstructured fact-inventories may have a certain appeal, as anyone who has ever introduced a preschooler to a magnifying glass will attest; but to transmute a collection of empirical observations into a body of empirical knowledge requires both an organizing conceptual framework and a purposeful investment in synthesis.

Thus, a second mode of empiricism might be termed *configurational*²²—seeking to discern and characterize recurring patterns or configurations in a body of empirical observations. Much of computational corpus linguistics, including topic modeling and collocation analysis, is configurational in this sense (as is personality typing in psychology, social network mapping in sociology, morphometric taxonomy in biology, and bit-string fingerprinting in pharmacology). Significantly, as Hamann and Vogel note, legal

20. See Mark C. Suchman & Elizabeth Mertz, *Toward a New Legal Empiricism: Empirical Legal Studies and New Legal Realism*, 6 ANN. REV. L. & SOC. SCI. 555 (2010).

21. William Outhwaite, *Naturalisms and Anti-Naturalisms*, in KNOWING THE SOCIAL WORLD 22, 35 (Tim May & Malcolm Williams eds., 1998).

22. The label “configurational analysis” has emerged largely independently (and with somewhat disparate meanings) in several distinct fields of scholarship. Here, I use it generically, to refer simply to the empirical identification of recurrent patterns; but this usage maps only loosely onto the terminology of any particular disciplinary tradition. See, e.g., Alan D. Meyer, Anne S. Tsui & C. R. Hinings, *Configurational Approaches to Organizational Analysis*, 36 ACAD. MGMT. J. 1175 (1993); Earl S. Schaefer, *A Configurational Analysis of Children’s Reports of Parent Behavior*, 29 J. CONSULTING PSYCHOL. 552 (1965); Andreas Wimmer, Lars-Erik Cederman & Brian Min, *Ethnic Politics and Armed Conflict: A Configurational Analysis of a New Global Data Set*, 74 AM. SOC. REV. 316 (2009).

linguistics in this configurational mode can be qualitative as well as quantitative, encompassing small-N methods such as hermeneutics, discourse analysis, and linguistic ethnography, as well as large-N methods of “calculating recurrent speech patterns . . . using big data and semi-automated algorithms.”²³ As an empirical agenda, these configurational approaches seek to problematize the familiar, revealing hidden regularities in texts that have hitherto been studied only for their argumentation, not their semantics. The patterns that emerge can then be sorted into typologies and their incidence can be mapped and theorized over time or across social and geographic space.

Whereas configurational analysis seeks to problematize the familiar, a third empirical mode, *interpretation*, seeks to familiarize the problematic. Most closely associated with the humanities and the “humanistic social sciences,”²⁴ interpretation is a common agenda in any field that takes culture (including linguistic culture) seriously. At its core, interpretation attempts to explore and clarify the meaning in social “texts”—broadly defined to include not only formal writings, artistic works, and staged performances, but also informal speech acts, visual displays, material artifacts, and customary practices. Given that “meaning” is ultimately an introspective psychic construct, many commentators dispute whether interpretive agendas qualify as “empirical” at all, and legal empiricists are often quite explicit in distinguishing their own preferred forms of scholarship from traditional doctrinal interpretation. But this reaction conflates two very different forms of interpretation: interpretation in the humanities and interpretation in the social sciences.

Interpretative research in the humanities, like doctrinal interpretation in law, is a largely nonempirical endeavor. The scholar may begin with systematic observations (readings) of a text; but the goal is less to capture an objective, external reality than to inspire and provoke new resonances in the scholar’s own mind and in the minds

23. Hamann & Vogel, *supra* note 1, at 1493.

24. The modifier “humanistic” is often used to designate those social science disciplines—including anthropology, history, and linguistics—that emphasize particularism versus generalization, description versus abstraction, cultural contingency versus physical determinism, empathetic understanding versus nomothetic prediction, and hence interpretivism versus positivism. Cf. George Casper Homans, *The Humanities and the Social Sciences*, AM. BEHAV. SCIENTIST, Apr. 1961, at 3, 3.

of his or her interpretive interlocutors. In this humanistic mode, there is no “wrong” interpretation of *Hamlet* (or of the U.S. Constitution); just more or less engaging ones.

In contrast, interpretive research in the social sciences is often objective and systematic to the core. The goal here is to uncover the “emic”²⁵ meaning of a particular text within a particular cultural community in a way that members of that community would recognize and endorse. In this latter mode, an interpretation is “wrong” if it cannot be confirmed by other researchers studying the same subject community with the same methods. Far from being nonempirical, such interpretive research is essential to understanding the recurring behavioral patterns of self-aware, culture-bearing human subjects. Even such apparently objective descriptive statistics as the race and gender ratios of a survey sample become empirically tractable only after one resolves (by pre-testing or by assumption) the inherently interpretive issue of what the relevant questionnaire items meant to the survey respondents. And interpretive empiricism is all the more necessary when attention turns to the intendedly meaningful texts produced and received by shifting discourse communities over protracted periods of time. Indeed, for many legal linguists, interpretive empiricism may be the main event, not a mere warm-up act.

Despite the importance and difficulty of configurational and interpretive empiricism, many researchers aspire to a fourth, even more demanding mode of analysis, seeking to use objective, systematic observation to resolve questions of *causality*. Causal claims take the form “if *X*, then *Y*”—or, more forcefully, “because of *X*, therefore *Y*”—and causal empiricism seeks to test such propositions by observing how various *X*s (independent variables) correlate with various *Y*s (dependent variables); preferably under statistically or experimentally controlled conditions that rule out various prior, mediating, coincidental, or reciprocal factors. Taken literally, this goal is more daunting than it might initially seem: the more science learns about quantum uncertainty, physio-psycho-social over-determination, and macro-historical path dependence, the more philosophers debate

25. Anthropologists traditionally distinguish between the emic understandings held by research subjects and the etic explanations offered by outside scholars. See Marvin Harris, *History and Significance of the Emic/Etic Distinction*, 5 ANN. REV. ANTHROPOLOGY 329 (1976).

whether causal propositions are ontologically meaningful and epistemologically knowable *even in principle*, let alone technically resolvable in practice. Nonetheless, humans seem predisposed to understand the world in causal terms, and empirical researchers routinely and unreflectively slip into the causal register even when neither the available methods nor the available data merit such presumptions. Currently, legal corpus linguistics seems somewhat less susceptible to this temptation than other empirical approaches that get less mileage out of uncovering novel configurations and interpreting shrouded meanings; but as the field's methods and data mature, its causal ambitions may grow as well. Hamann and Vogel posit that "legal norms . . . are *constructed* by and through speech and texts,"²⁶ and they envision an agenda for CAL² that "analyzes what judges and other jurists write down and how they thus *exert* power and authority."²⁷ These are causal claims about the force of words. To test such claims empirically, legal corpus linguists will need to address philosophical and methodological challenges that go well beyond anything attempted to date.²⁸

III. THE AMBIGUITY OF LINGUISTIC CAUSATION

Although I do not propose to resolve the causal ambiguities of corpus linguistics here, a brief illustration of one central puzzle may

26. Hamann & Vogel, *supra* note 1, at 1486 (emphasis added).

27. *Id.* at 1494 (emphasis added).

28. For a more concrete illustration of how our four empirical purposes differ, consider an imaginary traffic planner who arrives in the United States from a distant country. Walking the streets, she observes various plaques, mounted in various ways, and bearing various markings. Operating in the *descriptive* mode, she begins to record her observations in a notebook, perhaps using categories based on the (possibly irrelevant) practices of her native land. After a while of this, she switches to the *configurational* mode, noting that one particular pattern seems to recur more often than others—an octagonal plaque, painted red, bearing the white markings "STOP." Perhaps this is a culturally meaningful symbol, she thinks. So, slipping into the *interpretive* mode (and miraculously developing a fluency in spoken English), she begins to ask passersby what this configuration means to them. Most describe it as a stop sign, and some go on to explain the legal and normative implications. But our intrepid heroine understands the difference between beliefs and behaviors; and she knows that in her homeland, traffic regulations are often widely publicized, but equally widely violated. So, to round out her empirical project, she turns to the *causal* mode, designing a series of experiments to determine the effects of such signage on American motorists of various ages and genders, under various traffic conditions, at various times of day, and in various stages of inebriation. No single element of this imaginary research design would generate much knowledge on its own; however, skillfully conjoined, they might yield an empirical understanding of American traffic patterns that would surpass the lay intuitions of even a lifelong American driver.

help to demonstrate the magnitude and importance of the challenge. Specifically, any evenhanded approach to legal empiricism must acknowledge that what we commonly call “the Law” is neither exclusively a corpus of disembodied texts (constitutions, statutes, judicial opinions, etc.) nor exclusively a system of embodied practices (legislating, policing, litigating, adjudicating, punishing, etc.). Rather, Law is a complex institutional amalgam, blending words and actions in multiple causal (and noncausal) concoctions.

To grasp the implications for legal linguistics, consider a second simple typology of causes and effects:

Table 2: A typology of linguistic and nonlinguistic processes in law

Effect: Cause:	Actions	Words
Actions	1: Realism	2: Performances
Words	3: Performatives	4: Textualism

In Cell 1, we find the core domain of Legal Realism (and of many more recent social-scientific approaches to the study of law) where the focus falls not on what legal language says “in books” but on what legal institutions do “in action.”²⁹ In Oliver Wendell Holmes’s famous dictum, “If you want to know the law and nothing else, you must look at it as a bad man, who cares only for the material consequences which such knowledge enables him to predict.”³⁰ To Holmes and his fellow Realists, an excessive focus on legal “phraseology,” “language,” “talk[],” and “words” invites us “to drop into fallacy.”³¹ “[I]f we take the view of our friend the bad man we shall find that he does not care two straws for the axioms or deductions, but that he does want to know what the . . . courts are likely to do in fact.”³² Although

29. Roscoe Pound, *Law in Books and Law in Action*, 44 AM. L. REV. 12 (1910).

30. Oliver Wendell Holmes, Justice, Supreme Judicial Court of Mass., Address at the New Hall of the Bos. Univ. Sch. of Law: The Path of the Law (Jan. 8, 1897), *in* 110 HARV. L. REV. 991, 993 (1997).

31. *Id.*

32. *Id.* at 994.

Holmes's somewhat dogmatic presentation elides important aspects of real-world human psychology and organizational behavior, it stands as an important reminder that some aspects of law—the police officer's baton, the lawyer's bill, the defendant's race, the prison's zip code—do not “live[] in language.” Many respected researchers have devoted their careers to producing empirical knowledge about legal phenomena that have very little to do with “speech and texts,”³³ and it is almost impossible to imagine a linguistic research design that could support causal claims about the impact of legal text without controlling for such nontextual pathways.

This cautionary note is particularly important for linguists who purport to operate in Cell 4, examining the causal effects of particular textual patterns on other textual patterns. The big-data computational cachet of corpus linguistics encourages relatively bold leaps from correlational data about the co-occurrence of words to causal assertions about the force of words. But sometimes words are epiphenomena, mere garnishes on outcomes that are predetermined by other nonlinguistic factors. If an antidiscrimination statute says that physical abilities may not be used as hiring criteria unless they are “job-related for the position in question and . . . consistent with business necessity,”³⁴ then regulatory guidance documents, employee handbooks, and applicant rejection letters may also begin to use the phrase “job-related and consistent with business necessity”³⁵; but this may tell us almost nothing about changes in actual hiring and enforcement patterns or in cultural understandings of what constitutes a valid job requirement or a viable business model. Indeed, it may not even tell us whether the statutory change sparked the other documentary changes, or whether instead the legislative use of “job-related and consistent with business necessity” simply encapsulated an understanding that had already emerged in political protests, labor negotiations, and out-of-court legal settlements, before being verbalized in law. Certainly, textualist causation, of the sort envisioned by Cell 4, *does* sometimes occur; but legal empiricists may vastly

33. *Contra* Hamann & Vogel, *supra* note 1, at [114].

34. Americans with Disabilities Act of 1990, 42 U.S.C. § 12112(b)(6) (2012).

35. *See, e.g.*, Enforcement Guidance on Disability-Related Inquiries and Medical Examinations of Employees Under the Americans with Disabilities Act (ADA), EEOC Notice No. 915.002 (July 27, 2000).

overstate its frequency and importance if they become so enamored of linguistic corpora that they lose sight of the other cells in the table.

In all fairness, this cautionary note also works the other way around, and the Hamann and Vogel article, along with the rest of this symposium, should usefully alert conventional Cell 1 scholars to the possibility of textual causation—as well as to the ability of corpus linguistics to illuminate such causation. Significantly, the potential for productive dialog between conventional legal empiricists and linguistic legal empiricists may be greatest, not in Cell 1 (where linguists' skills may strike nonlinguists as unnecessary) nor in Cell 4 (where linguists' findings may strike nonlinguists as inconsequential), but rather in Cells 2 and 3, where words and actions intertwine.

In Cell 2, we find words as performances—scripted or improvised displays that flow from nonverbal events. At the extreme stand exclamations: the verbal “Ouch!” that follows a physical hammer striking a physical thumb. But of greater research interest are the verbal accounts that people provide for their actions,³⁶ and the verbal narratives that people construct around their experiences.³⁷ For obvious reasons, such accounts and narratives play a crucial role in interpretive inquiries into the emic meaning of nonverbal events. More subtly, accounts and narratives also play an important role in tracing how meanings emerge and propagate within a community—and for understanding how extra-legal injuries evolve into private legal disputes and public political movements, which in turn evolve into public policies and, thence, into new legal regimes.³⁸

Finally, in Cell 3, we find processes in which “the textual work of lawyers [should be] taken seriously and analyzed . . . as a . . . performative speech act.”³⁹ Performative texts neither merely evoke textual responses (as in Cell 4) nor merely describe states of the world (as in Cell 2) but rather call into existence *new* states of the world,

36. Courtroom testimony is an obvious example, but account-formation is a fundamental social process that occurs in innumerable nonlegal settings as well. Marvin B. Scott & Stanford M. Lyman, *Accounts*, 33 AM. SOC. REV. 46 (1968).

37. Patricia Ewick & Susan S. Silbey, *Subversive Stories and Hegemonic Tales: Toward a Sociology of Narrative*, 29 L. & SOC'Y REV. 197 (1995).

38. William L.F. Felstiner, Richard L. Abel & Austin Sarat, *The Emergence and Transformation of Disputes: Naming, Blaming, Claiming . . .*, 15 L. & SOC'Y REV. 631 (1980).

39. Hamann & Vogel, *supra* note 1, at 1494 (distinguishing CAL² from both traditional doctrinal analysis and traditional Legal Realism).

conditioned on the textual utterance. The archetype is the pronouncement of guilt, an utterance that instantaneously transmutes a defendant into a convict.⁴⁰ Other legal performatives are less celebrated but perhaps even more consequential: the definition that divides a previously undifferentiated extra-legal category into the legally sacred and the legally profane; the ruling that transfers the burden of proof from plaintiff to defendant; the subtle linguistic signaling that makes one phrase into a holding of law and another into mere obiter dicta; the opinion-of-counsel letter that makes a desultory background-check into “due diligence.” Indeed, a large amount of traditional doctrinal scholarship involves efforts (albeit usually nonempirical) to identify which legal utterances are performative and which are not.

Arguably, as Hamann and Vogel suggest, Cell 3 may be the domain in which corpus linguistics has the most to offer to empirical legal scholarship. Legal performatives have real-world causal force, yet they are rarely captured by traditional social science methodologies. In making such utterances objectively and systematically observable, corpus linguistics brings them within the perceptorium of legal empiricism and hence within the scope of socio-legal analysis. Nonetheless, much hard conceptual and methodological work remains to be done. In particular, performances and performatives often intertwine with one another and with the institutionalized practices that they accompany. Few would deny that legal institutions exhibit an elective affinity with performative speech, but we currently lack a strong theory of when and why legal causation works through such words—and when and why, instead, it works through the nonverbal devices of space and time and wealth and force.

IV. THE AMBIGUOUS PROMISE OF LEGAL CORPUS LINGUISTICS

Hamann and Vogel describe CAL² as aspiring to become “a new transdiscipline, combining big-data corpus research with its own epistemological focus on the language of legal practice, distinct from the plain meaning and original intent traditions.”⁴¹ This is a

40. It is interesting, in this context, to note that whereas jurors “announce” a verdict, performing a speech ritual that merely reports prior deliberative actions, judges “pronounce” a sentence, uttering performative words that invoke a new state of the world.

41. Hamann & Vogel, *supra* note 1, at 1475.

commendably ambitious agenda. If only more legal scholars would attempt such big things! However, Hamann and Vogel's ambition can be read in several distinct ways.

At times, they seem to envision CAL² as a scholarly method for better discerning the Law—that is, as an empirical toolkit for clarifying legal concepts by revealing when and how particular terms and constructions appear with one another or are used or avoided by particular speakers. This resembles the use, illustrated elsewhere in this symposium, of extra-legal corpora to assess “plain meaning” and “original intent”; and Hamann and Vogel acknowledge that these cognate American efforts “probably differ [from CAL²] more in their source material (laypeople language vs. expert language) than in their conceptual underpinning.”⁴² Significantly, this distinctly scholarly vision of legal corpus linguistics also resembles traditional doctrinal legal scholarship—albeit with a fortifying infusion of big-data empiricism. CAL², American “public meaning” research, and traditional doctrinalism all rest on the ontological premises that: (a) legal language has a single true meaning, (b) this true meaning is systematic enough to form an orderly conceptual structure, and (c) this conceptual structure provides a useful normative foundation (or at least a necessary degree of transparency and predictability) to guide public policy. CAL², “public meaning” linguistics, and doctrinalism also all rest on the epistemological premise that the true meaning of a legal text can be discerned by reading it in the light of other, related texts. Admittedly, the three approaches differ in their views about which types of texts one must include, how many texts one must examine, and to what degree the examination must be statistical versus impressionistic.⁴³ But these differences are no greater than the other divisions of topic, ideology, and style that typically divide law faculties; and peaceful coexistence—or even fruitful collaboration—seems

42. *Id.* at 1495.

43. To some degree, traditional doctrinalism also stands apart from corpus linguistics (whether of the CAL² or public-meaning variety) on the question of whether legal interpretation can and should be made empirical rather than introspective. As suggested above, many doctrinalists favor the humanistic variety of interpretivism, reading legal texts for subjective inspiration and insight, rather than for objective evidence. Few such scholars would embrace the proposition that corpus linguistics can render normative and philosophical matters empirical simply by toting up the number of utterances on each side. Indeed, some might argue, who needs doctrinal scholarship if majority usages are never wrong?

easily achievable. In this vision, corpus linguistics represents what business- strategy researchers would call a “competence-enhancing technological discontinuit[y]”⁴⁴: different and valuable, yes; but fundamentally disruptive, no.

At other times, however, Hamann and Vogel seem to envision CAL² as something more radical. Particularly when they link CAL² to the larger agenda of “evidence-based jurisprudence,”⁴⁵ they conjure an image of law scholarship that would shift legal academia away from its current affinity with moral philosophy, welfare economics, and literary interpretation, toward a more “praxeological”⁴⁶ concern with bottom-line lawyering. Hamann and Vogel posit that “[l]awyers like doctors are interested in reality (only) insofar as their practical decision-making requires.”⁴⁷ But the practical decision-making of lawyers differs from that of doctors in a fundamental way: Doctors’ decisions confront an extrinsic physical reality that must be understood empirically before it can be mastered clinically; lawyers’ decisions, in contrast, confront only the intrinsic socio-cultural reality of legal institutions staffed by fellow legal professionals. Doctors cure diseases; lawyers persuade colleagues. Given that legal persuasion is heavily verbal, corpus linguistics could, in fact, prove to be a useful legal tool; but in this more pragmatic vision, it would serve not the scholarly agendas of truth and justice but the lawyerly agendas of advocacy and effectuation. By determining empirically which word combinations elicit which legal outcomes, practitioners could argue more compellingly (or more performatively) toward any given ends whether or not those ends were morally just, socially beneficent, or logically coherent.⁴⁸ Using the label “Linguistic Legal Realism” and

44. Michael L. Tushman & Philip Anderson, *Technological Discontinuities and Organizational Environments*, 31 ADMIN. SCI. Q. 439, 450 (1986).

45. Hamann & Vogel, *supra* note 1, at 1495.

46. *Id.* at 1484.

47. *Id.* (quoting Hanjo Hamann, EVIDENZBASIERTE JURISPRUDENZ: METHODEN EMPIRISCHER FORSCHUNG UND IHR ERKENNTNISWERT FÜR DAS RECHT AM BEISPIEL DES GESELLSCHAFTSRECHTS 7–8 (Mohr Siebeck, Tübingen 2014) (Ger.)).

48. This, I take it, is what Hamann and Vogel mean by “a new corpus driven macro perspective on the constitution of dogmatics—language, knowledge and power.” *Id.* at 1495. Significantly, they contrast this dogmatic agenda with earlier efforts to develop “electronic brains” to generate “more objective, more predictable, more just and less complex legal decision-making.” *Id.* at 1494. In other words, whereas the earlier agenda sought to use legal corpora to improve the overall health of the system (as assessed by extrinsic criteria, such as predictability

citing Holmes's *Path of the Law*,⁴⁹ Hamann and Vogel implicitly acknowledge that this is a legal linguistics for the bad man: it cares less about what legal words mean than about what they do.⁵⁰ Used in this way, corpus linguistics would quite possibly represent a “competence-destroying”⁵¹ discontinuity within the legal academy, elevating clinical practice to new prominence while pushing doctrinal interpretation, economic efficiency analysis, and moral philosophy to the margins.

V. CONCLUSION

In all honesty, I must confess that neither of these agendas holds much appeal to me as an empirical sociologist. Nonetheless, legal corpus linguistics holds a great deal of appeal for separate reasons of my own.

I am skeptical about whether corpus linguistics can reveal the true meaning of legal language because, in accord with the Legal Realists, I am skeptical that language can ever have a single, stable, true meaning, independent of particular speakers, particular audiences, and the particular purposes of each. In a few very special cases, the speaker, the audience, and the purpose may be so explicit and narrow that something like a “plain meaning” becomes apparent; but most legal words are uttered by multiple speakers to multiple audiences in multiple contexts. As a result, multiple usages often enjoy substantial currency alongside one another, and several of the most popular usages are likely to have been familiar to both the speaker and the audience(s) at the moment when the utterance was made. Moreover, some of the most important utterances—constitutional compromises, landmark legislation, breakthrough contracts—are intentionally ambiguous, reflecting agreement on words but not on substance. In

and justice), the new agenda focuses on the more pragmatic project of determining how practicing lawyers “succeed at convincing or out-writing opponents.” *Id.* at 1496.

49. *Id.* at 1494.

50. This characterization may be overly harsh. If one believes that skillful legal representation generally leads to more just, predictable, or economically efficient outcomes, then the use of linguistic evidence to improve lawyers' skills may lead not only to direct benefits for clients (who will, at least occasionally, be “bad men”), but also to indirect benefits for society as a whole. The distinctive feature of the praxeological use of legal linguistics, however, is that it gives top priority to improving lawyers' rhetorical skills, while potentially de-emphasizing such other skills as logical deduction, cost-benefit optimization, and ethical judgement.

51. Tushman & Anderson, *supra* note 44.

effect, each party implicitly gambles that the chosen verbal formula will eventually be interpreted favorably to that party's interests by some unknown third-party audience at some unknown future date.⁵² Corpus linguistics may still be useful for determining "plain meaning" in such cases—but only after a prior determination (by ideology or by power) that certain speakers, audiences, contexts, and usages are relevant, while others are not. As a sociologist, that prior determination strikes me as the socio-legally interesting moment, after which the findings of the corpus analysis will be largely predetermined.

I am equally skeptical of Hamann and Vogel's more radical "praxeological" agenda, albeit on quite different grounds. First, I am skeptical that the causal efficacy of legal words can be gleaned from a research agenda that confines itself to Cell 1 in Table 1's typology of legal and extra-legal processes (Law → Law) and Cell 4 in Table 2's typology of linguistic and nonlinguistic processes (Text → Text). Although I do not deny the causal *potential* of legal words, I doubt that the actual *impact* of legal words can be accurately gauged without controlling for confounding variables in the nontextual aspects of law and the nonlegal aspects of social life. Second, I think that Hamann and Vogel underestimate the reflexivity of legal practice. Legal words are not a chemical vaccine operating on insentient microbes⁵³; they are communicative devices carrying both ostensible meanings and subtextual signals. Even if corpus linguistics succeeds in identifying distinctively efficacious utterances in prior legal speech, there is no guarantee that those utterances will remain efficacious in the same way and to the same degree once their distinctive efficacy becomes widely known and disingenuously exploited. Teaching shepherds to cry "wolf" may seem like a useful lesson, but only until a corpus linguist comes along to reveal how efficacious that utterance is even in the absence of actual wolves. Third, as a social scientist with no professional obligation to train legal practitioners, I am not overly excited about the project of teaching lawyers how to select the most efficacious power words, especially if this comes at the expense of teaching them how to recognize and promote just and orderly social

52. See, e.g., JOSEPH R. GRODIN, IN PURSUIT OF JUSTICE: REFLECTIONS OF A STATE SUPREME COURT JUSTICE 133–61 (1989); Vicki Eaton Baier, James G. March & Harald Saetren, *Implementation and Ambiguity*, 2 SCANDINAVIAN J. MGMT. STUD. 197 (1986).

53. Cf. ROBERT L. KIDDER, CONNECTING LAW AND SOCIETY: AN INTRODUCTION TO RESEARCH AND THEORY 112–43 (1983) (critiquing a "vaccine model" of legal impact).

arrangements—or at least how to recognize the scope and limits of power words in the face of other less loquacious social forces. In short, although I can imagine standing apart from this praxeological enterprise and analyzing its progress as a study in the sociology of knowledge, I cannot imagine embracing its agenda as my own.

Instead, my enthusiasm for legal corpus linguistics rests on very different disciplinary grounds. To me, legal documents and the corpora composed of them are intriguing as *social artifacts*—the intentional products of particular social communities at particular times.⁵⁴ As such, they provide a cultural window into the past, and their variations across time and place provide a fossil record of socio-legal change. Admittedly, words-as-artifacts are often harder to interpret than either objects-as-artifacts or words-as-text: Beneath their communicative plain meaning, documentary artifacts may have both performative technical properties and symbolic subtextual resonances. And the same document may serve different purposes in different communities: having one role for the private individuals who crafted it; other roles for the various publics who stand before it as subjects or view it from afar as spectators; and yet other roles for the specialized professionals who bear responsibility for its implementation, curation, modification, and diffusion. Nonetheless, legal corpora provide unparalleled opportunities for big-data, longue-durée⁵⁵ cultural analysis. Cross-sectionally, we can compare them to other contemporaneous corpora from literature, politics, science, or other domains; longitudinally, we can trace their evolution as the legal community expands, matures, and professionalizes. They may never tell us the “true” meaning of the Law, and they may never generate a comprehensive how-to manual for efficacious legal speech; but they may certainly shed empirical light on our social world and on the political, cultural, psychological, and economic forces that make it run.

In the end, however, the empirical value of legal corpora may depend less on methodological ambition than on methodological humility. Legal corpora seem destined to become a useful and versatile

54. Mark C. Suchman, *The Contract as Social Artifact*, 37 L. & SOC'Y REV. 91 (2003).

55. Fernand Braudel, *Histoire et Sciences Sociales: La Longue Durée*, 13 ANNALES. ÉCONOMIES, SOCIÉTÉS, CIVILISATIONS 725 (1958) (Fr.) (arguing that historians should analyze long-term shifts and continuities in social structure, rather than merely chronicle salient but fleeting historical events).

tool in the social-scientific toolkit; but in themselves, they can never be a panacea. To understand society empirically through legal linguistics, we must pursue linguistic empiricism in all its registers—descriptive, configurational, interpretive, and causal. We must also couple empirical knowledge about legal linguistics with empirical knowledge about the many other, nonlinguistic aspects of law and of legal institutions. And, perhaps most importantly, we must couple empirical knowledge about law in all its aspects with commensurately deep and subtle empirical knowledge about the other facets of social life. For socio-legal empiricists, “law and society” must always be a conjunction, not a mere intersection. Nothing in legal corpus linguistics changes this fundamental truth.

