

Evidence-Based Jurisprudence Meets Legal Linguistics—Unlikely Blends Made in Germany

Hanjo Hamann[†] & Friedemann Vogel[‡] *

German legal thinking is renowned for its hair-splittingly sophisticated dogmatism. Yet, some of its other contributions to research are frequently overlooked, both at home and abroad. Two such secondary streams recently coalesced into a new corpus-based research approach to legal practice: Empirical legal research (which had already developed in Germany by 1913) and research on language and law (following German pragmatist philosopher Ludwig Wittgenstein’s work of 1953). This Article introduces both research traditions in their current German incarnations (Evidence-Based Jurisprudence and Legal Linguistics) and shows how three common features—their pragmatist observation of social practices, their interest in dissecting legal authority, and their big data strategy—inspired a new, corpus-based research agenda, Computer Assisted Legal Linguistics (CAL²).

CONTENTS

I. INTRODUCTION	1474
II. EVIDENCE-BASED JURISPRUDENCE.....	1476
A. Origins: Genuinely “Legal” Empirical Studies	1476
B. Evidence-Based Approaches to Law.....	1482
C. Evidence-Based Jurisprudence.....	1484

† Senior Research Fellow at the Max Planck Institute for Research on Collective Goods, Bonn (Germany), and Young Fellow of the Bielefeld Center for Interdisciplinary Research (Junges ZiF).

‡ Professor for Media Linguistics at the University of Freiburg (Germany).

* Both authors jointly head the International Research Group “Computer Assisted Legal Linguistics” (CAL²) at the Heidelberg Academy of Sciences (HAW) in Germany, whose funding is gratefully acknowledged. We thank Isabelle Gauer and Yinchun Bai for their valuable input and helpful suggestions on an earlier draft of the paper, and Mark Suchman for his insightful discussion, which even our revised draft could incorporate only in part. Any non-English sources discussed are written in German. Where we cite sources published in German, quotations were translated by the first author.

III. LEGAL LINGUISTICS	1487
A. Origins: Analyzing the Medium of Social Regulation	1487
B. Modern Legal Linguistics	1489
C. Legal Linguistics in the Age of New Media and Digitalization.....	1493
IV. A MEETING OF THE MINDS: COMPUTER ASSISTED LEGAL LINGUISTICS.....	1495
A. Conceptual Novelties of CAL ²	1496
B. State of CAL ² and its Challenges	1498
C. Toward a Reference Corpus of Legal Language.....	1499
V. CONCLUSION	1500

I. INTRODUCTION

When one can no longer hide behind empirical uncertainties, one is compelled to state one's normative preferences. . . . Ergo, and somewhat ironically, the more positive knowledge there is, the more and better the law's normative work will be.

—Michael Saks¹

Corpus research in law is inherently interdisciplinary. It differs from traditional legal research in two important features: methodologically, it relies on big data empiricism rather than doctrinal analysis; substantially, it treats language not merely as the medium but as the object of study. Both of these features have been central to intellectual traditions originating some hundred years ago—which ought to be acknowledged, analyzed, and adapted for the field as a whole to progress. This is where some work remains to be done, and this Article attempts to do it.

Nowadays, empirical legal research in the United States exerts notable influence overseas, e.g., in Germany.² According to its most

1. Personal letters quoted in Peter H. Schuck, *Why Don't Law Professors Do More Empirical Research?*, 39 J. LEGAL EDUC. 323, 335 (1989).

2. See the dedication in HANJO HAMANN, EVIDENZBASIERTE JURISPRUDENZ: METHODEN EMPIRISCHER FORSCHUNG UND IHR ERKENNTNISWERT FÜR DAS RECHT AM

popular origins narrative, this “new empiricism” traces its roots “to the original Legal Realist movement of the 1930s”.³ Yet, even that may not be looking back far enough. In Germany, a popular type of legal empirical inquiry called legal fact research (*Rechtstatsachenforschung*) was in full swing by the early 1910s. Although this methodology did not go unnoticed by contemporary U.S. scholarship, it was never fully appreciated.

Regarding research on language and law, no origins narrative seems yet to have been proposed in the United States.⁴ In Germany, where research on language and law arguably got its earliest academic coverage, the most important treatise on language pragmatics (which seeded German research on language and law) dates from the early 1950s. This contribution, too, is commonly overlooked both in international (U.S.-centered) discourse and in German academia proper.

Revisiting these early origins will enable us to trace their trajectory into the current German research programs called Evidence-Based Jurisprudence (Part II) and Legal Linguistics (Part III), respectively. This will help us understand the intellectual underpinnings of corpus research in Germany and may inform the U.S. discourse as well. As a future perspective, we will also show how both traditions recently coalesced into a new transdiscipline that combines big data corpus research with its own epistemological focus on the language of legal practice, distinct from the plain meaning and original intent traditions: Computer Assisted Legal Linguistics (CAL²) (Part IV).

BEISPIEL DES GESELLSCHAFTSRECHTS p. V (Mohr Siebeck, Tübingen 2014) [hereinafter EVIDENZBASIERTE JURISPRUDENZ] (“to the memory of Theodore Eisenberg,” who had founded the Empirical Legal Studies [ELS] movement and passed away in the year of the book’s publication).

3. Mark C. Suchman & Elizabeth Mertz, *Toward a New Legal Empiricism: Empirical Legal Studies and New Legal Realism*, 6 ANN. REV. LAW & SOC. SCI. 555, 556–57 (2010).

4. However, Peter M. Tiersma came close in *A History of the Languages of Law*, in THE OXFORD HANDBOOK OF LANGUAGE AND LAW 13 (Peter M. Tiersma & Lawrence M. Solan eds., 2012). See also Stephen C. Mouritsen, *Corpus Linguistics in Legal Interpretation—An Evolving Interpretive Framework*, 6 INT. J. LANG. & LAW 67 (2017).

II. EVIDENCE-BASED JURISPRUDENCE

Evidence-Based Jurisprudence is a moniker for one of the most recent empirical research programs in Germany. While the terms “evidence-basing” and “law” were used in conjunction only rather recently (section B), their intellectual history stretches back into the early twentieth century (section A). Nowadays, this research develops an explicitly pragmatic account of legal empiricism that raises epistemological questions as well (section C).

A. Origins: Genuinely “Legal” Empirical Studies

One of the most influential forerunners of empirical work in U.S. jurisprudence is surely law and economics. As an over-quoted quip has it, “the man of the future is the man of statistics and the master of economics.”⁵ Judging by this quote and its popularity, it would seem that even from the very start, statistical methods of (quantitative) empirical research were tied up inextricably with the theory of economic analysis.

Yet in German-speaking countries, while economic analysis of law started even earlier than in the United States,⁶ the German variety of economic analysis never monopolized statistics (“econometrics”) to the extent it did in the United States. Instead, another tradition came to be regarded as *the* touchstone of empirical inquiry, legal fact research (*Rechtstatsachenforschung*),⁷ which both fed on and further fueled “a thorough change” of German legal methods at the onset of the twentieth century:

5. Oliver W. Holmes Jr., *The Path of the Law*, 10 HARV. L. REV. 457, 469 (1897).

6. See Izhak Englard, *Victor Mataja’s Liability for Damages from an Economic Viewpoint: A Centennial to an Ignored Economic Analysis of Tort*, 10 INT’L REV. L. & ECON. 173 (1990) (discussing Mataja’s 1888 book, *Das Recht des Schadensersatzes vom Standpunkte der Nationalökonomie*); Kristoffel Grechenig & Martin Gelter, *The Transatlantic Divergence in Legal Thought: American Law and Economics vs. German Doctrinalism*, 31 HASTINGS INT’L & COMP. L. REV. 295, 325–40 (2008) (citing an even earlier law and economics monograph by Friedrich Kleinwächter [1883], as well as work following Mataja’s by lawyer-turned-economics-professor Anton Menger [1890], and the German inventor of “principal-agent” and “nexus of contracts” reasoning, Erwin Steinitzer [1908]).

7. Arthur Nussbaum, *Fact Research in Law*, 40 COLUM. L. REV. 189, 207 (1940) [hereinafter *Fact Research in Law*] (Nussbaum’s terminology notwithstanding, we translate *Rechtstatsachenforschung* as “Legal Fact Research” for purposes of readability. Both terms are synonymous.).

Observation of the law from within has been integrated by observation from without Thus, two sets of facts are to be taken into account: the life situations to be regulated, and the politics and evaluations guiding the regulating activities. To know them both, the jurist must put himself outside of the body of the existing rules. He must closely observe the facts of life, economics and other social phenomena.⁸

One of the earliest pioneers of this new “observation from without” methodology before World War I was practicing lawyer, and soon to turn law professor, Arthur Nussbaum. Nussbaum advocated empirical research as a remedy to the “empty dogmatism” of German legal academia, of which he claimed his contemporaries had grown “weary.”⁹ He called for the “systematical scientific processing and comprehensive acquisition” of “a certain body of inductively discoverable facts . . . which have to be known to fully appreciate and properly apply legal norms”—in short, “legal facts” (*Rechtstatsachen*).¹⁰ While this approach did not fully anticipate the theoretical premises of the later “legal realism” movement (which denied the existence of applicable legal norms in the first place), it did anticipate much of legal realism’s methodology—as Nussbaum realized when he defined legal fact research as

the systematic search into the social, political and other fact conditions which give rise to the individual legal rules, and examination of the social, political and other effects of those rules. . . . No contest or doubt exists as to the necessity and urgency of such research among the various groups of realists, American or non-American. It is a first principle of realism of all shades.¹¹

8. Max Rheinstein, *Comparative Law and Conflict of Laws in Germany*, 2 U. CHI. L. REV. 232, 251–53 (1935).

9. ARTHUR NUSSBAUM, DIE RECHTSTATSACHENFORSCHUNG. IHRE BEDEUTUNG FÜR WISSENSCHAFT UND UNTERRICHT 1 (J.C.B. Mohr [Paul Siebeck], Tübingen 1914) [hereinafter DIE RECHTSTATSACHENFORSCHUNG] (“man ist der leeren Dogmatik überdrüssig geworden”). Other excerpts are translated to English in Elliott E. Cheatham et al., *Arthur Nussbaum: A Tribute*, 57 COLUM. L. REV. 1, 3 (1957) (explaining just “how revolutionary Nussbaum’s approach . . . was”).

10. DIE RECHTSTATSACHENFORSCHUNG, *supra* note 9, at 3, 6, 8.

11. *Fact Research in Law*, *supra* note 7, at 197 (also calling legal fact research “definitely a phase of the realistic movement”); Cheatham et al., *supra* note 9, at 1 (noting

With legal fact research slightly preceding the other legal realist movements, it spurred plenty of follow-up research in Germany¹² and equally infused the U.S. literature of its day.¹³ Even though its dissemination was tragically cut short by “the immediately following disasters, war, revolution, inflation, deflation, social trouble and again revolution,”¹⁴ it was later revisited and persists in Germany to this day.¹⁵

Equally persistent is the original disdain that legal fact research harbored for sophisticated statistics¹⁶ and for causal inquiry in general.¹⁷ Nussbaum envisioned a genuinely legal empirical research

that Nussbaum was “in the forefront of those who preached and practiced a new approach to the study of law: . . . the search for a more realistic approach . . . in which economic and social facts formed an integral part of legal doctrine and principle”). For other German “shades of realism,” see Rheinstein, *supra* note 8, at 251 (citing examples as early as 1903 for “[d]iscussions similar to those now going on here in the United States between traditionalists, legal sociologists and realists”).

12. See Cheatham et al., *supra* note 9, at 4 (“Among them were studies on cartels, sales and leases, international commercial transactions, standard contracts, which have become classics.”); *Fact Research in Law*, *supra* note 7, at 198; Robert A. Riegert, *Empirical Research About Law: The German Picture, with Comparisons and Observations*, 2 DICK. INT’L L. ANN. 1 (1983), library.law.psu.edu/psilr/vol2/iss1/2.

13. Prior to Nussbaum’s own reflections, see *supra* note 7, other authors who had taken note included Henning Holm-Nielson, *The Law of Torts in Denmark*, 15 J. COMP. LEGIS. & INT’L L. 3D SER. 176 (1933); Henning Holm-Nielson, *The Problem of Wage Earner Bankruptcies and its English Solution*, 9 J. NAT’L ASS’N REF. BANKR. 103 (1935); Sidney B. Jacoby, *Some Realism About Judicial Statistics*, 25 VA. L. REV. 528, 529 (1939); Edouard Lambert & Max J. Wasserman, *The Case Method in Canada and the Possibilities of its Adaptation to the Civil Law*, 39 YALE L.J. 1, 18 (1929); Rheinstein, *supra* note 8, at 251.

14. Holm-Nielson, *Wage Earner Bankruptcies*, *supra* note 13, at 106; see Cheatham et al., *supra* note 9, at 4 (“[Nussbaum’s] venture was cut short by the advent of the Nazi regime.”).

15. See the recent survey by HAMANN, *supra* note 2, at 40–45, the monograph series by Duncker & Humblot publishers, SCHRIFTENREIHE ZUR RECHTSSOZIOLOGIE UND RECHTSTATSACHENFORSCHUNG (2014) (ninety-six total volumes), and the series by JWV publishing house, EMPIRISCHE STUDIEN ZUM DEUTSCHEN UND EUROPÄISCHEN UNTERNEHMENSRECHT (2017) (fourteen volumes).

16. Cheatham et al., *supra* note 9, at 2 (“Professor Nussbaum, unlike some of the American realists, has never considered statistics as of more than auxiliary importance.”); *Fact Research in Law*, *supra* note 7, at 208–19 (identifying statistics as the “main” and “paramount” problem and attacking the “popular superstition” and “indiscriminate enthusiasm for statistics in legal science” resulting in “a rampant ‘rage du nombre’”).

17. DIE RECHTSTATSACHENFORSCHUNG, *supra* note 9, at 5–6 (insisting that “methods of natural scientific inquiry, indeed of any causal inquiry, cannot be methods of jurisprudence”).

tradition, distinct and independent from economics.¹⁸ It was meant to robustly inform legal decision-making, not to answer basic social science questions. So while legal fact research frankly acknowledged its intellectual heritage,¹⁹ it meant a genuinely legal heritage—“[a]t all times, legal writings have referred to economic and other social factors. . . . [T]heir references to factual elements were incidental in nature The call for fact research *per se* and in all fields of the law is of recent date.”²⁰

This reference to (and reverence for) “legal” writings rather than those of theory-testing social sciences emphasized a notably different approach from concurrent U.S. legal realism.²¹ Using a distinction suggested by Mark Suchman in discussing this Article,²² one could say that if “the law” is a network of words and actions affecting each other (with four basic causal relationships imaginable in a 2x2 typology), then legal fact research is concerned less with the effect of actions upon actions (when compared with U.S. legal realism) or with the effect of words upon words (when compared with academic linguistics), but focuses more on the “off-diagonal” cases in which words and actions interact, making causal claims notoriously difficult, if not futile altogether.

This focus may have resulted in Nussbaum’s more pragmatic and less epistemological approach to empiricism from the very start, and it can even be traced in current schools of German empirical legal research, as we will argue later. Hence “legal fact research” became something like a proprietary label for a certain type of non-doctrinal-but-legal inquiry that was systematic rather than incidental—but

18. *Id.* at 6–8 (“There can be no talk of transferring subject matter of economics into legal research The approach in these two subject areas is so different in essence . . . that large areas of economics must remain out of consideration for actual legal work.”).

19. *Fact Research in Law*, *supra* note 7, at 197 (“There is nothing brand new about fact research in law.”).

20. *Id.*

21. See Rheinstein, *supra* note 8, at 252 (“German legal scholars have . . . never confined themselves to an art of prophesying what judges are likely to do.”); *id.* at 252 n.87 (“The continental legal scholar reflects in advance on new problems which were never before decided by the Codes or the courts. He leads the practice. The Anglo-American text-book writer and teacher of law is inclined to confine himself to expounding what the courts did in the past.”).

22. See Mark C. Suchman, *The Power of Words: A Comment on Hamann and Vogel*, 2017 BYU L. REV. 1751.

emphatically distinct from “empirical” research. This was true both in Germany and in the United States, where “the value of legal fact research, which has so successfully been undertaken in this country in recent years, can be fully appreciated without subscribing to the realist doctrines as such and without stressing the empirical character of legal science to the exclusion of the other components of law.”²³

This specific understanding of legal fact research later blurred amid the 1970s flood of sociological inquiries into law, known as *Rechtssoziologie* in Germany or as “law and society,” “legal sociology,” or “sociolegal studies” in English-speaking jurisdictions. Legal sociologists in Germany broadened *Rechtstatsachenforschung* to include any of their non-theoretical endeavors.²⁴ By the mid-1980s, it had lost most of its discriminatory power, being variably employed as a synonym for empirical legal research as a whole, or as a generic term for different empirical approaches, or as a label for one particular approach.²⁵ Nowadays, legal fact research commonly denotes “observational studies with descriptive intentions and near-exclusive reliance on descriptive statistics.”²⁶

As “genuinely legal” empirical research struggled with its identity in Germany, it received new momentum on the other side of the Atlantic: a “new legal empiricism” developed from 1996 to 1997 onward, with Empirical Legal Studies (ELS) and New Legal Realism (NLR) trying “to legitimate empirical research within the legal academy itself”²⁷ by “reinserting empirical social science into the legal academy—not necessarily in the original idiom of the social

23. Jacoby, *supra* note 13, at 529; *see also* *Fact Research in Law*, *supra* note 7, at 197; Manfred Rehinder, *The Development and Present State of Fact Research in Law in the United States*, 24 J. LEGAL EDUC. 567, 567 (1972) (noting that “[n]on-doctrinal research includes empirical research, fact-research, behavior- and fact-inquiry”).

24. *See* Alfons Bora, *Sociology of Law in Germany: Reflection and Practice*, 43 J.L. & SOC'Y 619, 626 (2016) (discussing the “schism in thematic orientation between empirical legal research (*Rechtstatsachenforschung*) and sociological theory of law, which has become a powerful part of the official narrative of the field” (footnote omitted)).

25. *See* Bora, *supra* note 24, at 638 (noting the “semantic and conceptual fuzziness” of sociological empirical research in law); Riegert, *supra* note 12, at 7 (“It has been described as ‘non-doctrinal research’; ‘fact (or social-fact) research in law’; ‘social-science research in law’; ‘research about law’; ‘experimental, empirical, field, or quantitative research’; ‘law-related research’; ‘research in law and society’; ‘research in legal sociology,’ and ‘interdisciplinary legal research.’”).

26. EVIDENZBASIERTE JURISPRUDENZ, *supra* note 2, at 41 n.258.

27. Suchman & Mertz, *supra* note 3, at 556.

sciences, but in a dialect that the legal academy might actually be willing to hear.”²⁸ Behind this general mission united a vast and varied following, so twenty years later the literature on empirical studies in law and their epistemological challenges is boundless.²⁹ Apart from various methodological concerns, one issue seems to be particularly dear to the hearts of those who reflect on the new empiricism:

The issue of translation between law and social science is a core issue Our goal is to create translations of social science that will be useful even to legal academics and lawyers who do not wish to perform empirical research themselves, while also encouraging translations of legal issues that will help social scientists gain a more sophisticated understanding of how law is understood “from the inside” by those with legal training.³⁰

This new empiricism fully hit Continental European shores in 2016 with the first Conference on Empirical Legal Studies in Europe (CELSE) conference held in Amsterdam.³¹ Most recently, the German-based *Journal of Institutional and Theoretical Economics* (JITE) hosted a conference on “Empirical Methods for the Law” (Syracuse, 7–9 June 2017) where eminent empirical legal scholars from the United States and Germany took a step back from the

28. *Id.* at 565.

29. For textbooks and handbooks, see EMPIRICAL LEGAL ANALYSIS: ASSESSING THE PERFORMANCE OF LEGAL INSTITUTIONS (Yun-chien Chang ed., 2014); ROBERT M. LAWLESS ET AL., EMPIRICAL METHODS IN LAW (2010); THE OXFORD HANDBOOK OF EMPIRICAL LEGAL RESEARCH (Peter Cane & Herbert M. Kritzer eds., 2010). For reviews, see Michael Adler & Jonathan Simon, *Stepwise Progression: The Past, Present, and Possible Future of Empirical Research on Law in the United States and the United Kingdom*, 41 J. LAW & SOC’Y 173 (2014); Suchman & Mertz, *supra* note 3; Kathryn Zeiler, *The Future of Empirical Legal Scholarship: Where Might We Go from Here?*, 66 J. LEGAL EDUC. 78 (2016).

30. Howard Erlanger et al., *Is It Time For a New Legal Realism?*, 2005 WIS. L. REV. 335, 336; Suchman & Mertz, *supra* note 3, at 559 (stating that with a similar objective, albeit less epistemological reserve, “an ELS framing would reach out to the legal academy’s traditional constituencies, highlighting a study’s usefulness to practicing attorneys, relevance to judges and legislators, or bearing on doctrinal conundrums—and soft-pedaling any disciplinary debates that legal audiences might consider overly arcane”).

31. See *First Conference on Empirical Legal Studies in Europe (CELSE)*, UNIV. AMSTERDAM, <http://celse2016.acl.e.nl/> (last visited Jan. 19, 2018). The next CELSE will take place in Leuven Belgium, in May 2018. See *Conference on Empirical Legal Studies in Europe 2018*, KU LEUVEN, www.law.kuleuven.be/pub/en/cltej/CELSE (last visited Jan. 19, 2018).

statistical sophistication into which empirical legal research has matured, and returned to the question that had been central to Nussbaum even one hundred years ago: “Could there be empirical methods that are specific for legal research?”³² In Germany, such renewed interest met with a new paradigm to think about legal meta-empiricism, Evidence-Based Jurisprudence.

B. Evidence-Based Approaches to Law

Considering its name, evidence-based jurisprudence might be mistaken for one of various other research strands that used similar vocabulary. To clearly distinguish these strands, it is helpful to revisit the etymology of the term “evidence-basing” and its history in medical research.

In 1992, a group of Canadian doctors noticed a massive disconnect between the fast-growing clinical and epidemiological research evidence on the one hand and patient treatment on the other, which was still largely based on doctors’ rule-of-thumb experience and intuitive wisdoms. As a remedy, these doctors proposed translation procedures collectively called “evidence-based medicine” (EBM),³³ a term that was soon applied to legal practice concerning medical issues as well.³⁴ But the moniker’s journey did not end there. As it travelled on, its distinctiveness waned, just like “legal fact research” had done a couple of decades earlier. By 2005, an “evidence-based law approach” had appeared in the legal literature as an “inspiration” from medicine to promote “a systematization and verification of knowledge about legal doctrine.”³⁵ In subsequent years, authors from Canada, the Netherlands, and the United States coined similar composites of

32. Invitation Email by Conference Convener Christoph Engel (Dec. 2016) (on file with author). Conference contributions are forthcoming in 174 J. INSTITUTIONAL & THEORETICAL ECON. (forthcoming 2018); See also a German report by Hanjo Hamann, *Empirische Methoden für die Rechtswissenschaft. 35. Jahrestagung für juristische Institutionenökonomik vom 7. bis 10. Juni 2017, in Siracusa (Sizilien)*, 72 JURISTENZEITUNG [JZ] (forthcoming 2018).

33. Evidence-Based Med. Working Grp., *Evidence-based Medicine: A New Approach to Teaching the Practice of Medicine*, 268 J. AM. MED. ASS’N 2420 (1992).

34. Editorial, *Evidence-Based Morality*, 161 CAN. MED. ASS’N J. 229 (1999).

35. Bernard Trujillo, *Patterns in a Complex System: An Empirical Study of Valuation in Business Bankruptcy Cases*, 53 UCLA L. REV. 357, 363 n.17 (2005).

“evidence-basing” and “law,” seemingly independent of each other and without mutual citation or acknowledgement.³⁶

This disjointed development resulted in an equally disjointed usage of its central concept. Far from denoting a common agenda, the “evidence-based” label was used by lawyers for at least three distinct concepts:

- (1) creating and applying legal rules as informed by legal fact research³⁷ or empirical legal scholarship,³⁸
- (2) studying quantitatively how lawyers create and apply legal rules by “treating doctrine as a quantitative unit,”³⁹ and
- (3) having law-makers embrace “legal experimentalism,”⁴⁰ i.e., “that laws and policy initiatives are to be supported by research evidence and that policies are preferably introduced on a trial and error basis.”⁴¹

This proliferation of very different ideas related to the goal of putting more empirical systematicity and scientific rigor into law hampered the usefulness of the “evidence-based” label as applied to law. As Mark Suchman observed in discussing this Article, “[P]articularly in thinking about things like ‘empirical’ . . . the potential for misunderstandings over words having multiple meanings in different settings is fairly large.”⁴² Hence one of the

36. Rob van Gestel, *Evidence-Based Lawmaking and the Quality of Legislation Regulatory Impact Assessments in the European Union and the Netherlands*, in STATE MODERNIZATION IN EUROPE 139 (Heinz Schäffer & Julia Iliopoulos-Strangas eds., 2007); R.J. Cook et al., *Emergency Contraception, Abortion and Evidence-Based Law*, 93 INT’L J. GYNECOLOGY & OBSTETRICS 191 (2006); Kevin E. Davis, *Legal Universalism: Persistent Objections*, 60 U. TORONTO L.J. 537, 548 (2010); Jeffrey J. Rachlinski, *Evidence-Based Law*, 96 CORNELL L. REV. 901, 919 (2011).

37. Cook et al., *supra* note 36, at 192–93.

38. Rachlinski, *supra* note 36, at 904–05; *see also* Richard A. Posner, *The State of Legal Scholarship Today: A Comment on Schlag*, 97 GEO. L.J. 845, 852 (2009) (even earlier, explicitly calling ELS an “analogy” to evidence-based medicine “in placing law on a solid empirical basis”).

39. Trujillo, *supra* note 35, at 363.

40. Davis, *supra* note 36, at 548 (defining legal experimentalism as “a call for law-making to be based on evidence showing which legal instruments work and which ones don’t”).

41. Van Gestel, *supra* note 36, at 142.

42. Mark Suchman, Address at Inaugural BYU Law and Corpus Linguistics Conference (Feb. 2016). *See generally* Suchman, *supra* note 22.

present authors recently undertook not only to review the usages of the term “evidence-basing” in law but also to analyze more precisely what medical practitioners actually meant by “evidence-basing” and whether their specific ideas might be useful for the law.⁴³

C. Evidence-Based Jurisprudence

Evidence-based jurisprudence in the previously introduced sense starts by asking for conceptual overlaps between legal and medical practice, hence realizing

that jurisprudence and medicine share a common epistemological perspective: ‘The subject is called Jurisprudence, not Jurisscience. The practice of law [. . .] is thus obviously closely related to the two other major practical disciplines, technology and medicine.’ Lawyers like doctors are interested in reality (only) insofar as their practical decision-making requires. Also, lawyers and doctors both deal with pathological cases. . . and cannot defer their decision until reality is fully understood.⁴⁴

Starting from these parallels and the pragmatism that both disciplines require in their quest for empirical data, evidence-based jurisprudence takes medicine’s concept of “evidence-basing” seriously in its own right, not merely watered down to a metaphor for empirical research, as was previously done. The novel contribution then is to analyze and potentially adapt the “tremendous substance and discipline” of “[e]vidence-based practice[s]” to enrich legal studies.⁴⁵ In line with this praxeological impetus, the first distinction transferred from evidence-based practice into jurisprudence was one between the “‘doing’ mode,” the “‘using’ mode” and the “‘replicating’ mode” of empirical

43. EVIDENZBASIERTE JURISPRUDENZ, *supra* note 2; *see also* Hanjo Hamann, *Empirische Erkenntnisse in juristischen Ausbildungsarbeiten. Prüfungsschema, Zitier- und Arbeitshilfen für das Jurastudium und danach [Empirical Findings in Legal Examinations. A Test Checklist, Citation and Writing Aids for Legal Studies and Afterwards]*, 39 JURISTISCHE AUSBILDUNG [JURA] 759 (2017) (summarizing evidence-based methods for educational purposes).

44. EVIDENZBASIERTE JURISPRUDENZ, *supra* note 2, at 7–8 (citing DAS PROPRIUM DER RECHTSWISSENSCHAFT, at ix (Christoph Engel & Wolfgang Schön eds. 2007)).

45. Denise M. Rousseau, *2005 Presidential Address: Is There Such a Thing as “Evidence-Based Management”?*, 31 ACAD. MGMT. REV. 256, 258 (2006).

research⁴⁶—a distinction later mirrored in the “doer, user, . . . critic” typology by Shari Diamond.⁴⁷

Each of these modes requires different tools. For instance, given that most lawyers—just as most doctors—will not be able to produce empirical research according to empirical state-of-the-art standards,⁴⁸ their most promising strategy will resemble evidence-based medicine’s five-step procedure to (1) ask for, (2) access, (3) appraise, (4) apply, and (5) assess empirical evidence gathered elsewhere.⁴⁹ Thus, evidence-based jurisprudence is not merely another strand of empirical inquiry but actually meta-empiricism in the “using” mode (*Rezeption*), which observes primary empirical research and distills overarching findings that can reliably inform judicial practice.⁵⁰

Despite its inspiration from medicine, evidence-based jurisprudence rests on the shoulders of empirilegal theorizing and overtly acknowledges this fact.⁵¹ Like earlier approaches,⁵² evidence-based jurisprudence seeks to provide a genuinely legal interface that does not threaten the identity of legal academia or “require[] law professors to exit the library . . . to conduct their scholarship.”⁵³ Like

46. Sharon E. Straus & Finlay A. McAlister, *Evidence-Based Medicine: a Commentary on Common Criticisms*, 163 CAN. MED. ASS’N J. 837, 837, 839 (2000) (noting that “[a]lthough a minority of practitioners of evidence-based medicine also do research, its practice is . . . not a method for performing research”).

47. Shari Seidman Diamond, *Empirical Marine Life in Legal Waters: Clams, Dolphins, and Plankton*, 2002 U. ILL. L. REV. 803, 807–08.

48. Suchman & Mertz, *supra* note 3, at 566 (“Legal academics must train future lawyers, advise sitting policy makers, and assimilate an ongoing stream of daily legal developments. Within this world, even dual-degree social scientists find that shortened time lines and limited external funding opportunities make the standards of ‘normal social science’ difficult to meet.”); cf. Zeiler, *supra* note 29 (for currently suggested remedies).

49. Chris Del Mar et al., *Teaching Evidence Based Medicine Should Be Integrated into Current Clinical Scenarios*, 329 BRIT. MED. J. 989 (2004); cf. William Rosenberg & Anna Donald, *Evidence Based Medicine: An Approach to Clinical Problem-Solving*, 310 BRIT. MED. J. 1122 (1995) (for an earlier four-step procedure).

50. Hanjo Hamann, *Unpacking the Board: A Comparative and Empirical Perspective on Groups in Corporate Decision-Making*, 11 BERKELEY BUS. L.J. 1 (2014) [hereinafter *Unpacking the Board*] (spelling out a sample application of evidence-based methods, based on a chapter from EVIDENZBASIERTE JURISPRUDENZ, *supra* note 2).

51. EVIDENZBASIERTE JURISPRUDENZ, *supra* note 2, at VII (“Research requires firm foothold on the shoulders of the proverbial Giant. . . . The present treatise is no exception.”).

52. See *supra* Section II.A.

53. Suchman & Mertz, *supra* note 3, at 566 (earlier approaches included “the legal process school, critical legal studies, feminist legal theory, critical race theory, law and economics”); EVIDENZBASIERTE JURISPRUDENZ, *supra* note 2, at 32 (noting that evidence-

the most recent brand of U.S. legal empiricism, which “turned to the pragmatist tradition in philosophy,”⁵⁴ evidence-based jurisprudence also emphasizes its pragmatic origins.⁵⁵

As far as meta-empiricism relies on gathering and appraising empirical evidence, evidence-based medicine introduced a rule-of-thumb hierarchy of five evidence “levels,”⁵⁶ preferring controlled randomized trials (i.e., large-sample field experiments) and meta-analyses (i.e., “the statistical analysis of statistical analyses”)⁵⁷ over other ways of gathering empirical data. Insofar as these are lacking, however, any rigorous and replicable empirical study is assumed to outperform (or at least usefully complement) the lowest evidence level “mechanism-based reasoning,” i.e., experiential judgment.⁵⁸

Building on and going beyond these learnings of evidence-based medicine, law can develop its own canon of pragmatic rules-of-thumb to guide meta-empirical inquiry. Seven such rules were previously proposed:⁵⁹

- (1) All empirical research is implicitly normative.
- (2) Careful study design trumps statistical sophistication.
- (3) Without hypotheses, there are no results.
- (4) Results are not in the significance, but in the effect size.
- (5) A picture tells more than a thousand significances.

based jurisprudence generally “enables access to empirical research from a specifically legal perspective, while letting legal researchers use their traditional mode of operation as far as possible”).

54. Suchman & Mertz, *supra* note 3, at 561.

55. EVIDENZBASIERTE JURISPRUDENZ, *supra* note 2, at 15, 18–19, 53–55.

56. Jeremy Howick et al., OCEBM Levels of Evidence Working Grp., *Oxford Centre for Evidence-Based Medicine 2011 Levels of Evidence*, CEBM.NET, <http://www.cebm.net/wp-content/uploads/2014/06/CEBM-Levels-of-Evidence-2.1.pdf> (last visited Jan. 20, 2018).

57. PAUL D. ELLIS, THE ESSENTIAL GUIDE TO EFFECT SIZES: STATISTICAL POWER, META-ANALYSIS, AND THE INTERPRETATION OF RESEARCH RESULTS 94 (2010); cited by *Unpacking the Board*, *supra* note 50, at 30 (explaining that “[t]he process of meta-analysis helps identify bias . . . [and] may also increase the statistical power to detect small effects and reduce the reliance on fickle significance levels by estimating standardized effect sizes. In short, it systematizes literature reviews and remedies a number of shortcomings of the research and publication process.”).

58. Howick et al., *supra* note 56.

59. EVIDENZBASIERTE JURISPRUDENZ, *supra* note 2, at 106–26.

- (6) Statistics must not be magic, but ought to be MAGIC.⁶⁰
- (7) Synthesis trumps primary research, meta-analyses trump narrative reviews.

Generalizing from these and other evidence-based practices, three features of the evidence-based jurisprudence approach stand out: it emphasizes the *political embeddedness of empirical issues* and the importance of theory (rules 1–3); it advocates a *pragmatic approach to empirical practices*, explicitly assessing their actual impact (rules 4–6); and it adopts an *observational big data strategy* by preferring cumulated research over salient yet eclectic findings (rule 7). These three features can also be identified in another empirical research tradition hailing from Germany, namely Legal Linguistics,⁶¹ which is why a collaboration between the two research traditions has proven fruitful and provides an intriguing future perspective.⁶²

III. LEGAL LINGUISTICS

Legal Linguistics is a theoretical and empirical research approach in the wider field of “law and language” research. It originated in Germany (Section B), where most of its historical roots can be found (Section A). In the wake of globalization and digitalization, it developed a distinctly big data (corpus) driven approach (Section C) that is beginning to coalesce with empirical legal research more generally.⁶³

A. Origins: Analyzing the Medium of Social Regulation

Law lives in language. We cannot speak about legal norms without acknowledging that they are constructed by and through speech and texts. In fact, reflections on the relationship between signs and societal norms are even older than modern constitutional states: Plato’s philosophical text *Phaedrus* (fourth century B.C.) features a dialogue about whether speech and written papyrus can

60. See ROBERT P. ABELSON, *STATISTICS AS PRINCIPLED ARGUMENT* (1995) (developing five “MAGIC” criteria for convincing statistical analyses: Magnitude, Articulation, Generality, Interestingness, and Credibility).

61. See *infra* Part III.

62. See *infra* Part IV.

63. See *infra* Part IV.

appropriately communicate “the Good” in general and effectively regulate the *polis* (city-state) in particular.⁶⁴ Plato despised scriptures because he considered them prone to misunderstandings; only speech and oral dialogues were deemed fit for the dialectics of truth. Ever since this Platonic account, the medium of law—i.e., language—was often criticized (in the sense of Kant) both from the perspective of experts (lawyers, politicians, academics, theologians) and that of laypeople.⁶⁵

The most important steps towards modern legal linguistics were taken in eighteenth and nineteenth century German feudal states. Friedrich Carl von Savigny (1779–1861), founder of the German Historical School of Jurisprudence, wrote that “both law and language live in the soul of people.”⁶⁶ He and others, like jurist Joseph von Sonnenfels (1732–1817) and linguist Johann Christoph Adelung (1732–1806), developed a new style of official language (*Geschäftsstil*) to improve the linguistic quality of statutes, administrative decisions, etc.⁶⁷ So-called “style guides” (*Stillehren*) greatly impacted the education of jurists and administrators.⁶⁸

One of Savigny’s disciples, famous lawyer-philologist Jacob Grimm (1785–1863), also pioneered the empirical lexicography of legal language.⁶⁹ One of Grimm’s most important monographs was the first grammar of legal language.⁷⁰ Another contained a collection (corpus) of “wisdoms” (*Weisthümer*) that documented oral traditions of legal communication in the past, serving as a historical basis for

64. See PLATO, PHAEDRUS (James H. Nichols Jr. trans., 1998).

65. For details, see Friedemann Vogel, *Rechtslinguistik: Bestimmung einer Fachrichtung*, HANDBUCH SPRACHE IM RECHT 209 (Ekkehard Felder & Friedemann Vogel eds., 2017). See also various contributions on *Legal Language* in THE OXFORD HANDBOOK OF LANGUAGE AND LAW, *supra* note 4.

66. FREDERICK CHARLES VON SAVIGNY, OF THE VOCATION OF OUR AGE FOR LEGISLATION AND JURISPRUDENCE 52 (Abraham Hayward trans., 1831).

67. See Bernhard Asmuth, *Geschäftsstil. Seine Prägung durch Sonnenfels und Adelung um 1784*, HISTORISCHE RECHTSSPRACHE DES DEUTSCHEN 175 (Andreas Deutsch ed., 2013).

68. Gernot Kocher, *Rechtsvereinheitlichung und Rechtssprache von Maria Theresia bis Franz Joseph I*, HISTORISCHE RECHTSSPRACHE DES DEUTSCHEN, *supra* note 67, at 207.

69. See Heino Speer, *Das Deutsche Rechtswörterbuch. Historische Lexikographie einer Fachsprache*, LEXICOGRAPHICA 85 (Antonín Kučera et al. eds., 1989).

70. JACOB GRIMM, DEUTSCHE RECHTSALTERTHÜMER (Wissenschaftliche Buchgesellschaft 1965) (1899); see Ruth Schmidt-Wiegand, *Deutsche Sprachgeschichte und Rechtsgeschichte bis zum Ende des Mittelalters*, SPRACHGESCHICHTE 73-4 (Werner Besch et al. eds., 2d ed. 1998).

further empirical studies.⁷¹ Such historical lexicography of legal language continues to this day, with its most important heirs being the “German Law Dictionary” (*Deutsches Rechtswörterbuch*, DRW) established in 1896/97,⁷² and the “Concise Dictionary of German Legal History” (*Handwörterbuch zur deutschen Rechtsgeschichte*, HRG) established in 1971.⁷³

B. Modern Legal Linguistics

After these early forays into empirical lexicography, modern linguistic approaches to law really took off starting in 1953 with philosopher Ludwig Wittgenstein’s famous treatise initiating the *pragmatic turn* in linguistics.⁷⁴ This fundamental change in the research perspective on language in general (and legal language in particular) led to law being considered, no longer as a system of isolated (“autopoietic”) structures, but as a social practice with and within texts, taking place in the context of complex institutions. This view emphasized social roles of specific actors, considered their interactions with texts, and analyzed the structure of these interactions, thereby describing legal interpretation methods, communication routines in courts, and the creation of statutes as specialized text genres.

The legal implications of this new view began to be fully realized (and elaborated) by lawyers and linguists in the 1960s. In Germany, for instance, one law professor made his career on a monograph titled “Normative structure and normativity” (196). In this treatise, Friedrich Müller laid the foundations for a theory that he later continuously refined and developed: “Structuring Legal Theory” (*Strukturierende Rechtslehre*).⁷⁵ While still claiming to be a “work in progress,” this theory has exerted demonstrable influence in legal

71. JACOB GRIMM, WEISTHÜMER (Göttingen, Dieterichschen Buchhandlung 1866).

72. See DAS DEUTSCHE RECHTSWÖRTERBUCH: PERSPEKTIVEN [DRW] (Andreas Deutsch ed. Heidelberg Univ., Winter 2010), www.rzuser.uni-heidelberg.de/~cd2/drw.

73. See HANDWÖRTERBUCH ZUR DEUTSCHEN RECHTSGESCHICHTE [HRG] (2017), www.hrgdigit.al.de.

74. LUDWIG WITTGENSTEIN, PHILOSOPHICAL INVESTIGATIONS (G.E.M. Anscombe trans., MacMillan Co. 1953).

75. FRIEDRICH MÜLLER, STRUKTURIERENDE RECHTSLEHRE (Duncker & Humblot, Berlin 1984, 2d ed. 1994). For an informative excerpt translated into English, see Friedrich Müller, *Basic Questions of Constitutional Concretization*, 31 RUTGERS L.J. 325, 331 (2000).

theorizing in France, Spain, South Africa, Brazil, and even the United States.⁷⁶ It essentially describes a structural model of legal argumentation that includes social facts and empirical findings as essential building blocks in speaking about—and thereby establishing—law:

Starting from the facts of the case, the interpreter singles out those written rules that might apply. Text and case together allow him to develop the pertinent rule, which typically is more precise than the written rule.⁷⁷

In a way, this methodology implements linguistic pragmatism, despite having been developed prior to encountering the pragmatic turn in linguistics.⁷⁸ At the same time, it also speaks to legal realist audiences: “The interpreter is not only allowed to—but even asked to—look at reality when making up the rule. . . . It would certainly be a challenging task to contrast it with American realism and its aftermath.”⁷⁹

In 1984, Friedrich Müller (based in Heidelberg) and linguist Rainer Wimmer (from Trier) founded the “Heidelberg Working Group on Legal Linguistics,” now headed by one of the contributors to this Article. Today, the group remains the oldest and most influential working group on legal linguistics in Germany, consisting of about forty lawyers, linguists, philosophers, and media scientists. It meets four times a year to discuss current research in progress and has jointly authored numerous interdisciplinary monographs and research articles, most at the Duncker & Humblot Publishing House.⁸⁰ Other interdisciplinary working groups in Germany were founded more recently, such as one at the Berlin-Brandenburg Academy of Sciences (2001–2005), one at Regensburg University

76. See Hanjo Hamann, *Strukturierende Rechtslehre als juristische Sprachtheorie*, HANDBUCH SPRACHE IM RECHT 175 (Ekkehard Felder & Friedemann Vogel eds., 2017); Friedemann Vogel et al., *Computer-Assisted Legal Linguistics: Corpus Analysis as a New Tool for Legal Studies*, LAW & SOC. INQUIRY (forthcoming 2018) [hereinafter *Corpus Analysis*] (each introducing the theory with a graphical illustration). For an example of South African and U.S. reception, see Müller, *supra* note 75.

77. Christoph Engel, *The International Bookshelf*, 18 INT’L J. LEGAL INFO. 266, 267 (1990) (book review).

78. See MÜLLER, *supra* note 75, at 374 (cited in Hamann, *supra* note 76, at 180).

79. Engel, *supra* note 77, at 267–68.

80. For relevant references, see Hamann, *supra* note 76, at 184–86 (“Literatur”).

(around 2004), and one at the University of Halle-Wittenberg (2008).⁸¹

Regarding the substance matter of linguistic analyses of law, it is useful to distinguish two research strands that often get conflated in English nomenclature: legal linguistics (*Rechtslinguistik*) and forensic linguistics (*forensische Linguistik*). The latter refers to a sub-discipline of variational and applied linguistics, developing qualitative and computer supported methods of authorship detection.⁸² It typically studies everyday speech and text types like blackmail letters or voice records, aiming to improve the precision of authorship identification techniques; other methods (such as text type modelling, corpus building, and voice recognition) mainly serve auxiliary functions for this main task. Unlike forensic linguistics, legal linguistics is a specialized variety of communication studies. Legal linguists explore the relationship of legal practice and language as its own end, studying terminology and grammar, speech theories, semiotics and interpretation methods of law, discourse structures in courts, creation of statutes, or genesis of legal norms in general, for instance. The current state of legal linguistics research in Germany can be characterized by three observations.

First, legal linguistics research in Germany tends to be national. Although its subject becomes increasingly transnational and multilingual, there is very little exchange with legal linguistics of other countries. This is evident by just glancing over recent handbooks and introductions on legal linguistics.⁸³

Second, some areas of legal linguistics research in Germany have been more exposed to empirical research than others. For instance, much research has been done on lexical and grammatical characteristics of legal language,⁸⁴ interpretation methods in law and

81. See Vogel, *supra* note 65, at 215–16.

82. EILIKA FOBBE, FORENSISCHE LINGUISTIK: EINE EINFÜHRUNG (2011).

83. *E.g.*, HANDBUCH SPRACHE IM RECHT (Ekkehard Felder & Friedemann Vogel eds., De Gruyter, Berlin 2017); MONIKA RATHERT, SPRACHE UND RECHT (Winter, Heidelberg Univ., 2006) (German discourse), Tiersma & Solan, *supra* note 4 (Anglo-American discourse).

84. Bernd Jeand'Heur, *Die neuere Fachsprache der juristischen Wissenschaft seit der Mitte des 19. Jahrhunderts unter besonderer Berücksichtigung von Verfassungsrecht und Rechtsmethodik*, FACHSPRACHEN: EIN INTERNATIONALES HANDBUCH ZUR FACHSPRACHENFORSCHUNG UND TERMINOLOGIEWISSENSCHAFT 1286 (Lothar Hoffmann et al. eds., 1998).

legal semantics,⁸⁵ argumentation, decision making and semantic struggles,⁸⁶ discourse between professionals and laypeople in courts and administration,⁸⁷ and improving textual comprehensibility and its limitations.⁸⁸ Gaps remain in building a consistent model of legal text genres;⁸⁹ in transnational law and multilingualism in Europe;⁹⁰ in the

85. DIETRICH BUSSE, TEXTINTERPRETATION: SPRACHTHEORETISCHE GRUNDLAGEN EINER EXPLIKATIVEN SEMANTIK (1992); DIETRICH BUSSE, RECHT ALS TEXT: LINGUISTISCHE UNTERSUCHUNGEN ZUR ARBEIT MIT SPRACHE IN EINER GESELLSCHAFTLICHEN INSTITUTION (1992); DIETRICH BUSSE, JURISTISCHE SEMANTIK: GRUNDFRAGEN DER JURISTISCHEN INTERPRETATIONSTHEORIE IN SPRACHWISSENSCHAFTLICHER SICHT (2d ed. 2010); ZUGÄNGE ZUR RECHTSEMANTIK: INTERDISZIPLINÄRE ANSÄTZE IM ZEITALTER DER MEDIATISIERUNG ZWISCHEN INTROSPEKTION UND AUTOMATEN (Friedemann Vogel ed., 2015) [hereinafter ZUGÄNGE ZUR RECHTSEMANTIK].

86. THOMAS COENDET, RECHTSVERGLEICHENDE ARGUMENTATION (2012); EKKEHARD FELDER, JURISTISCHE TEXTARBEIT IM SPIEGEL DER ÖFFENTLICHKEIT (2003); HANS KUDLICH & RALPH CHRISTENSEN, DIE METHODIK DES BGH IN STRAFSACHEN (2009) [hereinafter DIE METHODIK DES BGH]; Hans Kudlich & Ralph Christensen, *Die Kanones der Auslegung als Hilfsmittel für die Entscheidung von Bedeutungskonflikten*, JURISTISCHE ARBEITSBLÄTTER 74 (2004); FRIEDRICH MÜLLER ET AL., RECHTSTEXT UND TEXTARBEIT (1997); RALPH CHRISTENSEN, WAS HEISST GESETZESBINDUNG? EINE RECHTSLINGUISTISCHE UNTERSUCHUNG (1989); Ekkehard Felder, *Semantische Kämpfe außerhalb und innerhalb des Rechts*, 49 DER STAAT 543 (2010).

87. DIE SPRACHE DES RECHTS UND DER VERWALTUNG (Ingulf Radtke ed., 1981); LUDGER HOFFMANN, KOMMUNIKATION VOR GERICHT (1983); RECHTSDISKURSE: UNTERSUCHUNGEN ZUR KOMMUNIKATION IN GERICHTSVERFAHREN (Ludger Hoffmann ed., 1989); Hans-Rüdiger Fluck, *Sprachliche Aspekte der Bürger-Verwaltungs-Kommunikation-Situationsbeschreibung und Forschungsperspektiven*, 114 MUTTERSPRACHE 193 (2004).

88. See Dietrich Busse, *Verständlichkeit von Gesetzestexten - ein Problem der Formulierungstechnik?*, 5 GESETZGEBUNG HEUTE (LEGES, BEM) 29 (1994); KARIN M. EICHHOFF-CYRUS & GERD ANTOS, VERSTÄNDLICHKEIT ALS BÜRGERRECHT?: DIE WRECHTS- UND VERWALTUNGSSPRACHE IN DER ÖFFENTLICHEN DISKUSSION (2008); Markus Nussbaumer, *„Es gibt nichts Gutes, außer man tut es“ - Arbeit an der Verständlichkeit von Gesetzestexten in der Schweizerischen Bundeskanzlei*, HERMES, J. LINGUISTICS 111 (2002); Stephanie Thiemé & Gudrun Raff, *Verständlichkeit von Gesetzestexten und ihre Optimierung in der Praxis: Der Redaktionsstab Rechtsprache beim Bundesministerium der Justiz und für Verbraucherschutz*, HANDBUCH SPRACHE IM RECHT 391 (Ekkehard Felder & Friedemann Vogel eds., 2017); RECHT VERSTEHEN: VERSTÄNDLICHKEIT, MISSVERSTÄNDLICHKEIT UND UNVERSTÄNDLICHKEIT VON RECHT (Kent D. Lerch ed., 2004).

89. Dietrich Busse, *Textsorten des Bereichs Rechtswesen und Justiz*, TEXT- UND GESPRÄCHSLINGUISTIK 658 (Klaus Brinker et al. eds., 2000).

90. FRIEDRICH MÜLLER & ISOLDE BURR, RECHTSSPRACHE EUROPAS: REFLEXION DER PRAXIS VON SPRACHE UND MEHRSPRACHIGKEIT IM SUPRANATIONALEN RECHT (2004); ISABEL SCHÜBEL-PEIFSTER, SPRACHE UND GEMEINSCHAFTSRECHT: DIE AUSLEGUNG DER MEHRSPRACHIG VERBINDLICHEN RECHTSTEXTE DURCH DEN EUROPÄISCHEN GERICHTSHOF (2004); Jan Engberg, *Durchschaubarkeit durch Vielfalt - Vorteile eines mehrsprachigen Rechtssystems und ihre linguistische Beschreibung*, 119 MUTTERSPRACHE 181 (2009).

relationship of law, media and digitalization;⁹¹ and in the creation of legal norms and statutes by the legislature.⁹²

The third observation concerns the practical impact of this research. Doctrinal analysis remains the mainstay of German legal academia, so empirical research strands are rarely perceived (let alone absorbed) by the field's core authorities. Legal linguistics thus remains a marginalized pastime for interdisciplinary zealots rather than a research tradition to which university chairs might be designated. (To the best of our knowledge, there is no legal linguistics chair at any German university.) Despite the long-standing existence of a (somewhat lengthy) textbook on legal methodology, which relies heavily on pragmatist legal linguistics and has seen eleven editions between 1971 and 2013,⁹³ linguistic knowledge barely ever infuses German legal education or vocational training.

Given these observations, it may not be very surprising that there are only a few instances in which legal linguistics research finds its way into practice, as in the case of the Editorial Office for Legal Language (*Redaktionsstab Rechtssprache*) by Stefanie Thieme at the German Ministry of Justice.⁹⁴

C. Legal Linguistics in the Age of New Media and Digitalization

Meanwhile, globalization and digitalization have changed research on language and law, both in Germany and elsewhere. After

91. POLITIK, [NEUE] MEDIEN UND DIE SPRACHE DES RECHTS (Friedrich Müller ed., 2007); THOMAS VESTING, DIE MEDIEN DES RECHTS: SCHRIFT (2011); Ralph Christensen & Kent D. Lerch, *Ein Urteil, wie es im Buche steht. Vom Aufstieg und Niedergang des Gesetzsbuchs*, POLITIK, [NEUE] MEDIEN UND DIE SPRACHE DES RECHTS, *supra* at 221; Martin Morlok, *Intertextualität und Hypertextualität im Recht*, ZUGÄNGE ZUR RECHTSSEMANTIK 69 (Friedemann Vogel ed., 2015); Jan C. Schuhr, *Datenbanken gerichtlicher Entscheidungen als Zugang zu juristischer Semantik?*, ZUGÄNGE ZUR RECHTSSEMANTIK 93 (Friedemann Vogel ed., 2015).

92. FRIEDEMANN VOGEL, LINGUISTIK RECHTLICHER NORMGENESE: THEORIE DER RECHTSNORMDISKURSVITÄT AM BEISPIEL DER ONLINE-DURCHSUCHUNG (2012) [hereinafter LINGUISTIK RECHTLICHER NORMGENESE].

93. FRIEDRICH MÜLLER & RALPH CHRISTENSEN, JURISTISCHE METHODIK I. GRUNDLEGUNG FÜR DIE ARBEITSMETHODEN DER RECHTSPRAXIS (11th ed., 2013).

94. See Stephanie Thieme & Gudrun Raff, *Verständlichkeit von Gesetzstexten und ihre Optimierung in der Praxis: Der Redaktionsstab Rechtssprache beim Bundesministerium der Justiz und für Verbraucherschutz*, HANDBUCH SPRACHE IM RECHT 391 (Ekkehard Felder & Friedemann Vogel eds., 2017).

about twenty-five years of personal computers, the Internet, and new discourse media, the language of social regulation has also changed and found new forms of expression. We have only just begun to understand the present and future effects of this technological development on legal practice and the way that texts are used by lawyers.⁹⁵

One of the first tangible ambitions was for “electronic brains” to automate legal processes. The promise of prejudice-free, more objective, more predictable, more just, and less complex legal decision-making brought legal “cybernetics,” later called “informatics” (*Rechtsinformatik*), to the fore—analyzing legal texts, no longer by the handful, but in thousands and millions. As one of the first legal informatics projects in Germany, a research group in Darmstadt in the 1970s tried to develop a stable, logical, calculable ontology for legal interpretation.⁹⁶ This turned out to be overly ambitious,⁹⁷ and present legal informatics proceeds more carefully,⁹⁸ reflecting the limitations imposed on automation by language: judicial decision-making, i.e., combining texts and arguments from legal literature and precedents with statutes and party writs, is an inherently cognitive-creative endeavor. While introspection must not become the exclusive source of interpretation in legal methods, the other extreme—replacing judges by so-called “subsumption automata,” thus creating “law ex machina”⁹⁹—is equally unwarranted.

German legal linguists thus prefer a middle ground: they use corpora to inform their analysis but continue to rely on fairly qualitative methods, especially hermeneutics, discourse or content analysis, and ethnographical studies. Most German corpora contain

95. See discussion in Morlok, *supra* note 91.

96. See PARAPHRASEN JURISTISCHER TEXTE (Dieter Rave et al. eds., 1971).

97. See Kyriakos N. Kotsoglou, *Subsumptionsautomat 2.0*, 69 JURISTENZEITUNG [JZ] 451 (2014); FRIEDEMANN VOGEL ET AL., RICHTERRECHT DER ARBEIT – EMPIRISCH UNTERSUCHT: MÖGLICHKEITEN UND GRENZEN COMPUTERGESTÜTZTER TEXTANALYSE AM BEISPIEL DES ARBEITNEHMERBEGRIFFS 227 (2015) [hereinafter RICHTERRECHT DER ARBEIT].

98. See, for example, the Munich-based research group Lexalyze at www.lexalyze.de (last visited Jan. 20, 2018).

99. OLIVER RAABE ET AL., RECHT EX MACHINA: FORMALISIERUNG DES RECHTS IM INTERNET DER DIENSTE (2012).

at most a couple hundred texts, which are then manually annotated and explored.¹⁰⁰ One of the contributors to this Article, Friedemann Vogel, was the first to create corpora of several thousand prepared legal texts around 2010–2012 (especially decisions of the German Constitutional Court and other federal courts) and to develop an approach combining qualitative hermeneutics and quantitative methods of computational and corpus linguistics. Calculating recurrent speech patterns and analyzing “sediments” of legal semantics using big data and semi-automated algorithms¹⁰¹ resulted in a new corpus-driven macro perspective on the constitution of doctrine—language, knowledge, and power. Several case studies of this approach have been published.¹⁰²

IV. A MEETING OF THE MINDS: COMPUTER ASSISTED LEGAL LINGUISTICS

As we have argued, evidence-based jurisprudence and legal linguistics have converged on three features: they pragmatically observe their discourse of interest, analyze how descriptions of social “reality” are normatively charged (and clandestinely assert political authority), and increasingly use big data methodology. Their natural fit on these dimensions inspired a new trans-discipline that recently emerged in Germany: Computer Assisted Legal Linguistics (CAL²).¹⁰³ The name’s surface similarity with earlier “Computer-Assisted Legal Research”¹⁰⁴ was unintended and purely coincidental.

100. *E.g.*, JING LI, „RECHT IST STREIT“: EINE RECHTSLINGUISTISCHE ANALYSE DES SPRACHVERHALTENS IN DER DEUTSCHEN RECHTSPRECHUNG (2011); DIE METHODIK DES BGH, *supra* note 86.

101. Friedemann Vogel, *Das Recht im Text: Rechtssprachlicher Usus in korpuslinguistischer Perspektive*, in KORPUSPRAGMATIK 314 (Ekkehard Felder et al. eds., 2012) [hereinafter *Das Recht im Text*].

102. *See generally id.*; LINGUISTIK RECHTLICHER NORMGENESE, *supra* note 92; RICHTERRECHT DER ARBEIT, *supra* note 97; Ekkehard Felder et al., *Patientenautonomie’ und ,Lebensschutz‘*, 44 ZEITSCHRIFT FÜR GERMANISTISCHE LINGUISTIK 1 (2016); Friedemann Vogel, *Das LDA-Toolkit*, 57 ZEITSCHRIFT FÜR ANGEWANDTE LINGUISTIK 129 (2012).

103. For general introductions, see Hanjo Hamann et al., *Computer Assisted Legal Linguistics (CAL²)*, in 29 LEGAL KNOWLEDGE & INFORMATION SYSTEMS 195 (Floris Bex & Serena Villata eds., 2016) [hereinafter *Computer Assisted Legal Linguistics (CAL²)*]; *Corpus Analysis*, *supra* note 76.

104. Eldridge Adams, *The Move Toward Modern Data Management in the Courts*, 23 U. FLA. L. REV. 250, 251 (1971); J. Roger Hamilton, *Computer-Assisted Legal Research*, 51 OR.

A. Conceptual Novelties of CAL²

There are two major differences between this new approach and previous work:

First, CAL² analyzes law, “its language, semantics, knowledge structure, and discourse patterns,” as a social praxis.¹⁰⁵ It focusses on specialized legal vocabularies¹⁰⁶ rather than ordinary language, which was the focus for much previous U.S. research in the plain meaning and original intent traditions.¹⁰⁷ In some sense then, CAL² constitutes a new linguistic legal framework: where legal realists analyzed “what the courts . . . do in fact, and nothing more pretentious,”¹⁰⁸ CAL² analyzes what judges and other jurists write down and how they thus exert power and authority. In this context, the textual work of lawyers is taken seriously and analyzed, but less as a source of law than as a performative speech act.¹⁰⁹ Social practices that aim to change social reality often enough succeed at convincing or out-writing opponents. From this bird’s-eye perspective, law appears as “nothing more pretentious” than “sediments” of legal writing.¹¹⁰ This is distinct from plain meaning inquiry using corpora, but these two methods of inquiry probably differ more in their

L. REV. 665 (1972); William G. Harrington, *A Brief History of Computer-Assisted Legal Research*, 77 LAW LIBR. J. 543 (1984); W. Ronald Robins, *Automated Legal Information Retrieval*, 5 HOUS. L. REV. 691, 694 (1968).

105. *Corpus Analysis*, *supra* note 76, at 5.

106. See Vijay K. Bhatia et al., *Legal Discourse: Opportunities and Threats for Corpus Linguistics*, in DISCOURSE IN THE PROFESSIONS: PERSPECTIVES FROM CORPUS LINGUISTICS 203, 222 (Ulla Connor & Thomas A. Upton eds., 2004).

107. See Stephen C. Mouritsen, *Hard Cases and Hard Data: Assessing Corpus Linguistics as an Empirical Path to Plain Meaning*, 13 COLUM. SCI. & TECH. L. REV. 156 (2011); James C. Phillips et al., *Corpus Linguistics & Original Public Meaning: A New Tool to Make Originalism More Empirical*, 126 YALE L.J. F. 21 (2016); Lawrence M. Solan, *Can Corpus Linguistics Help Make Originalism Scientific?*, 126 YALE L.J. F. 57 (2016); Lawrence M. Solan & Tammy A. Gales, *Finding Ordinary Meaning in Law: The Judge, the Dictionary or the Corpus?*, 1 INT’L J. LEGAL DISCOURSE 253 (2016); Lee J. Strang, *How Big Data Can Increase Originalism’s Methodological Rigor: Using Corpus Linguistics to Reveal Original Language Conventions*, 50 U.C. DAVIS L. REV. 1181 (2017).

108. Such was the oft-quoted commonplace from Oliver Wendell Holmes, Jr., *supra* note 5, at 461.

109. For early accounts of performativity and law, see, for example, Michael Hancher, *Speech Acts and the Law*, in LANGUAGE USE AND THE USES OF LANGUAGE 245 (Roger W. Shuy & Anna Shnukal eds., 1980); Henry E. Nicholson, *Review*, 7 PHIL. & RHETORIC 103, 104 (1974) (reviewing WALTER PROBERT, LAW, LANGUAGE AND COMMUNICATIONS (1972)).

110. RICHTERRECHT DER ARBEIT, *supra* note 97, at 319.

source material (laypeople language versus expert language) than in their conceptual underpinnings.¹¹¹

Second, CAL² does not attempt to create machine-readable, algorithmic, or even cybernetic models of the law, hence restricting computers to *assisting* rather than replacing hermeneutic inquiry.¹¹² This connects “the micro perspective of individual cases and legal arguments with the macro perspective of normative structure and patterns in legal argumentation”¹¹³ but differs on a very basic level from automated content analysis¹¹⁴ and quantitative machine learning.¹¹⁵ In CAL², quantitative data are used merely as a starting point for contextual qualitative reasoning, bridging the methods and epistemological challenges of law and linguistics,¹¹⁶ thereby both revealing and utilizing some of their surprisingly similar features in a globalized world of hypertextualized mass media.¹¹⁷ CAL² thus resembles what has more recently been described as linguistically informed “big data legal scholarship,”¹¹⁸ or “‘text-as-data’ research.”¹¹⁹

111. We thank Larry Solan for answering a clarifying question to this end.

112. Just as evidence-based jurisprudence did not attempt to “replace normative assessments” but to “underpin them empirically, viz. to ‘base them on’ evidence.” EVIDENZBASIERTE JURISPRUDENZ, *supra* note 2, at 14–15.

113. *Corpus Analysis*, *supra* note 76, at 6.

114. *E.g.*, Michael Evans et al., *Recounting the Courts? Applying Automated Content Analysis to Enhance Empirical Legal Research*, 4 J. EMPIRICAL LEGAL STUD. 1007 (2007).

115. *E.g.*, Jonathan Macey & Joshua Mitts, *Finding Order in the Morass: The Three Real Justifications for Piercing the Corporate Veil*, 100 CORNELL L. REV. 99 (2014).

116. See Friedemann Vogel, *Calculating Legal Meanings? Drawbacks and Opportunities of Corpus Assisted Legal Linguistics to Make the Law (More) Explicit*, in THE PRAGMATIC TURN IN LAW 287 (Dieter Stein & Janet Giltrow eds., 2017). For a similar approach, see David S. Law, *Constitutional Archetypes*, 95 TEX. L. REV. 153, 154 (2016) (analyzing “the world’s constitutional preambles using methods from computational linguistics”).

117. Friedemann Vogel et al., “*Begin at the Beginning*”—*Lawyers and Linguists Together in Wonderland*, 6 INT’L J. LANGUAGE & L. 90 (2017) [hereinafter *Begin at the Beginning*].

118. Frank Fagan, *Big Data Legal Scholarship: Toward a Research Program and Practitioner’s Guide*, 20 VA. J.L. & TECH. 1, 2 (2016); accord Frank Fagan, *From Policy Confusion to Doctrinal Clarity: Successor Liability from the Perspective of Big Data*, 9 VA. L. & BUS. REV. 391 (2015).

119. Wolfgang Alschner et al., *Text-as-Data Analysis of Preferential Trade Agreements: Mapping the PTA Landscape* 4 (Ottawa Faculty of Law Working Paper No. 2017-32), <https://ssrn.com/abstract=2999800> (assuming that “[t]he standardized nature of the legal language of international treaties provides particularly fertile ground for the deployment of text-as-data approaches”).

B. State of CAL² and its Challenges

Specific examples of such research have been discussed at two international conferences in Germany: in April 2013, “Legal Corpus Pragmatics: How Language and Media Theory Challenge the Law” in Freiburg,¹²⁰ and in March 2016, “The Fabric of Law and Language: Discovering Patterns through Legal Corpus Linguistics” in Heidelberg.¹²¹ Those conferences were followed by a session on “Computer Assisted Legal Linguistics” at the Freiburg meeting of the International Language and Law Association (ILLA) in September 2017.¹²² As a result of these conferences, German legal researchers have become interested in corpus linguistics to answer questions as varied as the meaning of “enterprise” (*Unternehmen*) in corporate law discourse, and empirical analyses of criminal statutory law.¹²³

One of the major challenges for computer-assisted corpus research in law (both in Germany and elsewhere) is to tap useful and reliable data sources. Commercial databases of legal text such as *Lexis*, *HeinOnline*, and *Westlaw* in the United States, or *Juris*, *Jurion*, and *Beck Online* in Germany provide plenty of relevant content but restrict users to limited document retrieval. They lack any programming interfaces required to harvest large chunks of “text-as-data”—which instead is usually prohibited by T&Cs—so they are of limited usefulness for big data research. Also, they are not

120. For conference reports (in German), see Hanjo Hamann, *Tagungsbericht: Juristische Korpuspragmatik*, 68 JURISTENZEITUNG 829 (2013); Hanjo Hamann & Jana Werner, *Juristische Korpuspragmatik-Konferenzbericht*, 41 DEUTSCHE SPRACHE 285 (2013). Select conference proceedings are also documented in ZUGÄNGE ZUR RECHTSSEMANTIK, *supra* note 85.

121. For a brief report, see *Begin at the Beginning*, *supra* note 117. For a more elaborate report (but in German), see Christoph Lukas, *Korpuslinguistik und Recht*, 103 ARCHIV FÜR RECHTS- UND SOZIALPHILOSOPHIE 138 (2017). For select conference proceedings, see Hanjo Hamann & Friedemann Vogel, *The Fabric of Language and Law—Towards an International Research Network for Computer Assisted Legal Linguistics (CAL²)*, 6 INT’L J. LANGUAGE & L. 101 (introduction to the proceedings, which are available online at <http://www.languageandlaw.de/jll/issue/viewIssue/6/1> (last visited Jan. 20, 2018)).

122. See *Relaunch Conference 2017*, INT’L LANGUAGE & L. ASS’N, www.illa.online/index.php/relaunch-conference-2017 (last visited Jan. 20, 2018).

123. See *Core Projects and Associated Projects*, CAL², <https://www.cal2.eu/index.php/core-projects-and-associated-projects> (last visited Jan. 20, 2018) (a continuously updated list of projects in progress).

designed to balance texts on linguistically relevant dimensions, and they contain textual errors (which barely hamper document retrieval but seriously encumber any quantitative approach).¹²⁴ Consequently, compiling research corpora has been tedious and time-consuming, as both contributors to this Article quickly noted—Vogel in his legal linguistics research and Hamann in a pioneering study of legal citations analysis.¹²⁵

C. Toward a Reference Corpus of Legal Language

The previously noted challenges to corpus research in law have sprung the idea of creating a curated and representative reference corpus of German legal texts for research on legal language and discourses.¹²⁶ This would be similar to the “Corpus of Contemporary American English” (COCA) commonly used in plain meaning inquiry,¹²⁷ or its German equivalent “Deutsches Referenzkorpus” (DeReKo),¹²⁸ but restricted to legal texts like the recently released “Corpus of US Supreme Court Opinions.”¹²⁹

Incidentally, such text databases for empirical analyses of representative samples of legal language had already been foreshadowed by Arthur Nussbaum. More than a century ago, he proposed that some central institution should collect “legally typical or otherwise significant contracts, statutes, broker notes, protocols, security forms, etc. . . . in a systematic and comprehensive manner.”¹³⁰ Nussbaum could not have had linguistic big data analysis in mind, but his idea of centrally collecting representative texts that circulated among professional lawyers and were typical enough to

124. See Schuhr, *supra* note 91.

125. LINGUISTIK RECHTLICHER NORMGENESE, *supra* note 92; Hanjo Hamann, *Die Fußnote, das unbekannte Wesen: Potential und Grenzen juristischer Zitationsanalyse*, 5 RECHTSWISSENSCHAFT 501, 533 (2014) (concluding that the study “required a three-digit number of working hours and an effort close to what a single individual can handle at all”).

126. *Das Recht im Text*, *supra* note 101, at 319.

127. CORPUS CONTEMP. AM. ENG., <http://corpus.byu.edu/coca> (last visited Jan. 20, 2018); see also *supra* note 107.

128. INSTITUT FÜR DEUTSCHE SPRACHE, <https://cosmas2.ids-mannheim.de> (last visited Jan. 20, 2018).

129. CORPUS US SUP. CT. OPINIONS, <http://corpus.byu.edu/scotus> (last visited Jan. 20, 2018).

130. DIE RECHTSTATSACHENFORSCHUNG, *supra* note 9, at 21–22.

serve as templates for the legal language of his day came very close to a reference corpus as we understand it.

In order to now finally create such a corpus, the present authors initiated the CAL² research group (www.cal2.eu) in 2012, funded by the Heidelberg Academy of Sciences. With support by colleagues from various disciplines in Germany and internationally, the group then carefully reviewed the state of the art of legal corpus research¹³¹ and incorporated its learnings into the construction and compilation of its own reference corpus. As the first tier of proof-of-concept funding ran out in mid-2017, the German-language corpus had reached a size of over 6,000 statutes, some 43,000 academic papers, and about 370,000 case law texts of all levels of the judiciary, totaling over 1.3 billion tokens (words).¹³² These texts were extensively checked for errors and inconsistencies, then stored in TEI P5 compliant xml,¹³³ with additional text layers containing linguistic annotations (parts-of-speech), and enriched with all available metadata.¹³⁴ This corpus is now being explored by the authors and cooperators from various institutions in a number of projects on language and law.¹³⁵

The second tier of funding for the CAL² group, which might furnish the corpus with a graphical user interface with built-in statistical tools, began in May 2017.

V. CONCLUSION

Tracing the roots of current German corpus research to the early twentieth century, this Article has shown how an acute awareness of intellectual predecessors can inspire new and original research. Far

131. *Corpus Analysis*, *supra* note 76 (providing a near-comprehensive survey of earlier European corpus projects).

132. Numbers taken from the most recent annual report on the group's activities: Friedemann Vogel & Hanjo Hamann, *Computergestützte Rechtslinguistik (CAL²)—Das Gewirk von Sprache und Dogmatik des Rechts am Beispiel des JuReko-Referenzkorpus*, HEIDELBERGER AKADEMIE DER WISSENSCHAFTEN: JAHRBUCH 2016, at 251, 252 (Uta Hüttig ed., 2017).

133. P5 is the current version of xml structuring guidelines published by TEI, the Text Encoding Initiative. See *TEI: P5 Guidelines*, TEXT ENCODING INITIATIVE, <http://www.tei-c.org/Guidelines/P5/> (last updated Nov. 29, 2017).

134. Details in a companion piece for the computer science audience: *Computer Assisted Legal Linguistics (CAL²)*, *supra* note 103, at 196–97.

135. See *Core Projects and Associated Projects*, *supra* note 123.

from being intellectual trivia, academic origin stories benefit a field in at least three ways: They foster progress by highlighting common roots and shared interests in seemingly disparate domains. They foster self-reflection by questioning the originality and novelty of ideas that may be traceable to even a century ago. And they foster interdisciplinarity by unearthing links and points of contact that got buried in today's subject matter specialization.

This Article illustrated these three benefits in the case of corpus research and its German incarnation called Computer Assisted Legal Linguistics (CAL²). Joining two different strands of research that each observed their discourse of interest pragmatically, all the while being sensitive to power implications, and preferring big data strategies, CAL² has resulted in a new transdiscipline that interlocks with similar U.S. projects. Looking back to get ahead, corpus research will escape from the long shadow of previous empirical traditions and prove to be a valuable “new tool for legal studies” in the toolbox of academic and practical lawyers.¹³⁶

136. *Corpus Analysis*, *supra* note 76 (quote from title).