Forensic Linguistics: Science or Fiction?

ABIGAIL SHIM[†]

The history of linguistics is meager and splintered due to the subject's interdisciplinary nature. In the postwar era, the discipline attempted to revive as a scientific one, spearheaded by Noam Chomsky and his theory of generative grammar. Linguistics consequently broke away from the predominant structuralist approach of the nineteenth century, returning to rationalist roots. But with the rise of computer technology, Chomsky's critiques of empirical, applicational linguistic approaches have lost their force. As academic linguistics splinters off again, loses its scientific edge, and regroups with the humanities, linguistics applied in the forensic context may implicate more questions than it answers, fundamental questions about humans and language that linguists are still unable to solve: What is *language? Do we use language in a way that is uniquely identifiable? Should we look at language use from a societal or individualized, psychological perspective? This Note seeks to reveal these tensions, by providing an overview of the historical development of forensic linguistics; highlights the theory of idiolect backing the use of forensic linguistic evidence; and critiques idiolect and forensic linguistics' statistical turn in light of linguistics' ebbing scientific status. As the larger epistemological questions behind forensic linguistic theory remain indeterminate, authorship identification may remain a question of weight, similarity, and difference for judges and juries to grapple with, highlighting the "sliding scale" problem of reliability in the forensic sciences.*

[†] J.D. Candidate 2023, University of California, Hastings College of the Law; Editor-in-Chief, *Hastings Law Journal*. I am wholly indebted to Professor Binyamin Blum, whose brilliant teaching and guidance incubated this Note. I am also immensely grateful for the exceptional work of Jesse Honig, Executive Notes Editor; Nikol Nesterenko, Executive Technology Editor; Arlen Moradi and Isabella Ysebrands, Executive Production Editors; Sam Dodson, Senior Production Editor; and our indispensable staff editors. Any remaining errors or omissions are my own.

TABLE OF CONTENTS

INTRODUCTION

In 1989, Robin Lakoff, a linguist at the University of California, Berkeley, offered unsettling reflections on the present state of linguistics:

> Suppose the social 'sciences' in general, and linguistics in particular, have yet to demonstrate the appropriateness of the scientific method to their subject matter, the working of the human mind. Suppose a large part of the work of this field is . . . still humanistic at heart: dedicated to figuring out what is individual . . . and therefore beyond the reach of statistics, of replicable experimentation. Then what happens if we graft the argumentative techniques that work so well for science upon a humanistic study? What will we get? Modern linguistics, I suggest—to its misfortune . . . .[1]

She reaffirmed these views in 2000, characterizing the field as a social rather than hard science in her book *The Language War*,[2] which seeks to use discourse analysis to "make a true rapprochement" between linguistics and other social sciences like literary analysis, psychology, anthropology, and political science.[3] She candidly admits that by doing so, she engages not with what is considered "linguistics," but "contemporary psychoanalysis and literary theory";[4] that linguistics proper rejects her approach as unscientific;[5] and that there is "no extrinsic, objective, 'scientific' test" by which to verify or falsify her claims.[6] She then reiterates the broader question raised in her 1989 memoir: "Are interpreters of human communication . . . engaged in scientific or humanistic enterprise?"[7]

Yet in 1997, Lakoff submitted an affidavit on behalf of Theodore Kaczynski supporting a motion to suppress seized writings linking him with the infamous Unabomber.[8] Lakoff's affidavit challenged the FBI's assessment of the "shared linguistic analysis" between Kaczynski's and the Unabomber's writings,[9] with unmistakable irony: a professional linguist unconvinced of the scientific rigor of her discipline offered her expertise to attack the lack of rigor in the opposing side's linguistic methodology. While it remains unclear whether Lakoff's findings would have been admissible at trial,[10] this anecdote illustrates the troubling dichotomy reflected in the history of forensic linguistics.

---

1. Robin Lakoff, *The Way We Were; or; the Real Actual Truth About Generative Semantics: A Memoir*, 13 J. PRAGMATICS 939, 967 (1989).

2. ROBIN TOLMACH LAKOFF, THE LANGUAGE WAR 5 (2000).

3. *Id.* at 5, 8 (defining "discourse analysis" as the analysis of the "processes by which we understand larger and more abstract units of language").

4. *Id.* at 6.

5. *Id.* at 8.

6. *Id.* at 9–10.

7. *Id.*

8. LAWRENCE M. SOLAN & PETER M. TIERSMA, SPEAKING OF CRIME: THE LANGUAGE OF CRIMINAL JUSTICE 162 (2005).

9. *Id.* at 163.

10. *Id.* at 164.

Forensic linguistics emerged only recently in the 1990s from the fragmentation of linguistics proper after the "Chomskyan revolution" of the 1960s[11] as one of many liaisons between linguistics and other social sciences.[12] While linguistics lost its dominance and internal unity,[13] linguistic evidence increasingly found its way into the courts.[14] To this day, some of forensic linguistics' foremost advocates are lawyers with PhDs in linguistics who focus exclusively on the intersection between linguistics and the law.[15]

In forensic linguistics' earliest stages, experts testified exclusively on an experiential basis, picking out idiosyncratic usage in the unidentified document to identify the author—much like handwriting experts[16]—and their methods were not considered "scientific."[17] With the advent of sophisticated computer modeling and a growing attention to corpus linguistics,[18] however, forensic linguistic techniques may experience a revival in authorship attribution cases.[19] Thanks to these developments, linguistic expert evidence has become

---

11. Roger Shuy, *Language and Law*, *in* THE HANDBOOK OF LINGUISTICS 627, 627 (Mark Aronoff & Janie Rees-Miller eds., 2d ed. 2017); PETER H. MATTHEWS, A SHORT HISTORY OF STRUCTURAL LINGUISTICS 25 (2001).

12. Lakoff, *supra* note 1, at 964.

13. Ernst F.K. Koerner, *On the Place of Linguistic Historiography Within the Sciences of Language, Again*, *in* ESSAYS IN THE HISTORY OF LINGUISTICS 3, 12 (E.F.K. Koerner ed., 2004).

14. *See* Shuy, *supra* note 11 ("[N]ow linguists are applying their field's knowledge to such areas as statutory law and interpretation, voice and authorship identification, jury instructions, the asymmetry of power in courtroom exchanges, lawyer-client communication, police interrogation practices, contract disputes, legal discourse, defamation, trademark infringement, courtroom interpretation and translation, copyright disputes, discrimination, commercial warning messages, and various types of criminal charges such as perjury, bribery, solicitation, money laundering, threatening, and fraud.").

15. Lawrence M. Solan, a law professor at Brooklyn Law School, holds a PhD in linguistics from the University of Massachusetts and has written extensively on the use of linguistic evidence in the courts. *See generally, e.g.*, Lawrence M. Solan, Dieter Stein & Peter M. Tiersma, *Introducing Language & Law*, 1 INT'L J. LANG. & L. 1 (2012). Peter M. Tiersma was a professor at Loyola Law School and held a PhD in linguistics from UC Berkeley. *See generally, e.g.*, PETER M. TIERSMA, LEGAL LANGUAGE (1999); PETER M. TIERSMA, SPEAKING OF LANGUAGE AND LAW (Lawrence M. Solan et al. eds., 1st ed. 2015). Janet Ainsworth is a professor of law at Seattle University. *See generally, e.g.*, Janet Ainsworth, *Who Wrote This? Modern Forensic Authorship Analysis as a Model for Valid Forensic Science*, 96 WASH. L. REV. 1159 (2019). Roger Shuy is a retired professor of linguistics at Georgetown University who has written numerous works on forensic linguistics. *See generally* Shuy, *supra* note 11. He also testified as Quality Inn's linguistic expert in the "McSleep" trademark case. *See generally* Quality Inns Int'l v. McDonald's Corp., 695 F. Supp. 198 (D. Md. 1988).

16. *See generally, e.g.*, DON FOSTER, AUTHOR UNKNOWN: TALES OF A LITERARY DETECTIVE (2000); JOHN OLSSON, WORD CRIME: SOLVING CRIME THROUGH FORENSIC LINGUISTICS (2009).

17. *See, e.g.*, United States v. Van Wyk, 83 F. Supp. 2d 515, 523 (D.N.J. 2000) (identifying the "lack of scientific reliability of forensic stylistics," a method that compares shared features and differences between sets of documents based on the examiner's experience and personal observation).

18. "Corpus linguistics," broadly defined, refers to "the study of language based on examples of real-life language use," typically by analyzing a collected dataset of speech. TONY MCENERY & ANDREW WILSON, CORPUS LINGUISTICS: AN INTRODUCTION 1 (2d ed. 2001) (internal quotation marks omitted).

19. Moshe Koppel, Jonathan Schler & Shlomo Argamon, *Computational Methods in Authorship Attribution*, 60 J. AM. SOC'Y INFO. SCI. & TECH. 9, 9 (2009) ("More recently, this problem of authorship attribution has gained greater prominence due to new applications in forensic analysis . . . and the development of computational methods for addressing the problem."); *see also* United States v. Clifford, 704 F.2d 86, 90–91 (3d Cir. 1983) (recognizing a linguistic expert's computerized, frequency-based comparison of writings as a "scientific process").

a mainstay in trademark cases,[20] a rising force in statutory and constitutional interpretation,[21] and the occasional appearance in authorship identification cases.[22] Yet the divergent uses, methodologies, and admissibility of linguistic evidence may reflect deeper divides within linguistics proper surrounding metaphysical questions of meaning and the mind.[23] And if various linguistic factions trenchantly disagree on a foundational theoretical basis, the question becomes whether forensic linguistic methodologies are truly "reliable."[24]

This Note traces the historical development of forensic linguistics to question the subliminal, classically forensic theory behind applying linguistic methodologies in authorship identification cases—namely, that each person has a distinct use of language by which they can be identified.[25] By reducing the study of language to corpus linguistic techniques slapped onto a patently forensic theory of language use, forensic linguistic evidence masquerades behind the appearance of empirical rigor, casting doubt upon its reliability as existing apart from the courts.[26]

Part I traces the historical emergence of forensic linguistics around the 1990s[27] with the downfall of the Chomskyan hegemony and the fragmentation of linguistics proper. It would be impossible to provide a comprehensive history, given the scope of this Note. But identifying several themes from certain chapters in linguistic history will illuminate issues with the use of linguistic evidence in court, particularly in the authorship identification context. Part II traces parallel developments in the admissibility of linguistic evidence in authorship identification cases. Finally, Part III challenges the theory of idiolect backing the use of forensic linguistic evidence in authorship identification cases in light of linguistics' broader historical development.

---

20. ROGER W. SHUY, LINGUISTIC BATTLES IN TRADEMARK DISPUTES 23 (2002).

21. *See generally*, *e.g.*, Thomas R. Lee & Stephen C. Mouritsen, *Judging Ordinary Meaning*, 127 YALE L.J. 788 (2018); 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); Jennifer L. Mascott, *Who Are Officers of the United States*, 70 STAN. L. REV. 443 (2018).

22. SOLAN & TIERSMA, *supra* note 8, at 151.

23. Koerner, *supra* note 13, at 7 ("[L]inguistics [is] a discipline in which the coexistence of diverging theoretical views and possibly contrasting methodological procedures constitutes perhaps the most important element . . . .").

24. *See* Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579, 590–95 (1993).

25. JOHN GIBBONS, FORENSIC LINGUISTICS: AN INTRODUCTION TO LANGUAGE IN THE JUSTICE SYSTEM 297 (2003).

26. Jennifer Mnookin, *Scripting Expertise: The History of Handwriting Identification Evidence and the Judicial Construction of Reliability*, 87 VA. L. REV. 1723, 1728–29 (2001).

27. *See* Malcolm Coulthard, *Author Identification, Idiolect, and Linguistic Uniqueness*, 25 APPLIED LINGUISTICS 431, 431 (2004).

I. THE HISTORY OF FORENSIC LINGUISTICS

A. METHODOLOGY

There are several significant challenges to conducting a historical survey of linguistics. The historiography of linguistics itself is as young as the emergence of linguistics as an independent discipline,[28] and there is a deficiency of scholarly work on the subject.[29] Moreover, because the history of linguistics is a narrative of "diverging theoretical views and . . . contrasting methodological procedures,"[30] there are lively debates over whether certain moments in linguistic history should be characterized as "scientific revolutions" or as series of continuities and discontinuities.[31] Linguistic historiographic methodology is still hotly debated, and no single "canon of [linguistic] historiographic research . . . [has been] widely accepted by the scholarly community."[32] These problems are even more apparent given linguistics' loss of autonomy and dissolution into diverse subfields.[33] The history of linguistics is also arguably "a history of misreadings," due to fundamentally differing ideological views on the basic goals of the discipline and assumptions about how language works.[34] These challenges are somewhat curbed by the constricted scope of this Part, which focuses on the latest chapter of linguistic history following the "Chomskyan revolution."[35]

Additionally, this Part focuses on Anglo-European linguistic traditions, for several reasons. First, it is "European science that has become international science," including in the realm of linguistics.[36] Second, the developing tradition of European linguistics from Hellenistic antiquity has been relatively linear and well documented, making it reasonable "to make the history of European

28. Ernst F.K. Koerner, *Toward a History of Americanist Linguistics*, *in* TOWARD A HISTORY OF AMERICAN LINGUISTICS 17, 18 (2002).

29. Koerner, *supra* note 13, at 5–7 ("It is therefore curious to note that, whereas the . . . natural sciences have enjoyed the establishment of courses devoted to the history of their own discipline, no comparable arrangement exists with regard to linguistics . . . .").

30. *Id.* at 7.

31. *Compare* THOMAS KUHN, THE STRUCTURE OF SCIENTIFIC REVOLUTIONS (4th ed. 2012), *with* Ernst F.K. Koerner, *Continuities and Discontinuities in the History of Linguistics*, *in* PRACTICING LINGUISTIC HISTORIOGRAPHY: SELECTED ESSAYS 69 (E.F.K. Koerner ed., 1989), *and* NOAM CHOMSKY, CARTESIAN LINGUISTICS: A CHAPTER IN THE HISTORY OF RATIONALIST THOUGHT 107–08 (3d ed. 2009), *and* JOHN EARL JOSEPH, FROM WHITNEY TO CHOMSKY 68 (2002) ("Through all the vagaries of Neogrammarianism, structuralism, and generativism, behaviourism and universalism, rationalism and empiricism, the path of development has continued unbroken.").

32. Ernst F.K. Koerner, *Persistent Issues in Linguistic Historiography*, *in* THE HISTORY OF LINGUISTICS, 1993: PAPERS FROM THE 6TH INTERNATIONAL CONFERENCE ON THE HISTORY OF THE LANGUAGE SCIENCES (ICHOLS VI), WASHINGTON, D.C., 9–14 AUGUST 1993, at 3–4 (Kurt R. Jankowsky ed., 1995).

33. Koerner, *supra* note 13, at 12.

34. JOSEPH, *supra* note 31, at 133; *see also* Koerner, *supra* note 32, at 6; Ernst F.K. Koerner, *Linguistics and Ideology: A Neglected Aspect of 19th and 20th Century Historiography*, *in* LINGUISTIC HISTORIOGRAPHY: PROJECTS & PROSPECTS 39, 54 (E.F.K. Koerner ed., 1999).

35. *See generally* András Kertész, *From 'Scientific Revolution' to 'Unscientific Revolution': An Analysis of Approaches to the History of Generative Linguistics*, 32 LANG. SCI. 507 (2010).

36. R.H. ROBINS, A SHORT HISTORY OF LINGUISTICS 7 (1967).

linguistics the foundation for a history of linguistics as a whole,"[37] if not the complete account.

Finally, this Part focuses on two monolithic figures, Chomsky and Saussure, and the two leading philosophies of language they espoused that shaped modern linguistics: structuralism from the 1930s to 1950s, and generativism from the 1960s to 1990s.[38] It is from the wane of "the Chomsky-centred linguistic universe" and linguistics' corresponding breakdown into vying subfields that forensic linguistics emerges.[39]

B. FROM PHILOSOPHY TO THE SCIENCE OF LANGUAGE: STRUCTURALISM AND SAUSSURE

The fabric of nineteenth and twentieth century linguistics is convoluted, yet defined by a key development: the discipline's break from the humanities and alignment with the natural sciences, thanks to the structuralist movement.[40] By the nineteenth century, efforts to shift linguistics from the humanities into the natural sciences were well underway.[41] These efforts went hand in hand with the broader modern turn to empiricism, separating philosophy from the empirical sciences and leading to the specialization of knowledge, especially among academic institutions.[42] Previously, linguistics was simply an aspect of subjects like philosophy, rhetoric, aesthetics, pedagogy, and philology.[43] It was linguistics' disassociation from philosophy and achievement of scientific status that principally allowed it to establish itself as a robust, independent discipline.[44]

The scientific legitimization of linguistics rode on the back of structuralism and its foremost thinker, Saussure. Around the turn of the twentieth century, knowledge became increasingly specialized. Intellectuals saw "the differing nature of inquiry in the natural sciences . . . and in the human sciences," leading to the emergence of psychology, sociology, and political economy as autonomous disciplines, a shift bound to powerfully impact linguistics.[45] Amidst this paradigm shift, Saussure's structuralism decisively moved linguistics into

37. *Id.*

38. Antonio Pennisi, *The Beginnings of Psycholinguistics: Natural and Artificial Signs in the Treatment of Language Disorders*, *in* HISTORICAL ROOTS OF LINGUISTIC THEORIES 85, 86 (Lia Formigari & Daniele Gambarara eds., 1995); JOSEPH, *supra* note 31, at 48.

39. JOSEPH, *supra* note 31, at 48.

40. MATTHEWS, *supra* note 11, at 153.

41. Lyle Campbell, *The History of Linguistics*, *in* THE HANDBOOK OF LINGUISTICS 81, 106 (Mark Aronoff & Janie Rees-Miller eds., 2d ed. 2017).

42. Lia Formigari, *Linguistic Historiography Between Linguistics and Philosophy of Language*, *in* HISTORICAL ROOTS OF LINGUISTIC THEORIES 1, 2 (Lia Formigari & Daniele Gambarara eds., 1995).

43. Campbell, *supra* note 41, at 97; Ernst F.K. Koerner, *Pilot and Pirate Disciplines in the Development of Linguistic Science*, *in* PRACTICING LINGUISTIC HISTORIOGRAPHY: SELECTED ESSAYS 245, 246 (1989).

44. Formigari, *supra* note 42, at 5.

45. Ernst F.K. Koerner, *The Neogrammarian Doctrine: Breakthrough or Extension of the Schleicherian Paradigm. A Problem in Linguistic Historiography*, *in* PRACTICING LINGUISTIC HISTORIOGRAPHY: SELECTED ESSAYS 79, 95 (1989).

the natural sciences and toward independence by characterizing the objective of the discipline as the scientific study of language as a system.[46]

Saussure's structuralism is defined by several important features. First, Saussure drew a distinction between viewing language synchronically rather than diachronically.[47] A synchronic approach focuses on discerning the universal principles of language from the study of a single language and views language as a totality, where none of a language's terms are viewed in isolation apart from the system.[48] This view diverged from the contemporary diachronic approach, where linguists compared different languages to study their historical change through time.[49] In Saussure's synchronic accounting, the goal of linguistics was to intensively analyze an individual language in order to make out "a pure linguistics of the language system, to which the dimension of time and history is irrelevant."[50] This marked the rise of descriptive linguistics, or the "study of the formal properties of language abstracted from the actuality of their use."[51]

Second, Saussure identified language as semiotic, or as a system of signs.[52] Saussure believed that the "relationship between the signifier and the signified is arbitrary," supporting "the relative autonomy of language in relation to reality."[53] Lastly, Saussure drew a distinction between *langue*, the individual language, and *parole*, the communicative speech act.[54] Saussure primarily viewed language as a social phenomenon, skirting the physical and physiological aspects of the individual speech act and attempting instead to discern the language system from the perspective of the community of speakers.[55] After all, language exists in group consciousness, and individual speech acts vary tremendously from person to person, reflecting the synchronic view that looking at a single instance of language use outside of the system is futile.[56] Saussure thus characterized the linguist's objective as understanding the *langue* of each community.[57]

Saussure's greatest contribution to modern linguistics was structuralizing the study of language through formal reasoning, fencing off linguistics as an

46. FERDINAND DE SAUSSURE, COURSE IN GENERAL LINGUISTICS 6, 11 (1959).

47. *See* Joshua T. Katz, *Saussure at Play and His Structuralist and Post-Structuralist Interpreters*, 68 CAHIERS FERDENAND DE SASSURE 113, 121 (2015).

48. JOHN LECHTE, FIFTY KEY CONTEMPORARY THINKERS: FROM STRUCTURALISM TO POSTMODERNITY 150 (1994).

49. JOSEPH, *supra* note 31, at 48.

50. MATTHEWS, *supra* note 11, at 10; *see also* BORIS GASPAROV, BEYOND PURE REASON: FERDINAND DE SAUSSURE'S PHILOSOPHY OF LANGUAGE AND ITS EARLY ROMANTIC ANTECEDENTS 118 (2013).

51. Henry Widdowson, *Disciplinarity and Disparity in Applied Linguistics*, *in* VOICES AND PRACTICES IN APPLIED LINGUISTICS: DIVERSIFYING A DISCIPLINE 33, 35 (Claire Wright et al. eds., 2019).

52. SAUSSURE, *supra* note 46, at 16.

53. LECHTE, *supra* note 48, at 150.

54. MATTHEWS, *supra* note 11, at 11–12.

55. *Id.* at 12; LECHTE, *supra* note 48, at 151.

56. LECHTE, *supra* note 48, at 19.

57. ROBINS, *supra* note 36, at 225.

independent, scientific discipline. As a result, linguistics became increasingly reliant upon crafting robust, all-encompassing theories for its legitimacy and autonomy, removing it from its prior, more practical applications of teaching grammar and compiling dictionaries.[58] This reliance on theory was pivotal, as theory would define the according field of linguistic investigation and methodology.[59] Under Saussure's new paradigm, "linguistics began to change from a largely parasitic discipline, . . . a field borrowing terms, concepts, and methods from other fields, to a pilot discipline, . . . a field of scientific investigation offering its insights, procedures, and results to other disciplines."[60]

C. INSTITUTIONALIZING LINGUISTICS

Linguistics coming into its own in both popular and intellectual consciousness was far more than a matter of theory, however. The discipline's new theory-oriented goals also coincided with the growth of universities, transatlantic European and American scholarship, and increasing academic publication.[61] While Saussure promulgated structuralism and descriptive linguistics in Europe, American universities continued the cause into the interwar decades in ways that would radically shape the linguistic landscape.[62]

The First World War brought a "widespread sense of liberation from a century of German linguistic dominance[,] . . . [and from] the 1920s onward a national linguistics came to mean a more or less original theoretical position held by a nation's leading linguists."[63] Linguistics is thus "a relatively young academic discipline in America . . . [and] did not fully take root in the academic system until after World War II."[64] As supporting evidence, the Linguistic Society of America (LSA) and its accompanying periodical *Language* were established in 1924.[65] The world war efforts and an increasing sense of globalization, accompanied by dedicated military and government funding, helped linguistics flourish in America.[66]

This interest continued into the Cold War with developments in machine translation, granting linguists funding at prestigious centers like Berkeley, MIT, and Georgetown.[67] Thanks to linguistics' tremendous success throughout the interwar years, the discipline experienced massive growth in the postwar era with burgeoning departments at universities across the nation and the emergence

58. Koerner, *supra* note 43, at 247.
59. *Id.*
60. *Id.* at 254.
61. ROBINS, *supra* note 36, at 189.
62. *Id.* at 235–36.
63. JOSEPH, *supra* note 31, at 16.
64. Janet Martin-Nielsen, *A Forgotten Social Science? Creating a Place for Linguistics in the Historical Dialogue*, 47 J. HIST. BEHAV. SCIS. 147, 159 (2011).
65. Leonard Bloomfield, *Why a Linguistic Society?*, 1 LANG. 1, 1–5 (1925).
66. Martin-Nielsen, *supra* note 64, at 151–52.
67. *Id.* at 153.

of professional societies like the LSA.[68] While other social sciences suffered in the politically charged climate of the times, linguistics' myopically theoretical focus reinforced the discipline's "scientificization."[69] Linguistics' mathematical turn and the increasing association of science with academic prestige[70] supported this recharacterization.

It was in this climate that Noam Chomsky rose from MIT's Research Laboratory of Electronics,[71] dramatically furthering these developments with a scathing review of Skinner that spearheaded the "neo-empiricist revolution" outmoding behaviorism,[72] reviving mentalism, and launching the rise of cognitive science and cognitive psychology.[73] With Chomsky, moreover, began the "evolution of the American linguist as a professional figure."[74]

D. THE "CHOMSKYAN REVOLUTION"

"Chomsky is currently among the ten most-cited writers in all of the humanities (behind only Marx, Lenin, Shakespeare, the Bible, Aristotle, Plato, and Freud) and the only living member of the top ten."[75] Chomsky's work has been credited as "one of the first serious attempts on the part of a linguist to construct within the tradition of scientific theory-construction a comprehensive theory of language which may be understood in the same sense that a chemical [or] biological theory is ordinarily understood by experts in those fields."[76] Hence, the next chapter of modern linguistics is often called the "Chomskyan revolution,"[77] after one of the foremost intellectuals of our time.

Before Chomsky, American structuralist linguistics was largely behaviorist, led by Leonard Bloomfield and his incredibly influential work, *Language*.[78] Bloomfieldian structuralism resisted mentalism,[79] positing that language structure could only be discerned from the speech act itself, and therefore that discerning the system of language revolved

68. *Id.* at 154. *See generally* Archibald A. Hill, *The Linguistic Society of America and North American Linguistics, 1950–1968*, 18 HISTORIOGRAPHIA LINGUISTICA 49 (1991).

69. Martin-Nielsen, *supra* note 64, at 162; JOSEPH, *supra* note 31, at 47.

70. JOSEPH, *supra* note 31, at 47.

71. Martin-Nielsen, *supra* note 64, at 153.

72. *See generally* Noam Chomsky, *A Review of B.F. Skinner's* Verbal Behavior, *in* READINGS IN PHILOSOPHY OF PSYCHOLOGY, VOLUME I, at 48–63 (Ned Block ed., 1983).

73. Martin-Nielsen, *supra* note 64, at 163.

74. *Id.* at 169.

75. STEVEN PINKER, THE LANGUAGE INSTINCT 23 (1994).

76. Robert Lees, *Syntactic Structures by Noam Chomsky*, 33 LANG. 375, 377 (1957) (book review).

77. *See* Kertész, *supra* note 35.

78. *See generally* LEONARD BLOOMFIELD, LANGUAGE (1935).

79. *See generally* Leonard Bloomfield, *Language or Ideas?*, 12 LANG. 89 (1936). "Mentalism" can be defined as "[a]n orientation to the study of behavior, which holds that a unique, a necessary, and the primary contribution to the causal explanation of behavior consists in proposing various internal acts, states, mechanisms or processes, presumed to be operating in neural, conceptual, or psychic dimensions." Jay Moore, *On Mentalism, Privacy, and Behaviorism*, 11 J. MIND & BEHAV. 19, 20 (1990).

around studying stimuli and responses.[80] Whether the Chomskyan regime was the inevitable response to Bloomfieldian behaviorism,[81] and whether it was truly a "revolution" in linguistics[82]—two hotly contested questions in linguistic historiography[83]—Chomsky "brought fully-blown structuralism to American linguistics for the first time,"[84] reinventing American linguistics as a theoretical science.[85]

In 1959, Chomsky wrote a scathing review of Skinner that heralded the generativist turn.[86] When *Syntactic Structures* was published shortly thereafter,[87] the shift was realized with the advent of Chomsky's chief contribution to linguistics: the transformational generative grammar (TGG), a grammar[88] "generat[ing] a set of structural descriptions, each of which . . . incorporates a deep structure, a surface structure, a semantic interpretation (of the deep structure), and a phonetic interpretation (of the surface structure)."[89] The "deep structure" of a given sentence refers to its units of meaning, while the "surface structure" governs the actual words used to achieve those meanings and their sounds. Chomsky generally avoided semantics,[90] though he intimated that the deep structure of his TGG theory could explain derivative semantic meaning through the pairings of deep and surface structures

80. *See* Leonard Bloomfield, *On Recent Work in General Linguistics*, 25 MOD. PHILOLOGY 211, 212 (1927); Leonard Bloomfield, *A Set of Postulates for the Science of Language*, 15 INT'L J. AM. LINGUISTICS 195, 196 (1949). *See generally* Leonard Bloomfield, *Secondary and Tertiary Responses to Language*, 20 LANG. 45 (1944); Leonard Bloomfield, *Linguistics as a Science*, 27 STUD. PHILOLOGY 553 (1930); Leonard Bloomfield, *Linguistic Aspects of Science*, 2 PHIL. SCI. 499 (1935).

81. MARCUS TOMALIN, LINGUISTICS AND THE FORMAL SCIENCES: THE ORIGINS OF GENERATIVE GRAMMAR 13 (2006). "Behaviorism" refers to "[a]n orientation to the study of behavior that assumes it must be possible, in principle, to secure a full, lawful explanation of any instance of behavior, including verbal behavior, in terms of present and past behavioral, physiological, and environmental variables, without mentioning the realm of the mental." Moore, *supra* note 79, at 22.

82. Ernst F.K. Koerner, *The Chomskyan 'Revolution' and Its Historiography: Observations of a Bystander*, *in* PRACTICING LINGUISTIC HISTORIOGRAPHY: SELECTED ESSAYS 101, 134 (1989). *See generally* Frederick J. Newmeyer, *Has There Been a 'Chomskyan Revolution' in Linguistics?*, 62 LANG. 1 (1986).

83. Some historians resist characterizing this chapter in linguistic history as a "revolution," attributing Chomsky's dominance to his "slash and burn" treatment of previous linguistic movements, charisma, and tight organizational control. *See* Ernst F.K. Koerner, *On 'Influence' in Linguistic Historiography: Morphophonemics in American Structuralism*, *in* ESSAYS IN THE HISTORY OF LINGUISTICS 65, 69 (Ernst F.K. Koerner ed., 2004); TOMALIN, *supra* note 81, at 21, 51; Lakoff, *supra* note 1, at 946 (disputing the accuracy of characterizing Chomsky's influence as a Kuhnian paradigm shift). While they are correct in the sense that Chomsky's views have failed to unify the discipline, his intellectual dominance in the field is nevertheless established by the fact that contemporary linguistic developments were very much reactions to his work. *See* ROBINS, *supra* note 36, at 260–63; Lakoff, *supra* note 1, at 941.

84. JOSEPH, *supra* note 31, at 157; MATTHEWS, *supra* note 11, at 144; David Golumbia, *The Language of Science and the Science of Language: Chomsky's Cartesianism*, 43 DIACRITICS 38, 43 (2015).

85. ASA KASHER, THE CHOMSKYAN TURN 99 (1991).

86. *See generally* Chomsky, *supra* note 72.

87. *See generally* NOAM CHOMSKY, SYNTACTIC STRUCTURES (1957).

88. A "grammar" can be defined as the idealized system of rules governing the "ideal speaker-listener." NOAM CHOMSKY, ASPECTS OF THE THEORY OF SYNTAX 3 (1969).

89. NOAM CHOMSKY, LANGUAGE AND MIND 126 (1972).

90. CHOMSKY, *supra* note 88, at 52–53.

and their transformational relationships.[91] Instead, Chomsky based his science of linguistics almost exclusively on syntax,[92] using complex formal trees generated by his phrase-structure rules.[93]

Chomsky thus sought to find an all-encompassing theory that would explain the basic rules of a language system, the goal of which was to account for all possible grammatical utterances of the language-speaker's "competence."[94] This was an ambitious task, given the infinite possible grammatical uses of language and arbitrary complexity of language-users' sentences.[95] But this is also why competence—the internalized, innate language faculty of the language-speaker—was Chomsky's subject of analysis rather than language "performance," the actual, observable instances of language use.[96] Bloomfield and Saussure both struggled with the problem of language variation between speakers.[97] Chomsky's generative grammar sought to address this problem by explaining variation as grounds for adjusting the rules of his theoretical grammar.[98] Nevertheless, variation would continue to present a recurring thorn in TGG's side.[99]

Chomsky's envisioned goal for linguistics was a marked departure from the "controlled division of utterances into phonemes, morphemes, [and] phrases" then dominant in America.[100] Chomsky replaced behaviorism's view that language grammaticality was constructed from repetition and training with the concept of a language-speaker's creativity, which could account for how native language users immediately understand and craft new language formulations they have never encountered. Thus, it was not empirical data, but a speaker's own introspection that provided chief guidance on

91. *Id.* at 27, 52.

92. Golumbia, *supra* note 84, at 42; *see also* CHOMSKY, *supra* note 88. "Syntax" refers to how smaller units of language (individual words) form larger units (phrases and sentences). Lakoff, *supra* note 1, at 941.

93. *See* TOMALIN, *supra* note 81, at 53. Chomsky's phase-structure rules describe a language's "grammar," which attempts to comprehensively explain how meaning and sound interact. *See generally* CHOMSKY, *supra* note 88.

94. CHOMSKY, *supra* note 89, at 23, 102–03.

95. *Id.* at 105.

96. *Id.* at 98, 102.

97. MATTHEWS, *supra* note 11, at 29.

98. NOAM CHOMSKY, CURRENT ISSUES IN LINGUISTIC THEORY 54–55 (1964) ("It is necessary to distinguish between exceptions to grammar, and counter-examples to a proposed general theory of linguistic structure. . . . Examples that contradict the principles formulated in some general theory show that, to at least this extent, the theory is incorrect and needs revision. Such examples become important if they can be shown to have some bearing on alternative conceptions of linguistic structure.").

99. *See* Lakoff, *supra* note 1, at 960.

100. MATTHEWS, *supra* note 11, at 98. The definitions of "morpheme" and "phoneme" and their interrelationship are complex, endlessly debated, and far beyond the scope of this Note. *See generally, e.g.*, Andreas Koutsoudas, *The Morpheme Reconsidered*, 29 INT'L J. AM. LINGUISTICS 160 (1963); Zellig S. Harris, *From Phoneme to Morpheme*, 30 LANG. 190 (1955); Laurie Bauer, *What Is a Morpheme?*, *in* INTRODUCING LINGUISTIC MORPHOLOGY 110–21 (2d ed. 2003). For current purposes, it suffices to define a phoneme as "a unit of spoken language," and a morpheme as an abstract unit comprised of "morphs," or units that are "segment[s] of a word-form." William F. Twaddell, *On Defining the Phoneme*, 11 LANG. 5, 5 (1935); *see also* Bauer, *supra*.

grammaticality.[101] Chomsky's focus on syntax and mentalism, as opposed to language's external manifestations, thereby "shifted attention from . . . recorded [corpus] data . . . to the system of knowledge . . . underl[ying] the production and understanding of language, and, further, to the general theory of human language . . . behind this knowledge."[102]

In doing so, Chomsky "positioned his science directly in the midst of one of the central distinctions in Western philosophical history, and one most relevant to scientific inquiry and method: empiricism versus rationalism."[103] Whereas Bloomfield's conception of linguistics as a science was empirical,[104] Chomsky was a rationalist after Descartes.[105] Indeed, Chomsky aligned with "Aristotelian, a priori reasoning and opposition to experiment" as the only proper scientific way of approaching language, completely eschewing empirical methods.[106] It is then fascinating that Chomsky's theories dominated the linguistic field in the second half of the twenty-first century, given their departure from the general empirical trend in modern scientific thought.[107] Yet by aligning with a rationalist approach, Chomsky could "identify the 'transcendental' element [of language] . . . with the organic structure of man," as opposed to studying how language conditioned the speaker and the world.[108] In later years, Chomsky became more explicit about his interest in language as a gateway to the mind,[109] going so far as to characterize linguistics as "simply the subfield of psychology that deals with the aspects of [the] mind."[110] Chomsky grew even more radical with his "Cartesian rationalism";[111] he would eventually link language competence not only to psychology, but also to biology,[112] a view divorced from the general humanist conceptualization of language as a social phenomenon.

Whether or not Chomsky's theories revolutionized linguistics, they certainly initiated an intense focus on generating a robust theory that could dictate linguistics' objectives and derivative methodology. Rising above the morass of competing theories and skirmishes over various approaches and goals of the discipline, Chomsky's theoretical focus continues to haunt linguistics to this day.

---

101. MATTHEWS, *supra* note 11, at 23.
102. Campbell, *supra* note 41, at 111.
103. Golumbia, *supra* note 84, at 42.
104. ROBINS, *supra* note 36, at 242, 261.
105. *See generally* CHOMSKY, *supra* note 31.
106. Golumbia, *supra* note 84, at 42.
107. Formigari, *supra* note 42, at 2.
108. *Id.* at 5.
109. *See, e.g.*, NOAM CHOMSKY, THE MINIMALIST PROGRAM 4 (1995).
110. CHOMSKY, *supra* note 89, at 25.
111. *See generally* James S. Fulton, *The Cartesianism of Phenomenology*, 35 CONT'L PHIL. REV. 433 (2002).
112. *See generally* Noam Chomsky, *Of Minds and Language*, 1 BIOLINGUISTICS 9 (2007); Noam Chomsky, *Biolinguistic Explorations, Design, Development, Evolution*, 15 INT'L J. PHIL. STUD. 1 (2007); Noam Chomsky, *Universals of Human Nature*, 74 PSYCHOTHERAPY & PSYCHOSOMATICS 263 (2005).

E. THE RETURN TO EMPIRICISM, THE BREAKDOWN OF THE CHOMSKYAN REGIME, AND THE RISE OF FORENSIC LINGUISTICS

While Chomsky's ideas dominated linguistic intellectual consciousness well into the 1990s, they did not do so exclusively; other schools of linguistic thought carried on quietly and concurrently, and Chomsky's followers gradually fell away and were absorbed into other subdisciplines.[113] The first problems that began to manifest cracks in Chomsky's regime were not only his highly controversial theories over the genetic origins of language,[114] but also TGG's inability to comprehensively account for language variation.[115] The critiques flowed fast and strong; there were increasing doubts over whether a single grammar could truly account for all aspects of language, including speaker-relative concepts like intentions or assumptions.[116] Some of Chomsky's earliest and most enthusiastic students led the first break toward generative semantics, but this too soon devolved into various sub-approaches, including cognitive linguistics, itself "a flexible framework rather than a single theory of language."[117] "What began as a compact, in-house disagreement over a single hypothesis within Chomskyan linguistics [had] mushroomed into foundational proportions."[118] Linguists were suddenly rehashing fundamental issues Chomsky had supposedly put to rest, returning to old disputes over "the definition of the . . . field, the scope of language study, [and] the answer to the question, *What is linguistics?*"[119]

Robin Lakoff, one of Chomsky's closest and most devoted disciples, has attributed the split to TGG's inevitable failure to unite the mathematicians and logicians attracted to the theory's formality and scientific rigor with the humanists, who sought to discover language as a window into the mind.[120] Again, language variation was at the heart of the split. Chomsky's grammar simply could not account for more intricate constructions.[121] The formalist diehards rejected such variation in their continued search for the perfect theory, but the humanists embraced it,[122] leaving Chomsky to study pragmatic intention, discourse, and psychological influences on syntax.[123] To Lakoff, the breakdown represented a departure from "forc[ing] language into the Procrustean

113. Frederick J. Newmeyer, *Outside of the Palace Walls: Generative Linguists in the 1970s and 1980s*, 96 LANG. 173, 196 (2020).

114. RANDY ALLEN HARRIS, THE LINGUISTICS WARS 67 (1993).

115. Golumbia, *supra* note 84, at 46; Lakoff, *supra* note 1, at 960, 965, 981; JOSEPH, *supra* note 31, at 63.

116. FREDERICK J. NEWMEYER, GENERATIVE LINGUISTICS: A HISTORICAL PERSPECTIVE 118–19 (1996).

117. Dirk Geeraerts & Hubert Cuyckens, *Introducing Cognitive Linguistics*, *in* THE OXFORD HANDBOOK OF COGNITIVE LINGUISTICS 3, 4 (Dirk Geeraerts & Hubert Cuyckens eds., 2007).

118. HARRIS, *supra* note 114, at 9; *see also* JOSEPH, *supra* note 31, at 63.

119. HARRIS, *supra* note 114, at 7; *see also* Lakoff, *supra* note 1, at 946 ("It was not a theory-internal conflict . . . . Rather, the disagreement was about the subject-matter of the theory: what it should encompass, what language was.").

120. Lakoff, *supra* note 1, at 944–45.

121. *Id.* at 960, 965.

122. *Id.* at 946, 964.

123. *Id.* at 955–56.

bed of science."[124] In order to account for the "interconnectedness" of language, the mind, and the world,[125] linguists had to sacrifice the prized features of the sciences: "quantification, . . . falsification, [and] replication."[126] After a long stint as a science, linguistics was finally unmasked as having "the focus of the humanities but attempting the methods of the sciences."[127]

Indeed, today "many linguists . . . are willing to entertain the idea of abandoning the autonomous status of linguistics altogether, to wit Chomsky's proposal to develop the study of linguistic structure as a chapter of human psychology[,] . . . thus reversing the efforts of Saussure and generations of linguists before him."[128] Linguistics' autonomy has ebbed as linguists have reacted,[129] much like Lakoff, back toward a humanist orientation, feeding into the "contemporary rise of a large number of alternative approaches to language,"[130] including cognitive science,[131] pragmatics,[132] and corpus linguistics.[133]

But there were also other problems with changing times. Chomsky speculated that empirical research based on computation would not further "any significant advance in our understanding of the use or nature of language."[134] Within just a few decades, however, technology had advanced to the point where it was too tempting not to engage in the kind of empirical experimentation Chomsky repudiated. Moreover, as linguistics rose to prominence, there was greater interest in finding practical applications of linguistic methodology to solve real-world problems.[135] It was soon evident that linguistics' practical applications were at extreme tension with abstract conceptions of competence.[136] This practical interest was bolstered by other factors like a saturated academic job market and increasing numbers of linguists seeking

124. *Id.* at 985 (internal quotation marks omitted).

125. *Id.* at 984.

126. *Id.* at 985, 964.

127. *Id.* at 967.

128. Koerner, *supra* note 43, at 246–47 (internal citation and quotation marks omitted); *see also* ROBINS, *supra* note 36, at 269–70 ("All of this allows and even encourages the linguist to concentrate his attention on some one or other module, no longer looking at language as a unitary subject. . . . [S]pecialists will more and more direct their research and teaching to that part of the study of language which appeals most to them . . . ."); Koerner, *supra* note 32, at 14 ("[W]e are currently witnessing a diversity of views in matters concerning both general theory and the treatment of specific aspects of linguistic analysis . . . .").

129. ROBINS, *supra* note 36, at 151 ("The most obvious comment is that the discipline is no longer unified.").

130. JOSEPH, *supra* note 31, at 66.

131. *Id.* at 69.

132. "Pragmatics" can be defined as "the study of the relation between language forms and language function." LAKOFF, *supra* note 2, at 4.

133. *Id.* at 177; *see also supra* note 18.

134. CHOMSKY, *supra* note 89, at 4.

135. Widdowson, *supra* note 51, at 33–34.

136. *Id.* at 39–42.

employment in industry, pressuring the field to give more attention to application over theory.[137]

This helped shift linguists' attention from questions relating to language knowledge and acquisition to actual language use, transitioning linguistics away from exclusively studying speaker intuition to testing new theoretical hypotheses against computerized data.[138] Linguists moved from typical Chomskyan, intuition-centric linguistic evidence to look at performance data like speech errors and aphasic speech—in other words, to chiefly consider empirical data and variation.[139] And by tracking new experimental data like reaction-time measurement, eye tracking, and brain imaging, linguists increasingly turned to statistical models of language structure, acquisition, and use.[140]

II. LINGUISTICS AND AUTHORSHIP IDENTIFICATION IN THE COURTS

It is difficult to chart the historical development of the admissibility of forensic linguistic evidence, given the dearth of reported cases on its recent emergence and yet uncertain status in the courtroom. The first to coin the term "forensic linguistics" was linguistics professor Jan Svartvik in 1968, who was also among the first to advocate its use in court.[141] But this was not the first time courts encountered what later came to be known as forensic stylistic evidence, which originally cropped up in handwriting cases.

In 1901, the Supreme Court in *Throckmorton v. Holt* ruled that witnesses could not corroborate contested handwriting by relying on their familiarity with the style and composition of the alleged author.[142] Nevertheless, courts freely permitted corroborating comparisons of spelling and punctuation to determine the author of contested handwriting.[143] For example, in 1976, the Ninth Circuit in *United States v. Pheaster* permitted a comparison of spelling errors to support an opinion on the handwriting, reasoning that "[t]he manner of spelling a word is no less an 'identifying characteristic' than the manner of crossing a 't' or

137. Thomas Wasow, *Generative Grammar: Rule Systems for Describing Sentence Structure*, *in* THE HANDBOOK OF LINGUISTICS 119, 137 (Mark Aronoff & Janie Rees-Miller eds., 2d ed. 2017).

138. *Id.*

139. KASHER, *supra* note 85, at 95.

140. Wasow, *supra* note 137, at 119, 137.

141. *See generally* JAN SVARTVIK, THE EVANS STATEMENTS: A CASE FOR FORENSIC LINGUISTICS (1968).

142. 180 U.S. 552, 570 (1901).

143. *See, e.g.*, United States v. Van Wyk, 83 F. Supp. 2d 515, 523 (D.N.J. 2000) ("[C]ourts uniformly have admitted evidence of known writings, recognizing that the particular or peculiar use of grammar and spelling, for example, can be observed and identified to establish authorship."); Detroit Fire & Marine Ins. Co. v. Gagliardi, 32 P.2d 832, 838–39 (Colo. 1934) (comparing the spelling of certain words to corroborate a conclusion on handwriting); State v. Much, 287 P. 57, 61 (Wash. 1930) (similar); *see also* Carole E. Chaski, *Forensic Linguistics, Authorship Attribution, and Admissibility*, *in* FORENSIC SCIENCE AND LAW: INVESTIGATIVE APPLICATIONS IN CRIMINAL, CIVIL, AND FAMILY JUSTICE 505, 510–11 (Cyril H. Wecht & John T. Rago eds., 2006).

looping an 'o'[;] [a]ll may tend to identify a defendant as the author of a writing."[144]

Yet that same year, the United States District Court for the Northern District of California in *United States v. Hearst* excluded forensic linguistic expert testimony in a criminal case involving the disputed authorship of certain writings.[145] The court excluded the evidence principally on the finding that psycholinguistics had not yet achieved general acceptance in the scientific community. While the court made this determination under the old *Frye v. United States* test,[146] it also questioned the testimony's reliability in showing "a reasonable degree of certitude that . . . the defendant could not have authored the writings [at issue]."[147] The court also excluded the expert testimony on the basis that it was not relevant to the key issue of the defendant's state of mind, even if it supported the secondary issue of authorship.[148] Other courts have similarly excluded psycholinguistic testimony going to the speaker's intent.[149]

While courts have consistently rejected forensic psycholinguistic evidence, the admissibility of forensic stylistic evidence is less settled. In 1979, Roger Shuy sat next to a lawyer on a flight, leading Shuy to testify as an expert witness in his first case and eventually to become the "pioneer" of forensic linguistics.[150] By the 1990s, "a time of great expansion in the field of language and the law in the US,"[151] forensic linguistics had established its own academic organization, the International Association of Forensic Linguistics, and its own journal.[152] It was also during this time, however, that a "serious rift" developed in forensic authorship identification methodology, even as forensic linguistic evidence took off in trademark cases.[153] By 1997, the buzzy Unabomber case in the United States District Court for the Eastern District of California, a sister court of the court that rejected the psycholinguistic expert testimony in *Hearst*, thrust forensic linguistics into the national consciousness. The case has since been

144. 544 F.2d 353, 372 (9th Cir. 1976); *see also* United States v. Campbell, 732 F.2d 1017, 1021 (1st Cir. 1984) ("[S]pelling may be an identifying characteristic no less than handwriting idiosyncrasies." (emphasis omitted)).

145. 412 F. Supp. 893 (N.D. Cal. 1976).

146. 293 F. 1013 (D.C. Cir. 1923).

147. *Hearst*, 412 F. Supp. at 895.

148. *Id.*

149. *See, e.g.*, United States v. Kupau, 781 F.2d 740, 745 (9th Cir. 1986) (affirming the district court's exclusion of Shuy's expert linguistic testimony on the defendant's intent); United States v. Schmidt, 711 F.2d 595, 598–99 (5th Cir. 1983) (affirming the district court's exclusion of psycholinguistic expert testimony going toward the falsity of the appellant's statements); United States v. Washington, 20 F. Supp. 3d 777, 821–22 (W.D. Wash. 2006) (granting the motion to strike expert linguistic testimony on the ambiguity or intent of the speaker's language).

150. Jack Hitt, *Words on Trial*, THE NEW YORKER (July 16, 2012), https://www.newyorker.com/magazine/2012/07/23/words-on-trial.

151. Lawrence M. Solan, *Legal Linguistics in the US: Looking Back, Looking Ahead*, *in* LEGAL LINGUISTICS BEYOND BORDERS: LANGUAGE AND LAW IN A WORLD OF MEDIA, GLOBILISATION AND SOCIAL CONFLICTS 19, 20 (2019).

152. Shuy, *supra* note 11, at 627.

153. Solan, *supra* note 151, at 23.

made into a documentary and recently experienced some revived interest.[154] One of the FBI profilers working on the case, James Fitzgerald, later became the FBI's first trained forensic linguist.[155]

Although secondary sources indicate that Shuy and others testified as forensic stylistics experts in authorship identification cases,[156] there are nearly no reported cases, suggesting a "lack of prior judicial approval of this area of expertise."[157] In fact, in 2000, the court in *United States v. Van Wyk* noted that "no case law, treatise, or law review article has dealt with this precise issue of whether the examination of text analysis to resolve a litigated question related to disputed authorship . . . , or forensic stylistics, constitutes sufficiently reliable scientific evidence admissible in court."[158] The court then determined that because forensic stylistics, "much like handwriting analysis," lacks a known rate of error, recognized standards, meaningful peer review, and a system of accrediting experts in the field, [it] cannot "definitively establish[] . . . that a particular person is 'the' author of a particular writing."[159] Accordingly, the court permitted Fitzgerald to testify as to similarities and differences between the documents in question, but did not permit him to offer his opinion on their authorship.[160]

Around ten years later, in another authorship identification case involving Fitzgerald—who by this time had received a master's degree in linguistics—the court reached the same conclusion on the admissibility of forensic stylistic expert testimony, despite concluding that Fitzgerald now possessed the "requisite experience and education to qualify as an expert."[161] The defendants attacked the reliability of Fitzgerald's methodology, urging the court to exclude Fitzgerald's testimony because he failed to use scientific methods.[162] Citing *Kumho Tire Co. v. Carmichael*,[163] the court concluded that Fitzgerald's methods were valid because their reliability could be evaluated through "personal knowledge or experience rather than strict scientific methods."[164] Nevertheless, as in *Van Wyk*, the court permitted only Fitzgerald's testimony on similarities

154. *See Unabomber - in His Own Words* (Netflix, 2018); Alston Chase, *Harvard and the Making of the Unabomber*, THE ATLANTIC, June 1, 2000, at 41; Jake Hanrahan, *Inside the Unabomber's Odd and Furious Online Revival*, WIRED (Jan. 8, 2018, 7:00 AM), https://www.wired.co.uk/article/unabomber-netflix-tv-series-ted-kaczynski.

155. Chi Luu, *Fighting Words with the Unabomber*, JSTOR DAILY (Aug. 1, 2017), https://daily.jstor.org/fighting-words-unabomber/.

156. *See, e.g.*, Hitt, *supra* note 150; United States v. Van Wyk, 83 F. Supp. 2d 515, 520 n.6 (D.N.J. 2000).

157. *Van Wyk*, 83 F. Supp. 2d at 523.

158. *Id.* at 520.

159. *Id.* at 523.

160. *Id.* at 524.

161. United States v. Zajac, 748 F. Supp. 2d 1340, 1349 (D. Utah 2010).

162. *Id.* at 1351.

163. 526 U.S. 137 (1999).

164. *Zajac*, 748 F. Supp. 2d at 1351 (citing *Kumho Tire Co.*, 526 U.S. 137).

between the writings[165] and, by citing *Kumho Tire*, ostensibly discounted Fitzgerald's expertise as scientific.[166]

However, in 2020, a Florida district court in an unreported case denied the defendant's motion to exclude the linguistic expert testimony of Dr. Robert Leonard.[167] The court permitted Dr. Leonard not only to testify on the shared links between the two sets of documents in question, but also to offer his opinion that "the Q documents [were] consistent with the language patterns found in the documents known to have been written by [the defendant]."[168] Moreover, the court rejected the defendant's assertions that Leonard's methodology was unreliable under *Daubert v. Merrell Dow Pharmaceuticals, Inc.*,[169] highlighting that Leonard's methodology was peer-reviewed, "accepted and relied upon by federal courts in major cases," and "based on sufficient facts and data."[170] While the court noted the "robust debate among academics regarding best practices for authorial attribution," it determined that a "debate in the scientific community . . . is not a basis for exclusion," and therefore that any issues with Leonard's testimony should go to the weight of his opinion rather than to admissibility.[171] Observing that "[o]ther federal courts have rejected efforts to exclude expert testimony on the basis that the forensic stylistics approach is unreliable," the court denied the defendant's motion to exclude Leonard's opinion testimony.[172]

Despite the lack of definitive precedent, law enforcement investigators now commonly utilize forensic linguistic evidence to provide a profile of the speaker or writer,[173] while forensic linguists valiantly argue for greater inclusion in court.[174] Thus, "specialism in forensic linguistics . . . is [still] a relatively recent trend, with an increasing amount of research attention being paid to the validity

165. *Id.* at 1354.

166. *Id.* at 1351; *see also* United States v. Van Wyk, 83 F. Supp. 2d 513, 520 (D.N.J. 2000) ("Fitzgerald . . . arguably qualifies as an 'observational' expert . . . [and] qualifies as an expert in text analysis under the flexible requirements of *Daubert*, *Kumho*, and *Paoli II*.").

167. Kleiman v. Wright, No. 18-cv-80176, 2020 WL 6729362, at *32–34 (S.D. Fla. Nov. 16, 2020).

168. *Id.* at *32.

169. 509 U.S. 579 (1993).

170. *Kleiman*, 2020 WL 6729362, at *33.

171. *Id.* (internal citation and quotation marks omitted).

172. *Id.* at *34 (citing Dutcher v. Bold Films LP, No. 15-cv-110, 2019 WL 181353, at *1 (D. Utah Jan. 11, 2019)).

173. GIBBONS, *supra* note 25, at 307; *see also* Roger W. Shuy, *To Testify or Not to Testify?*, *in* LANGUAGE IN THE LEGAL PROCESS 3–18 (Janet Cotterill ed., 2002).

174. *See* Samuel Larner, *Forensic Linguistics*, *in* THE PALGRAVE HANDBOOK OF APPLIED LINGUISTICS RESEARCH METHODOLOGY 703, 705 (Aek Phakiti et al. eds., 2018) ("As such, there is now a strong research tradition surrounding the most appropriate ways to produce linguistic evidence . . . ."). *See generally, e.g.*, Robert A. Leonard, Juliane E.R. Ford & Tanya Karoli Christensen, *Forensic Linguistics: Applying the Science of Linguistics to Issues of the Law*, 45 HOFSTRA L. REV. 881 (2017); Coulthard, *supra* note 27; Ainsworth, *supra* note 15; Janet E. Ainsworth, *Linguistics as a Knowledge Domain in the Law*, 54 DRAKE L. REV. 651 (2006); Peter Tiersma & Lawrence M. Solan, *The Linguist on the Witness Stand: Forensic Linguistics in American Courts*, 78 LANG. 221 (2002); Lawrence M. Solan & Peter M. Tiersma, *Authorship Identification in American Courts*, 25 APPLIED LINGUISTICS 448 (2004).

and reliability of methods in authorship analysis."[175] Yet there remains a gaping divide between theoretical linguistics and forensic linguistics, and "it is not always possible to judge forensic testimony against ordinary practices among linguists, because linguists do not ordinarily engage in the activities that generate the expert testimony."[176] At the same time, new technology and the rise of the internet have led to a greater interest in computerized solutions to forensic authorship identification.[177]

III. CRITIQUING THE THEORY OF FORENSIC LINGUISTICS: IDIOLECT

The recent development of forensic linguistics tells the story of how forensic science has capitalized upon linguistics' loss of autonomy and failure to produce a unifying theory explaining language. Linguistics' status as a hard science is increasingly unsure, and the discipline has shifted back into the realm of the social sciences, evidenced by the rise of pragmatics, stylistics, and discourse analysis. But the history of forensic linguistics also illustrates "the real difficulty in applying the Galileian model[:] . . . [t]he more central [a]re features to do with the individual, the more impossible it bec[omes] to construct a body of rigorously scientific knowledge."[178]

The fact that forensic linguistics stems from the breakdown of the generativist theory seems almost natural because generative linguistics addressed the language-user in a most contradictory way. Chomsky argued that linguists should unlock the individual language-user's cognitive language faculties, while advocating abstract theories that fastidiously ignored variation and treated language strictly as a formal system.[179] The language-user was therefore at the heart of a paradox that depended on the individual for research but allowed him "no real effect on the foundations of linguistics. Though pretending to take him into account, linguistics ignores him."[180]

Applying linguistic methodologies in court further exacerbates the tensions already inherent in the discipline. On one hand, linguistic applications can only maintain their academic and scientific respectability by crafting robust

175. Larner, *supra* note 174.

176. Lawrence M. Solan, *Intuition Versus Algorithm: The Case of Forensic Authorship Attribution*, 21 J.L. & POL'Y 551, 561 (2013).

177. Rui Sousa-Silva, *Computational Forensic Linguistics: An Overview of Computational Applications in Forensic Contexts*, 5 LANG. & L. 118, 119 (2018).

178. Carlo Ginzburg & Anna Davin, *Morelli, Freud and Sherlock Holmes: Clues and Scientific Method*, 9 HIST. WORKSHOP J. 5, 19 (1980).

179. Raffaele Simone, *The Language User in Saussure (and After)*, *in* HISTORICAL ROOTS OF LINGUISTIC THEORIES 233, 234–35 (Lia Formigari & Daniele Gambarara eds., 1995).

180. *Id.* at 235; *see also* LAKOFF, *supra* note 2, at 8 ("Traditionally linguistics has been unwilling to consider the processes by which we understand larger and more abstract units of language (text or discourse) as a part of a speaker's knowledge of language . . . .").

theoretical solutions to more global questions about language.[181] On the other hand, practical linguistic applications can answer real-world problems, but fail to meet the requisite threshold of reliability by inherently focusing on variation.[182] Forensic linguistics seeks to tether its methods to a general theory of language[183] but takes these tensions to even greater extremes by using linguistic methods to appropriate a strictly forensic, individual-centric theory, idiolect.

A. THE THEORY OF IDIOLECT

Forensic linguistic methodology rests upon the theoretical assumption of "linguistic fingerprinting," or that each individual's speech has uniquely idiosyncratic, identifiable linguistic "impressions."[184] While the fingerprinting analogy may be an overstatement,[185] "[a]uthorship attribution . . . has always worked on the assumption that . . . one can discriminate texts by analyzing . . . language differences [that] would point to different authors," and that "an individual's language use can be determined and described by identifiable features."[186] Put differently, forensic linguistics rests upon the assumption that "language users have individual preferences and habits that determine their use of language."[187]

Idiolect backs the practice of linguistic experts who compare two sets of writing samples, an undisputed sample against the disputed sample, to identify any linguistic differences and similarities that might surface shared, idiosyncratic linguistic features unique to the author of the disputed document.[188] Linguistic experts might look, for example, to lexical density and richness, syntactic and morphological patterns, and deep structure to draw similarities between two documents that seem substantially unique.[189] Experts then measure these idiosyncratic features against interwriter differences[190] to gauge the uniqueness of the idiosyncrasy and evaluate whether it manifests the author's idiolect.[191] However, "[t]here is still no empirical method for demonstrating that each person has his or her own idiolectal variation that is

181. Simone, *supra* note 179, at 234 ("[Non-]user-centered theories are viewed as . . . formal, being able to use formalisms apt to make the procedures of analysis rigorous and full-fledged.").

182. *Id.* ("The main inconvenience with the user is that he intrudes a typical factor of disturbance, from which theoretical linguistics has always striven to stay off—variation.").

183. Carole E. Chaski, *Best Practices and Admissibility of Forensic Author Identification*, 21 J.L. & POL'Y 333, 342 (2013).

184. Coulthard, *supra* note 27, at 431–32 (emphasis omitted).

185. *Id.* at 432.

186. Eilika Fobbe, *Text-Linguistic Analysis in Forensic Authorship Attribution*, 9 J. LANG. & L. 93, 95 (2020).

187. Ainsworth, *supra* note 15, at 1165.

188. M. Teresa Turell & Núria Gavaldà, *Towards an Index of Idiolectal Similitude (or Distance) in Forensic Authorship Analysis*, 21 J.L. & POL'Y 495, 495–96 (2013).

189. *Id.* at 496–97.

190. *Id.* at 496.

191. SOLAN & TIERSMA, *supra* note 8, at 167.

uniquely identifiable, [and] author identification merely has to recognize intrawriter vs. interwriter variation strong enough to differentiate authors from each other."[192]

B. IDIOLECT, STATISTICS, AND THE CORPUS

There are powerful critiques of current approaches to forensic authorship attribution. While academic programs in forensic linguistics have increased over the years, these programs seldom offer sophisticated education in linguistics. "As a result, much of the linguistic training comes in the context of methods that might seem useful in forensic analysis . . . ."[193] Yet another problem is that there is still no established consensus on where to focus the idiosyncrasy inquiry,[194] whether by analyzing stylemarkers, vocabulary richness, the occurrence of function words, sentence length, word tokens, syntactic structure, or a combination of factors and approaches.[195] Forensic linguist Carole Chaski has suggested that "abstract syntactic structures can differentiate between authors and identify documents from one author," excluding potential authors from the pool based on failing statistical differentiation.[196] Chaski's statistical approach has been criticized, however, and it is dangerous for linguists to dabble in statistical testing without any formal training.[197] Similarly, corpus modeling requires a certain level of statistical expertise lacking in current corpus studies,[198] and stylometry has been attacked for misappropriating statistics and falling prey to the source fallacy.[199]

---

192. Chaski, *supra* note 183, at 348–49.

193. Solan, *supra* note 151, at 24.

194. *Id.* at 28.

195. *See* Chaski, *supra* note 183, at 349. *See generally, e.g.*, Frederick Mosteller & David L. Wallace, *Inference in an Authorship Problem: A Comparative Study of Discrimination Methods Applied to the Authorship of the Disputed* Federalist Papers, 58 J. AM. STAT. ASS'N 275 (1963); G. Udny Yule, *On Sentence-Length as a Statistical Characteristic of Style in Prose: With Application to Two Cases of Disputed Authorship*, 30 BIOMETRIKA 363 (1939); R. Harald Baayen & Antoinette Renouf, *Chronicling the Times: Productive Lexical Innovations in an English Newspaper*, 72 LANG. 69 (1996); Carole Chaski, *Who Wrote It? Steps Toward a Science of Authorship Identification*, 233 NAT'L INST. JUST. J. 15 (1997); E. Stamatatos, N. Fakotakis & G. Kokkinakis, *Computer-Based Authorship Attribution Without Lexical Measures*, 35 COMPUTS. & HUMANS. 193 (2001); Shlomo Argamon-Engelson, Moshe Koppel & Galit Avneri, Style-Based Text Categorization: What Newspaper Am I Reading? (1998) (unpublished manuscript), https://www.aaai.org/Papers/Workshops/1998/WS-98-05/WS98-05-001.pdf.

196. Chaski, *supra* note 143, at 516 (internal citation omitted).

197. Larner, *supra* note 174, at 711.

198. HANS LINDQUIST & MAGNUS LEVIN, CORPUS LINGUISTICS AND THE DESCRIPTION OF ENGLISH 37 (2018).

199. GIBBONS, *supra* note 25, at 303–04; Jonathan J. Koehler, *Linguistic Confusion in Court: Evidence from the Forensic Sciences*, 21 J.L. & POL'Y 515, 537 (2013); Joseph Rudman, *The State of Authorship Attribution Studies: Some Problems and Solutions*, 31 COMPUTS. & HUMANS. 351 (1998). Still other studies have produced results that undermine the idiolect hypothesis. *See, e.g.*, Max M. Louwerse, *Semantic Variation in Idiolect and Sociolect: Corpus Linguistic Evidence from Literary Texts*, 38 COMPUTS. & HUMANS. 207, 217 (2004) ("In sum, for some authors[] similarity in content can be found . . . [but] [f]or other authors texts differ within one author.").

Moreover, the data churned out by corpus linguistics studies can vary tremendously depending on research design, and current tools are still behind.[200] There are also bitter debates over what makes a truly representative corpus[201] and how large the corpus must be to achieve accurate results—particularly a problem in authorship attribution cases, where the initial dataset, such as a ransom note, is limited.[202] And does frequency really indicate individuality to the exclusion of all else,[203] in light of the haunting Chomskyan critique that a finite corpus can never represent all language use, making common constructions seem to occur less frequently in the chosen set than in actual use?[204] Lastly, there is the danger of confirmation bias. While intuition might indicate that identifying a sufficiently wide range of unusual linguistic features between two sets of documents would produce accurate results,[205] some studies have shown that applying a subset of selected features fares better,[206] meaning that "[w]ithout a predefined algorithm, an expert runs the significant risk of preferencing aspects that confirm her initial hypothesis over those that disprove it."[207]

C. IDIOLECT, HISTORY, AND THE COURTS

Idiolect and the focus on corpus-centered approaches reveal that forensic linguistics aligns more closely not with linguistics—whose history underscores the uneasy tensions in forensic linguistics' individualized focus—but with the work of literary and textual criticism identifying the authorship of the works of Homer, Shakespeare, and even the *Federalist Papers*.[208] Frederick Mosteller and David Wallace's study distinguishing the stylistic features of Madison and Hamilton's writing in the *Federalist Papers* illustrates the modern corpus linguistics techniques forensic linguistics seeks to exploit.[209]

But as the case of the Unabomber demonstrates, studies of literary or highly stylized writings may be inherently easy to identify even without turning to

200. Laurence Anthony, *A Critical Look at Software Tools in Corpus Linguistics*, 30 LINGUISTIC RSCH. 141, 144 (2013); Donald L. Drakeman, *Is Corpus Linguistics Better Than Flipping a Coin?* 109 GEO. L.J. ONLINE 81, 84 (2020).

201. TONY MCENERY & ANDREW HARDIE, CORPUS LINGUISTICS: METHOD, THEORY AND PRACTICE 2 (2012).

202. GRAEME KENNEDY, AN INTRODUCTION TO CORPUS LINGUISTICS 62, 66 (1998).

203. PAUL BAKER, SOCIOLINGUISTICS AND CORPUS LINGUISTICS 19 (2010) (Paul Kerswill & Joan Swann eds., 2020) ("*Frequency* is the bedrock of corpus linguistics.").

204. MCENERY & WILSON, *supra* note 18, at 75; *see also* DYLAN GLYNN & KERSTIN FISCHER, QUANTITATIVE METHODS IN COGNITIVE SEMANTICS: CORPUS-DRIVEN APPROACHES 11 (2010) ("There is, in reality, no such thing as a balanced corpus and no corpus can ever hope to be representative of a language.").

205. Solan & Tiersma, *supra* note 174, at 458–59.

206. *See, e.g.*, Farkhund Iqbal, Rachid Hadjidj, Benjamin C.M. Fung & Mourad Debbabi, *A Novel Approach of Mining Write-Prints for Authorship Attribution in E-Mail Forensics*, 5 DIGIT. INVESTIG. S42, S44 (2008).

207. Edward K. Cheng, *Being Pragmatic About Forensic Linguistics*, 21 J.L. & POL'Y 541, 550 (2013).

208. MALCOLM COULTHARD, ALISON JOHNSON & DAVID WRIGHT, AN INTRODUCTION TO FORENSIC LINGUISTICS: LANGUAGE IN EVIDENCE 152–53 (2d ed. 2017); Chaski, *supra* note 183, at 333 ("Forensic computational linguistics developed out of linguistic theory and computational linguistics.").

209. *See generally* Mosteller & Wallace, *supra* note 195.

rigorous statistical, linguistic methods. In fact, that is how such texts have been identified for centuries before linguistics was an autonomous discipline that could lend its methods to forensic authorship attribution. Similarities might therefore surface not because of idiolect, but because a highly stylized document like a manifesto or a ransom note may naturally, and perhaps even purposely, rely on a repeating set of linguistic features. Perceived similarities or differences may therefore reflect genre, topic selection, and style rather than an individualized authorial footprint.[210] Put still in other terms:

> [I]t remain[s] unclear how a linguistic element acquires its stylistic value[,] . . . reveal[ing] a lack of differentiation between language and style. . . . The problem with automated systems is not their method of feature extraction or evaluation but the underlying theory of style, which still cannot provide a sufficient explanation of how a feature's frequent occurrence in itself points to an individual author.[211]

On the intrawriter end, this problem persists when comparing a highly stylized piece of writing such as a ransom note with average samples of the suspect's writing. That certain linguistic features like word choice or syntactical phrasing are present in both samples may neither manifest idiolect nor conclusively identify a single authorial source. To the contrary, the comparison may be meaningless, because when the writings are so different, it may be just as likely that the match is wholly coincidental.[212] And if, as historian Carlo Ginzburg identifies, there are significant, qualitative differences between the contrived piece and the natural, minute details which best reflect individuality,[213] overlap may not say anything meaningful about a common source.

To return once more to the example of the Unabomber: Lakoff pointed out in her supporting affidavit that "when two people write about the same thing, the words they use are likely to overlap."[214] However, the Unabomber's unconventional usage of the idiom "eat one's cake and have it," while at first thought to be a shared idiosyncrasy between Kaczynski and the terrorist, was in fact occasionally used in the press.[215] Chaski thus admits that "there are still no empirical data supporting the essential claim of forensic stylistics that each individual has a cluster of stylemarkers which is different from everyone else's, actually unique to individuals."[216] She admits further that "even if one subscribes to the existence of individual style, it does not necessarily mean that

210. Chaski, *supra* note 183, at 338 ("Research that focuses on literary classics or edited newspaper articles may develop accurate methods, but these methods must be tested on forensically feasible data before they are borrowed across-the-board for forensic authorship identification.").

211. Fobbe, *supra* note 186, at 95–96.

212. *See* Hitt, *supra* note 150 ("[I]t can be tricky to compare different genres of even a single person's writing.").

213. Ginzburg & Davin, *supra* note 178, at 11.

214. SOLAN & TIERSMA, *supra* note 8, at 162.

215. *Id.* at 161–62.

216. Chaski, *supra* note 143, at 515.

it has to be always detectable or would always develop an individualizing function."[217]

It is significant, then, that forensic linguistics turns primarily to the theory and techniques of the recently disgraced forensic identification sciences rather than to theoretical linguistics. Like other forensic identification sciences, forensic linguistics often starts with drawing similarities from experience, the rudimentary method behind handwriting and fingerprint analysis.[218] Again, the case that launched forensic linguistics, the case of the Unabomber, represents the era in which forensic linguistic evidence was offered purely on an anecdotal, experiential basis.[219] Even now, while forensic linguistics claims to rely upon linguistic techniques, it often simply hides behind linguistics' name and academic status while turning to statistical modeling analogous to DNA matching.[220] Moreover, using corpus linguistic techniques to identify similarities and differences does not sterilize the process from relying on experience because experts must still choose markers by which to gauge uniqueness.

In this respect, critiques leveled at forensic linguistic evidence in the authorship identification context mirror those leveled at forensic handwriting expertise.[221] This may seem surprising, given that forensic linguistics is accompanied by a greater sense of legitimacy than recently discredited forensic graphology.[222] Yet as the case law indicates, "[f]orensic stylistics developed out of traditional forensic handwriting identification."[223] While linguistics far outdates the use of linguistic evidence in court, linguistic authorship identification is more forensic than we might think. Indeed, the "linguistic" theory of idiolect directly replicates the methodology of interwriter

217. Fobbe, *supra* note 186, at 98.

218. *See generally* Jonathan J. Koehler, *Fingerprint Error Rates and Proficiency Tests: What They Are and Why They Matter*, 59 HASTINGS L.J. 1077 (2008); Mnookin, *supra* note 26; Jennifer L. Mnookin, *Fingerprint Evidence in an Age of DNA Profiling*, 67 BROOK. L. REV. 13 (2001); Simon A. Cole, *Fingerprinting: The First Junk Science?*, 28 OKLA. CITY U. L. REV. 73 (2003); D. Michael Risinger & Michael J. Saks, *Science and Nonscience in the Courts:* Daubert *Meets Handwriting Identification Expertise*, 82 IOWA L. REV. 21 (1996); Michael J. Saks & Jonathan J. Koehler, *The Individualization Fallacy in Forensic Science Evidence*, 61 VAND. L. REV. 199 (2008); SVARTVIK, *supra* note 141.

219. Coulthard, *supra* note 27, at 433; SOLAN & TIERSMA, *supra* note 8, at 163.

220. *See* Dale A Nance, *Naturalized Epistemology and the Critique of Evidence Theory*, 87 VA. L. REV. 1551, 1610–12 (2001); Chaski, *supra* note 143, at 516–18; Shuy, *supra* note 11.

221. *See, e.g.*, United States v. Van Wyk, 83 F. Supp. 2d 515, 522 (D.N.J. 2000) ("Various judicial decisions regarding handwriting analysis, while not identical to text analysis, are instructive because handwriting analysis seems to suffer similar weaknesses in scientific reliability, namely the following: no known error rate, no professional or academic degrees in the field, no meaningful peer review, and no agreement as to how many exemplars are required to establish the probability of authorship.").

222. *See* Mnookin, *supra* note 26, at 1791.

223. Chaski, *supra* note 183, at 333 (citing GERALD R. MCMENAMIN, FORENSIC STYLISTICS 45–46 (1993)); *see also* Throckmorton v. Holt, 180 U.S. 552, 567–70 (1901) (holding witness opinion testimony on composition and style to corroborate familiarity with handwriting inadmissible). *But see* State v. Kent, 74 A. 389, 390–91 (Vt. 1909) (noting punctuation similarities and differences in a hand-carving comparison); United States v. Pheaster, 544 F.2d 353, 371–72 (9th Cir. 1976) (permitting a comparison of spelling errors between handwriting exemplars to identify the defendant).

and intrawriter similarities in handwriting analysis.[224] Thus, both history and current forensic linguistic identification techniques reveal that forensic linguistics was wholly constructed in court.[225]

But while the forensic identification sciences have generally fallen from grace,[226] with increasing interest in sophisticated computational techniques for matching text samples, reasons for excluding forensic linguistic evidence over nonexpert alternatives may appear less cogent. This in turn may lead judges to assess the evidence on the "sliding scale" of reliability Jennifer Mnookin suggests,[227] tending to support her contention that reliability fails to exist exogenous to the law and is instead constructed by it.[228] At the same time, *Daubert* continues to present a formidable hurdle to forensic linguistic evidence in authorship identification cases, which, when parading under the mantle of science, must produce reliable results through sufficient testing in the case at hand.[229] The new algorithmic model of forensic linguistic evidence thus faces the significant challenge of whether it can consistently produce superior results to the commonsense inferences of experience.[230] But even beyond that, as this Note demonstrates, the tangled history of the study of language challenges whether any amount of analysis or research can conclusively trace why and how people use language in the way that they do. Therefore, perhaps the best any language expert can do is merely point out similarities and differences.[231]

Robert Leonard, a professor at Hofstra University and mentee of Roger Shuy's,[232] is one of the most frequently appearing linguistic experts in recent

224. Mnookin, *supra* note 26, at 1805–06.

225. *Id.*

226. *See generally* Conor Friedersdorf, *CSI Is a Lie*, THE ATLANTIC (Apr. 20, 2015), https://www.theatlantic.com/politics/archive/2015/04/csi-is-a-lie/390897/; Michael J. Saks, *Merlin and Solomon: Lessons from the Law's Formative Encounters with Forensic Identification Science*, 49 HASTINGS L.J. 1069 (1998); EXEC. OFF. OF THE PRESIDENT'S COUNCIL OF ADVISORS ON SCI. & TECH., REPORT TO THE PRESIDENT: FORENSIC SCIENCE IN CRIMINAL COURTS: ENSURING SCIENTIFIC VALIDITY OF FEATURE-COMPARISON METHODS (2016), https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf.

227. Mnookin, *supra* note 26.

228. *Id.* at 1728–29.

229. 509 U.S. 579, 590–95 (1993); Solan, *supra* note 151, at 28–29.

230. Solan, *supra* note 151, at 29.

231. *See* Chaski, *supra* note 143, at 518; Koehler, *supra* note 199, at 539 ("Source claims . . . should be avoided. Such a modest approach will not only help forensic linguistics gain a place in the courtroom, but it will also reduce the risk that jurors will overweigh this potentially important, but as yet untested, evidence."). Indeed, this currently seems to be the prevailing approach taken by courts. *See e.g.*, People v. Coleman, 24 N.E.3d 373, 402 (Ill. App. Ct. 2014) (allowing the expert to testify on similarities without offering an opinion on the ultimate issue of authorship); United States v. Zajac, 748 F. Supp. 2d 1340, 1353 (D. Utah 2010) (same); United States v. Van Wyk, 83 F. Supp. 2d 515, 523–24 (D.N.J. 2000) (same). *But see* Kleiman v. Wright, No. 18-cv-80176, 2020 WL 6729362, at *32 (S.D. Fla. Nov. 16, 2020) (permitting the expert to offer an opinion that shared language patterns in a set of documents "[were] consistent with the language patterns found in the documents known to have been written by [the defendant]").

232. Hitt, *supra* note 150.

case law.[233] Leonard took the stand in 2012 in a case where the defendant, Brian Hummert, was charged with strangling his wife.[234] Among the evidence offered against Hummert were a string of letters and a death note left by an alleged stalker whom Hummert claimed had committed the murder.[235]

> As an expert witness, Leonard testified about Hummert's prose style, noting the rare use of . . . "ironic repetition" . . . [and] a linguistic habit that, Leonard testified, he had found nowhere else: a tendency to use contractions in negative statements . . . but not in positive ones. . . . The jury was out for forty-five minutes and returned a verdict of guilty.[236]

The jury verdict was affirmed on appeal,[237] and at the close of 2020, the Supreme Court of Pennsylvania unanimously and summarily denied Hummert's petition for review.[238] While other circumstantial evidence strongly linked Hummert to the murder, the intermediate appellate court's opinion noted that "[Hummert's] attempt to blame his wife's murder on a fictional stalker that he created demonstrated his consciousness of guilt."[239]

CONCLUSION

The history of forensic linguistics reveals that forensic linguistic evidence in the authorship identification context—as with the rest of the forensic identification sciences—suffers from the shared assumption of identifiable individual idiosyncrasy, for which there may be no real scientific basis.[240] While hiding behind linguistics' scientific status, forensic linguistics instead aligns itself with literary authorship attribution. But as linguistics' modern turn demonstrates, the discipline's scientific status may be tainted, and its history suggests that the questions forensic linguistics claims to answer unequivocally are just as contentious, unsettled, and unknown as ever. Forensic linguistics' "wishlist" is a widely adopted, predefined algorithm; a large, random sample

233. *See* D.H. Pace Co. v. Aaron Overhead Door Atlanta LLC, 526 F. Supp. 3d 1360, 1365 (N.D. Ga. 2021) (testifying on trademark genericness); Lodestar Anstalt v. Bacardi & Co., No. 16-cv-06411, 2019 WL 8105378, at *8 (C.D. Cal. July 3, 2019) (similar); PODS Enters., Inc. v. U-Haul Int'l, Inc., No. 12-cv-01479-T-27MAP, 2014 WL 12628664, at *1 (M.D. Fla. June 27, 2014) (same); Apple, Inc. v. Amazon.com Inc., No. 11-cv-1327, 2011 WL 2638191, at *3 (N.D. Cal. July 6, 2011) (same); Nat'l W. Life Ins. Co. v. W. Nat'l Life Ins. Co., No. A-09-CA-711, 2011 WL 840976, at *1 (W.D. Tex. Mar. 3, 2011) (similar); *Kleiman*, 2020 WL 6729362, at *32–34 (offering expert testimony in an authorship identification case); Ragbir v. United States, No. 17-1256, 2019 WL 9522419, at *29 (D.N.J. Jan. 25, 2019) (testifying on whether a written confession was a verbatim transcript); Babos v. Welch, No. 09-cv-908, 2017 WL 9673713, at *12 (N.D. Ohio, Oct. 27, 2017) (offering expert testimony in a voice identification case); United States v. Chaudhry, No. 06-771, 2007 WL 9725142, at *3 (D.N.J. July 30, 2007) (offering expert testimony on an IM-messaging dialogue).

234. Hitt, *supra* note 150.

235. *Id.*

236. *Id.*

237. Commonwealth v. Hummert, 237 A.3d 454, 454 (Pa. Super. Ct. 2020) (unpublished table decision).

238. Commonwealth v. Hummert, 240 A.3d 875 (Pa. 2020) (per curiam).

239. *Hummert*, 237 A.3d at 454.

240. Saks & Koehler, *supra* note 218, at 202.

of known exemplars; and a proven theory of language use.[241] But as history has shown, this is a wishlist the most brilliant minds of academic linguistics have sought and failed to attain for the field generally.

"In a social structure of ever-increasing complexity like that of advanced capitalism, befogged by ideological murk, any claim to systematic knowledge appears as a flight of foolish fancy."[242] The modern page of linguistic history seems to affirm this view, with its interdisciplinary dilution and failure to generate an all-encompassing theory of language. It may seem ironic that "scientific investigation accepts more uncertainty than does the legal process, since it is the legal system's assumption that scientific knowledge is crisp and factual that makes it attractive in the first place. Nonetheless, that is often the case."[243] Or perhaps what this brief historical survey reveals is that the applied sciences are simply far less reliable than we think:

> Our desire to idealize science runs, I fear, rather deep; we do not actually want science to be muddy, complex, pragmatic, methodologically imperfect and messy. When the science offered in court is all of these things, as it so often is, we therefore tend to blame the science itself, rather than our own unrealistic desires.[244]

The history of linguistics reveals what the discipline's forensic variant has failed to solve: the ongoing quandary of whether interpreting human communication is a scientific or humanistic enterprise, and whether "the methods and perspectives of those very different discover systems [can] be welded together into a harmonious whole that yields reliable results."[245] More simply and poignantly put, perhaps no theory can "reliably" account for what is the "strangeness, beauty, and import of human speech."[246]

241. Cheng, *supra* note 207, at 545.

242. Ginzburg & Davin, *supra* note 178, at 27.

243. Lawrence M. Solan, *The Forensic Linguist: The Expert Linguist Meets the Adversarial System*, *in* THE ROUTLEDGE HANDBOOK OF FORENSIC LINGUISTICS 349, 350 (Malcolm Coulthard et al. eds., 2d ed. 2021) (internal citation omitted).

244. Jennifer L. Mnookin, *Idealizing Science and Demonizing Experts: An Intellectual History of Expert Evidence*, 52 VILL. L. REV. 763, 767 (2007).

245. LAKOFF, *supra* note 2, at 10.

246. BLOOMFIELD, *supra* note 78, at vii.