Vector Semantics, William Empson, and the Study of Ambiguity

Michael Gavin

The aim of this essay is to introduce humanists to *vector semantics*, a subfield of computational linguistics that uses statistical measurements to model the meanings of words. Mathematicians and information theorists such as Warren Weaver and Claude Shannon and linguists like Zellig Harris and John Firth first theorized its basic methods in the 1940s and 1950s. For much of the later twentieth century, attention in linguistics turned away from statistical, data-based approaches, focusing instead on studies of syntax and grammar like those of Noam Chomsky. Since the 1990s, however, computational semantics has blossomed into a massive field, in large part because of the commercial demand for internet search engines and targeted advertising. The basic notion that informs this work is vector representation, which defines a word as a sequence of numbers that record how often it appears near other words. Much like social-network graphs that track relations among people, semantic models expose the many points of connection that words share in high-dimensional space. Across these connections, words can be shown to cluster into concepts or topics, but those concepts never form discrete entities; instead, vector-space models place words in a vast, interconnected space of meaning.

In this regard, I'll argue, vector semantics share a set of assumptions with literary critic William Empson, who devoted his career to explaining how poets played with words' many meanings. Words were, in his view, "compacted doctrines" that always carried their various senses as latent semantic potential.[1] Empson's method of close reading broke words into

1. See William Empson, *The Structure of Complex Words* (New York, 1951); hereafter abbreviated *SCW*.

Critical Inquiry 44 (Summer 2018)

their putatively constituent units of connotation, and vector semantics pushes this conceit to an extreme he would have found as baffling and exhilarating as his first reviewers found him.

Although the method used in this essay could be described as "algorithmic criticism" (to borrow Stephen Ramsay's phrase)[2] or "computationally assisted close reading" (mine),[3] I want to tread carefully while entering the muddle of contention that surrounds those phrases. I'll use instead a more precise technical term: *disambiguation*. To disambiguate is to identify multiple possible meanings and to differentiate among them. In literature, the study of ambiguity owes much to Empson's pioneering work. His *Seven Types of Ambiguity* (1930) and *The Structure of Complex Words* (1951) established disambiguation as a primary critical activity by reconceiving its purpose; rather than disambiguate words by discarding their irrelevant or contradictory meanings, Empson showed how those meanings could be held in productive interpretive juxtaposition. Disambiguation of this kind is so primary to literary criticism, in fact, that it's often conflated with close reading when scholars explicate irony, paradox, or ideology. In the field of computational linguistics, disambiguation's pedigree is just as illustrious, though largely unknown to humanists. Machine translation and information retrieval depend on training computers to discriminate among the ambiguities of natural language. The technical jargon for this is *word-sense disambiguation*—an accident of terminology, perhaps, but one that suggests an unexplored interdisciplinary region shared by literary criticism and natural language processing. Vector-space models can be used in much the same way that Empson used the dictionary, not to reduce words to a meaning in context, but to expose their many possible connotations. By taking disambiguation rather than close reading as my frame, I mean to bracket off debates internal to literary studies and highlight instead this area of overlap.

2. See Stephen Ramsay, *Reading Machines: Toward Algorithmic Criticism* (Chicago, 2011).
3. See Michael Gavin, "The Arithmetic of Concepts: A Response to Peter de Bolla," modelingliteraryhistory.org/2015/09/18/the-arithmetic-of-concepts-a-response-to-peter-de-bolla/

Michael Gavin is a professor of English at the University of South Carolina. He is the author of *The Invention of English Criticism, 1650–1760* (2015).
Research for this essay was supported by the Early Modern Digital Agendas institute, hosted in 2015 by the Folger Shakespeare Library and funded by the National Endowment for the Humanities. During that institute, the analogy between William Empson's criticism and my earliest attempts at semantic modeling were suggested to me by Mary Erica Zimmer.

However, there's another advantage in returning to Empson now, when humanities computing is just beginning to emerge as a new paradigm of criticism. I want to make close reading seem strange again, to acknowledge some of its now-forgotten embarrassments, and to recover something of its original crudity and automaticity. Though he was recognized as a leading voice of New Criticism from its inception, Empson remained a solitary and eccentric figure whose methods were subject to fierce objection. Professional critic and amateur linguist, poor historian, ambivalent disciple, controversialist, Orientalist, freethinker, dogmatist, atheist, and sexual deviant, William Empson is the patron saint of polysemy and the inspiration for this project. He worked through poems like a human language-processing machine. He read methodically, relentlessly—we might say, algorithmically. At the same time, few could keep pace with his playful and imaginative intelligence. His writings exemplify the exuberant potential of a tenacious adherence to critical process.

1. Critical Disambiguation: William Empson

Allow me to begin by reading Empson's *Seven Types of Ambiguity* through the eyes and mind of Elder Olson. A poet, critic, and teacher, Olson was perhaps the most Chicagoan member of the Chicago school. He was born in the city in 1909, earned his undergraduate degree from the University of Chicago in 1934 and his master's the next year, and defended his doctorate in 1938 before joining the faculty of the English Department in 1942.[4] As an undergraduate, Olson would almost certainly have studied with R. S. Crane, and in any case he became one of Crane's most vocal acolytes and therefore a dedicated opponent of New Criticism.

When Olson read *Seven Types of Ambiguity*, he paused at length over an example in the second chapter, where Empson quotes *Macbeth*. Planning murder, Macbeth wonders aloud, "If th' Assassination / Could trammel up the Consequence, and catch / With his surcease, Success; that but."[5] The verb "trammel" suggests an odd analogy that Empson offers as an example of his second type of ambiguity, "when two or more meanings are resolved into one."[6] Empson explains,

> *Consequence* means causal result, and the things to follow, though not causally connected, and, as in 'a person of consequence,' the di-

4. See Thomas E. Lucas, *Elder Olson* (New York, 1972), p. 11.

5. Quoted in Elder Olson, "William Empson, Contemporary Criticism, and Poetic Diction," in *Critics and Criticism: Ancient and Modern*, ed. R. S. Crane (Chicago, 1952), p. 49; hereafter abbreviated "W."

6. Empson, *Seven Types of Ambiguity* (New York, 1966), p. 48; hereafter abbreviated *S*.

> vinity that doth hedge a king. *Trammel* was a technical term used about netting birds, hobbling horses in some particular way, hooking up pots, levering, and running trolleys on rails. *Surcease* means completion, stopping proceedings in the middle of a lawsuit, or the overruling of a judgment; the word reminds you of 'surfeit' and 'decease,' as does *assassination* of hissing and 'assess' and, as in 'supersession,' through *sedere*, of knocking down the mighty from their seat. . . . The meanings cannot all be remembered at once, however often you read it. [*S*, p. 50]

Empson's key critical move here—the technique that would draw Olson's ire—is to conceive meaning, not as the intention of an author nor speaking character, but as a property of words themselves. Like a lexicographer gathering testimonies, Empson doesn't paraphrase what Macbeth means but what words mean, and their meanings are seemingly detached from the dramatic situation. Considered comprehensively and synthetically, meanings are quite literally unthinkable. "The meanings cannot all be remembered at once," Empson says, yet somehow they're resolved into one complex unit of ambiguity and beauty. Olson remarks,

> Such a passage as this needs only attentive reading to make manifest its utter absurdity; but then that very absurdity in a fashion protects it, and gains a certain credence for it; it is so absurd that we in a measure believe it, merely because we are loath to believe that anything could be so absurd. . . . We are actually being asked to believe that the speech actually *means* all these various things; that Macbeth, trembling on the brink of murder, and restrained only by his fears of what may follow, is babbling of bird-nets, pothooks, levers, trolleys, assessments, lawsuits, and what not; and all this on the shadowy grounds that the OED, or whatever dictionary, lists alternative meanings for "trammel," "surcease," and "assassination," and that poetic language is ambiguous.[7] ["W," p. 49]

For Olson, Empson's perverse method of disambiguation is an intolerable kind of deformation that reduces language to nonsense. To tease out words' various meanings without discarding the false ones unravels the grammatical and rhetorical structure of communication; a sentence that actually made explicit all of its possible associations would be an absurd, babbling sentence. Clearly, William Shakespeare's sentences aren't absurd,

7. Richard Strier defends Empson against Olson's critiques in *Resistant Structures: Particularity, Radicalism, and Renaissance Texts* (Berkeley, 1995), pp. 14–15.

so Empson's must be. Olson also hints at two criticisms that he develops in greater detail elsewhere. First, that Empson is constrained by dogmatic adherence to a single principle ("poetry is ambiguous"), which he applies to all poetry, regardless of form. Second, that he relies too heavily on an untrustworthy model of language ("the OED, or whatever dictionary") and improperly uses it as a proxy for meaning in the poem.

Though written with a flair for the obtuse, Olson's criticisms actually capture well many of the interpretive dynamics that informed *Seven Types of Ambiguity* and, later, *The Structure of Complex Words*. The most important involves Empson's use of the *Oxford English Dictionary* (*OED*). Olson notes, "its presence everywhere is neither invisible nor subtle" ("W," p. 48). And Empson was indeed quite open about his reliance on the dictionary, remarking in *The Structure of Complex Words* that his work "has been almost entirely dependent on using the majestic object" (*SCW*, p. 391). The dictionary intervenes in Empson's reading process by providing him with a model of the lexical universe poems occupy. By itself, there's nothing much unusual about looking up words in the dictionary. Empson's procedure, however, was novel. Rather than immediately winnow down to the appropriate meaning of a word, Empson pauses over the various paths of meaning each entry offers, using its historical apparatus to recover all possibly relevant connotations.[8] Thus, words like *trammel* imply as much about fishing nets as they do about wrapping corpses. Even more troubling is Empson's automaticity. He doesn't just look up "'trammel,'" but consults the *OED* for "'surcease,'" "'assassination,'" and "'consequence'" as well. It's a rigid procedure, as if Empson simply followed a set of self-imposed instructions for analyzing each poem as a bag of words.[9] Disam-

8. Although Empson's *Seven Types of Ambiguity* does not have a reputation as a work of historicist criticism, his use of the *OED* is an exercise in historical semantics. As Donna Lee Berg has emphasized, the *OED* "is a diachronic dictionary; that is, it not only defines the language of the present day, it also records its use at any period within the Dictionary's coverage" (Donna Lee Berg, *A Guide to the Oxford English Dictionary* [New York, 1993], p. 3).

9. This description may not be far from the mark. According to legend, the bulk of the manuscript was written in a weeklong burst of enthusiasm while Empson was a student in Cambridge under I. A. Richards. Richards recounted the story:

> At about his third visit he brought up the games of interpretation which Laura Riding and Robert Graves had been playing with the unpunctuated form of "The expense of spirit in a waste of shame." Taking a sonnet as a conjurer takes his hat, he produced an endless swarm of lively rabbits from it and ended by "You could do that with any poetry, couldn't you?" This was a Godsend to a Director of Studies, so I said, "You'd better go off and do it, hadn't you?" A week later he said he was still slapping away at it on his typewriter. Would I mind if he just went on with that? Not a bit. The following week there he was with a thick wad of very illegible typescript under his arm—the central 30,000 words or so of the book. [Quoted in John Haffenden, *Among the Mandarins*, vol. 1 of *William Empson* (New York, 2008), p. 207]

biguation can't really be equated to close reading because so much of its emphasis goes beyond the poem itself. One might say Empson doesn't use the dictionary to read poems but uses poems to read the dictionary. Olson found it distasteful, calling it a "mechanical method . . . capable of all the mindless brutality of a machine" ("W," p. 48).

The epithet "mindless" recalls Empson's uncomfortable associations with his mentor I. A. Richards and, later, the New Critics.[10] In *Principles of Literary Criticism* and *The Meaning of Meaning*, Richards emphasized poetry's fundamental continuity with other forms of human activity: to read a poem is to have a sensory and emotional experience. Richards focused less on literary history than on the psychology of reception, allowing him to disregard authorial intention in most cases and rely instead on imagining the text as a stimulus of mental action. Empson was always sympathetic to this framework (though he resisted many of its details) because the muddled minds of readers provided him with an apt justification for his basic technique. Ambiguities are most intuitively grasped as a latent potential in the reading mind. "I . . . shall think relevant to my subject," Empson wrote, "any verbal nuance, however slight, which gives room for alternative reactions to the same piece of language" (*S*, p. 1). When analyzing individual poems, he often suggested a range of possible reactions readers might have, and he freely referred to the images or ideas that a passage might call to mind. Ambiguities represent cognitive possibilities that needn't even be latent in the unconscious—they are merely potential. Call it subjunctivist phenomenology. Loose speculation about what readers might think provided Empson with an axiomatic shorthand for his semantics. At the same time, though, the main thrust of his analyses were relentlessly textual and indifferent to psychology or history. He rejected the basic tenets of New Criticism on principle, but so little of his work would need to be thrown out by the intentional fallacy or the affective fallacy that it's no wonder he became associated with writers like Cleanth Brooks and William Wimsatt. Therefore, although his thinking remained powerfully informed by Richards, Empson's criticism in prac-

For a detailed analysis of similarities between Empson's work and Laura Riding and Robert Graves's *A Survey of Modernist Poetry* (1927), see Donald J. Childs's *The Birth of New Criticism: Conflict and Conciliation in the Early Work of William Empson, I. A. Richards, Robert Graves, and Laura Riding* (Montreal, 2013).

10. For an account of Richards's psychology and its relation to "mindless" forms of interpretation that discount authorial intentionality, see Joshua Gang, "Behaviorism and the Beginnings of Close Reading," *ELH* 78 (Spring 2011): 1–25 and "Mindless Modernism," *Novel* 46 (Spring 2013): 116–32.

tice always stumbled near an inhuman, routinized, even mindless style of reading.

Empson undoubtedly would object to such a characterization, but mindlessness is an apt description for the theory of meaning outlined in his longest and most rigorously argued work of semantics, *The Structure of Complex Words*, in which he argues that words are compound objects that contain, not only their different senses, but normative and emotive propositions about the interactions among those senses. "A word may become a sort of solid entity, able to direct opinion, thought of as like a person; also it is often said (whether this is the same idea or not) that a word can become a 'compacted doctrine', or even that all words are compacted doctrines inherently" (*SCW*, p. 39).[11] To study ambiguity, then, doesn't stop with the elucidation of meanings but with an account of those meanings' various interrelations. Empson glimpsed this problem early on, commenting in a 1930 letter to Richards that "words get their usefulness, not from a mere ambiguity, but from being like an unattainable limit, or like a continuous scale that can be pivoted at any required point."[12] Describing prepositions in *Seven Types of Ambiguity*, Empson says they have "not so much a number of meanings as a body of meaning continuous in several dimensions; a tool-like quality, at once thin, easy to the hand, and weighty, which a mere statement of their variety does not convey. In a sense all words have a body of this sort; none can be reduced to a finite number of points, and if they could the points could not be conveyed by words" (*S*, p. 5). The playful exuberance of Empson's method tries to capture something of words' malleability and extensibility. Meanings aren't discrete things, even though they inhere to words; instead they unfold over many dimensions of continuously scaled variation. Words have bodies and agency, Empson argues. Even a sort of personhood. They occupy an invisible lexical "thoughtspace" where they break apart and recombine to form superstructures, molding opinions and otherwise forging human experience.

In the next section, I'll follow ambiguity along an orthogonal trajectory through the field of computational linguistics, and I'll argue that key aspects of Empson's technique can, in fact, be automated. This automation does not provide convenience but serves more important intellectual pur-

11. Alan Durant and Colin MacCabe summarize Empson's views: "Words accumulate strata of senses and implications and assert propositions or arguments, even as they conceal such complexities by appealing to common-sense understanding" (Alan Durant and Colin MacCabe, "Compacted Doctrines: Empson and the Meanings of Words," in *William Empson: The Critical Achievement*, ed. Christopher Norris and Nigel Mapp [New York, 1993], p. 173).

12. Empson, letter to I. A. Richards, 29 Dec. 1930, in *Selected Letters of William Empson*, ed. Haffenden (New York, 2006), p. 34.

poses: it heightens the visibility of words' many polysemous associations; it further sensitizes those associations to historical nuance; and it places the entire method on altogether firmer theoretical ground. By scattering poetic statements over the lexical space of the dictionary, Empson showed that the capacity for words to be broken up into pieces—their disassemblability—is what makes possible their meaningful recombination. Disambiguation creates meaning by semantically disaggregating and then resuturing linguistic parts, often with a perverse disregard for the subjective experience of communication. Words are compacted entities made up of innumerable points that together form poetry's conceptual substrate, which by itself defies encapsulation through words alone. What Empson lacked was a mathematical theory appropriate to his intuitions.

2. Critical Disambiguation: Machine Translation to Vector Space

When Empson published *The Structure of Complex Words*, he put forward ideas about meaning that he believed were "not known to professional linguists."[13] Semantics had existed as a recognized field of inquiry for half of a century, and inquiries into the nature and process of meaning was active across several disciplines.[14] As already noted, Richards had published on the topic extensively.[15] Nonetheless, Empson had some good rea-

13. Empson, letter to Karunakar Jha, 20 June 1971, in *Selected Letters of William Empson*, p. 510.

14. Michel Bréal's *Essai de Sémantique* (1897) is often cited as defining the field; see Michel Bréal, *Semantics: Studies in the Science of Meaning*, trans. Mrs. Henry Crust (New York, 1900). One important strand of research into semantics that I cannot do more than acknowledge here can be found in analytic philosophy. Alexius Meinong's "Über Gegenstandstheorie" (1904) provoked reply by Bertrand Russell the next year in "On Denoting," which launched decades of debate about the relationships among mental states, language, and reality; see Alexius Meinong "The Theory of Objects," in *Realism and the Background of Phenomenology*, trans. Isaac Levi, D.B. Terrell, and Roderick Chisholm, ed. Roderick Chisholm (Glencoe, Ill., 1960), pp. 76–117. For a critical heritage of their debate, see *Russell vs. Meinong: The Legacy of "On Denoting,"* ed. Nicholas Griffin and Dale Jacquette (New York, 2009). One path connecting this intellectual tradition to computational linguistics follows Rudolph Carnap and Fred Dretske, who both wrote on information theory following Claude Shannon. Another path follows Russell through his student, Ludwig Wittgenstein, who in turn taught Margaret Masterman, who became a leading figure in machine translation and information retrieval.

15. In *Mencius on the Mind: An Experiment in Multiple Definition* (London, 1932), Richards expresses the desire to develop "a generalized technique by which meanings of all kinds of all necessary occasions can be systematically displayed" (p. 28). Writing in *Meaning of Meaning*, Branislaw Malinowski argues that "a word without linguistic context is a mere figment and stands for nothing by itself, so in the reality of a spoken living tongue, the utterance has no meaning except in the *context of situation*." (Branislaw Malinowski, "The Problem of Meaning in Primitive Languages," in *The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism*, ed. C. K. Ogden and Richards (New York, 1923), p. 307. Understanding meaning to be constrained or informed by

son to think of himself as an outsider to the field.[16] In the same year, Harris published *Methods in Structural Linguistics*, proposing a systematic disregard for meaning altogether. In meaning's place, Harris advocated for statistical measurements of linguistic distribution; words like *Albuquerque* and *applesauce* will appear in different contexts, and that (measurable!) difference, Harris argued, was a more appropriate object of study than subjective judgments about what words seem to denote. Harris was therefore among the leaders of a mid-century effort to secure linguistics on mathematically and scientifically sound principles. George Zipf's *The Psycho-Biology of Language* (1935) and Shannon and Weaver's *Mathematical Theory of Communication* (1949) sparked enthusiasm for statistical approaches to language, suggesting that trustworthy mathematical models of human communication and culture were not just possible but within close reach. This belief gained further momentum in the immediate wake of World War II, especially for American and British code breakers, whose success proved that human language could be subject to machine-based analysis and manipulation.

Among the American cryptographers was Weaver, director of the Rockefeller Foundation Natural Sciences Division, who proposed in 1949 a set of computational techniques for translating among natural languages. The futuristic promise of machine translation captured the imagination of scholars like Harris, and, over the next ten or fifteen years, machine translation became a major research field that swallowed up millions of dollars in federal grant funding. At the University of Pennsylvania, Harris's group

context goes back at least to Bréal, who, in a chapter titled "Polysemia," noted that "words are placed in surroundings which predetermine their import" (Bréal, *Semantics*, p. 141). The question of *how* contextual situations determined meaning remained largely open. Leonard Bloomfield attempts to get around this problem by defining meaning strictly as the situation of the utterance and the "response which it calls forth in the reader," but he admits that the "statement of meanings is . . . the weak point in language-study" (Leonard Bloomfield, *Language* [New York, 1933], pp. 139, 140).

16. Empson's sense of isolation from linguistics is echoed by Stephen Ullmann in *Principles of Semantics* (1951), first published the same year as Empson's *The Structure of Complex Words*. Ullmann, a linguist, presents his research into meaning—which he defines as a "reciprocal relation between name and sense"—as an effort to introduce philology and linguistics into a field of inquiry dominated by critics like Richards and philosophers like Russell (Stephen Ullmann, *Principles of Semantics* [New York, 1957], p. 70). Ullmann's work, particularly in his emphasis on semantic fields and the structure of concepts, points to another strand of intellectual work against which computational semantics can be usefully read: the theory of concepts and conceptuality. A recent book that brilliantly combines concept theory with computational semantics (while stubbornly refusing to acknowledge either tradition) is Peter de Bolla's *The Architecture of Concepts: the Historical Formation of Human Rights* (New York, 2013). Distributional semantics and the history of concepts are compared in Gavin et al., "The Spaces of Meaning: Conceptual History, Vector Semantics, and Close Reading," in *Debates in the Digital Humanities 2018*, ed. Matthew Gold and Lauren Klein (forthcoming).

tried to bring his linguistic theories to bear on the project, and similar efforts were underway by Anthony Oettinger at Harvard, Erwin Reifler at the University of Washington, Margaret Masterman in Cambridge, and Yehoshua Bar-Hillel, Victor Yngve, and Chomsky at the Massachusetts Institute of Technology (MIT). Many believed early on that Fully Automatic High Quality Translation—or FAHQT, as it was sometimes abbreviated—was a reasonable goal; soon, machines would not only translate scientific papers (the immediate aim) but also political communication and literature, even poems. Automatic translation would promote international communication and cooperation and thus help usher in a new, peaceful world order.[17] This vision was not realized, however. A report by the Automatic Language Processing Advisory Committee (ALPAC) in 1966 recommended that the funding stream be cut, and in 1968 Chomsky declared the entire project a misguided failure.[18]

With the benefit of hindsight, machine translation looks much more successful now than it did in the late 1960s. Modern internet-based translation engines work better than most people then believed possible, and these engines, much like the field of computational linguistics in general, trace their roots back to postwar efforts. More to our purposes, though, is that this research focused squarely on problems of meaning and ambiguity. Scholars who hoped to make automatic translation a reality faced an obvious obstacle—many words have more than one meaning, so translating among languages would involve, first and foremost, teaching machines to disambiguate words. As Masterman wrote in 1960, "The basic problem in Machine Translation is that of multiple meaning, or *polysemy*."[19]

One way to describe the origins and development of computational semantics is to trace it back to a tension in Harris's linguistic philosophy. When he proposed a structuralist disregard for meaning, Harris suggested a method for breaking language down to primitive simples and evaluating patterns across them. Like any "regularities in selected aspects of human

17. See Warren Weaver, "Machine Translation," in *Readings in Machine Translation*, ed. Sergei Nirenburg, Harold Somers, and Yorick Wilks (Cambridge, Mass., 2003), pp. 13–18, and Erwin Reifler, "The Mechanical Determination of Meaning," in *Readings in Machine Translation*, pp. 21–36.

18. A succinct and accessible history of machine translation can be found in Brian Lennon, "Machine Translation: A Tale of Two Cultures," in *A Companion to Translation Studies*, ed. Sandra Bermann and Catherine Porter (Malden, Mass., 2014), pp. 135–46. Lennon draws primarily from the work of John Hutchins, who wrote *Machine Translation: Past, Present, Future* (Chichester, 1986) and edited the important collection *Early Years in Machine Translation: Memoirs and Biographies of Pioneers* (Amsterdam, 2000).

19. Margaret Masterman, "Mechanical Pidgin Translation," in *Readings in Machine Translation*, p. 177.

behavior," he wrote, language could be studied by "associating discrete elements with particular features or portions of continuous events, and then stating the interrelations among these elements."[20] To study language, Harris argued, linguists should break the continuous field of talk into small parts and then describe the relations that pertain among those parts. This approach suggested two very different lines of inquiry, both of which Harris pursued. On the one hand, his theory suggested that sentences could be broken down into semantic "kernels"—simple, conceptual primitives—which were transformed by deep structural processes into human statements.[21] Harris's student Chomsky extended this line of inquiry into a complex but theoretically finite set of transformational rules through which a theoretically infinite number of statements might be made.[22] This grammatical, rule-based method investigated language through very small, hypothetical sentences to examine how they are or aren't meaningful.[23]

The other side of Harris's research agenda—and the work for which he is now most well known—involved statistical measurements of corpora. In 1954, Harris formulated what has since come to be known as the "distributional" hypothesis.[24] Rather than rely on subjective judgments or dictionary definitions, linguists should model the meanings of words with statistics because "difference of meaning correlates with difference of distribution." He explains,

> If we consider *oculist* and *eye doctor* we find that, as our corpus of actually occurring utterances grows, these two occur in almost the same environments. . . . In contrast, there are many sentence environments in which *oculist* occurs but *lawyer* does not; e.g., *I've had my eyes examined by the same oculist for twenty years*, or *Oculists often have their*

20. Zellig S. Harris, *Methods in Structural Linguistics* (1951; Chicago, 1963), p. 22.

21. According to Harris, "The transformations operating on the kernels yield the sentences of the language, either by modifying the kernel sentences of a given set (with the same modification for all kernels in the set) or by combining them (in fixed ways) with other kernel sentences" (ibid., pp. vi–vii).

22. Chomsky first developed these views in *Syntactic Structures* (1957), written while employed by MIT on its machine-translation initiative. Yehoshua Bar-Hillel compares Harris's and Chomsky's approaches in "The Present Status of Automatic Translation of Languages," in *Readings in Machine Translation*, pp. 50–51.

23. Indeed, computational linguists tended to work with extremely small corpora during most of the later twentieth century. Yorick Wilks, Brian M. Santor, and Louise M. Guthrie report on a conference from the early 1990s in which researchers in natural language processing admitted that their systems contained, on average, thirty-six words. See Wilks, Brian M. Santor, and Louise M. Guthrie, *Electric Words: Dictionaries, Computers, and Meanings* (Cambridge, Mass., 1996), p. 2.

24. See Harris, "Distributional Structure," in *The Structure of Language: Readings in the Philosophy of Language*, ed. Jerry A. Fodor and Jerrold J. Katz (Englewood Cliffs, N.J., 1964), pp. 33–49.

> *prescription blanks printed for them by opticians*. It is not a question of whether the above sentence with *lawyer* substituted is true or not; it might be true in some situation. It is rather a question of the relative frequency of such environments with *oculist* and with *lawyer*.[25]

This passage reflects a subtle shift in Harris's thinking. In *Structural Linguistics* he had proposed that linguists attend to distributional structure as a theoretical principle. Here the emphasis is slightly different. He says that distribution "correlates" to meaning. *Oculist* and *eye doctor* mean pretty much the same thing, and they appear in sentences with pretty much the same words. Rather than turn away from meaning, Harris here offers a statistical proxy for it. His theory was picked up and expanded upon by Firth, whose 1957 comment, "You shall know a word by the company it keeps," has become a truism of computational semantics.[26] The apparent simplicity of the distributional hypothesis fanned enthusiasm for machine translation, as Weaver and Erwin Reifler had proposed techniques that shared these assumptions for disambiguating polysemous words. Theory and engineering seemed to align.

However, statistical approaches to word meaning faced at least three major obstacles. The first was theoretical. Bar-Hillel objected on principle, claiming that many statements depend for their legibility on contextual knowledge that only conscious minds can hold. Giving the example of a child playing with a toy box while in a playpen, Bar-Hillel suggested that a sentence like "'The box was in the pen'" could never be translated by machine.[27] No statistical profile of the word *pen* could help a machine differentiate between possibly relevant meanings. The second problem involved a lack of data. To be reliable, every word would need to be included many times in all of its variations, and no such corpus existed.[28] Third, the limits of computer memory kept the datasets small. Machines were still operat-

25. Ibid., p. 43.

26. J. R. Firth, "A Synopsis of Linguistic Theory, 1930–1955," in *Studies in Linguistic Analysis* (Oxford, 1962), p. 11.

27. Bar-Hillel, "The Present Status of Automatic Translation of Languages," p. 74. For what it's worth, Google Translate manages the sentence just fine because increased computing power allows it to compile statistical profiles not just of words but of whole phrases.

28. The problem of data sparseness is a serious one for corpus-based word-sense discrimination, as Nancy Ide and Jean Véronis argue in "Introduction to the Special Issue on Word Sense Disambiguation: The State of the Art," *Computational Linguistics* 24 (Mar. 1998): 17. They point out that the Brown Corpus of one million words, released in 1967, contains only eight instances of the word *ash*, and that common meanings of the word are underrepresented or excluded altogether. For researchers today, the easy availability of massive online sources, like Wikipedia (2.9 billion English words), lessens this problem significantly, but it remains an issue for many humanists, whose datasets are often much smaller.

ing on punch cards, and researchers often resorted to simulating decision-tree algorithms by hand.[29] These problems were all insurmountable at the time, and so it wasn't until the late 1980s and 1990s that corpus-driven research became popular in linguistics.[30]

One way proposed to get around these problems involved developing semantic dictionaries that organized words under larger headings, much like a thesaurus. At Cambridge University, Masterman was a leading proponent of this technique. A philosopher and former student of Ludwig Wittgenstein, Masterman founded the Cambridge Language Research Unit (CLRU), which was an important center for the study of computational semantics through the 1980s. Along with Karen Spärck Jones, Masterman developed the first computer-based thesaurus, drawn from *Roget's*, for modeling word meaning; words, Masterman believed, distributed meaning through a corpus in a lattice-shaped network (fig. 1).[31] On the bottom of the lattice was the language as a whole, from which words were selected to make statements. Words used in a sentence (*W*, *Y*, *X*, *Y*, and *Z*) corresponded up the network to their various locations in the thesaurus, where co-occurring words were themselves organized into larger groups, which she called archeheads (*A*, *B*, *C*, and *D*), and which together formed an abstract conceptual superstructure that divided the language into headings, much like a subject index in a library catalog. With researchers at the CLRU, Masterman designed algorithms for reading over the lines of the thesaurus's semantic network to paraphrase English sentences and convert them into Latin and vice versa.[32]

During the 1970s and 1980s, after the promise of machine translation had seemed to falter and while corpus-based linguistics was relatively in-

29. Victor H. Yngve recounts such an experiment in "Early Research at M.I.T.: In Search of Adequate Theory," in *Early Years in Machine Translation: Memoirs and Biographies of Pioneers*, ed. W. John Hutchins (Amsterdam, 2000), p. 44.

30. For contemporary accounts of this development, see Nancy Ide and Donald Walker, "Introduction: Common Methodologies in Humanities Computing and Computational Linguistics," *Computers and the Humanities* 26 (Dec. 1992): 327–30, and Kenneth W. Church and Robert L. Mercer, "Introduction to the Special Issue on Computational Linguistics Using Large Corpora," *Computational Linguistics* 19 (Mar. 1993): 1–24. For a retrospective overview, see Tony McEnery and Andrew Hardie, *Corpus Linguistics: Method, Theory, and Practice* (Cambridge, 2012).

31. This view is echoed, without the algebraic formalization, by Werner Hüllen: "The essential gain to be had from thesauri is, therefore, the awareness that the lexis of a language is a huge semantic web in which every unit (word) is dependent on many others, or—carrying the idea to an extreme—where each unit is determined by every other one. The lexis of a language has no central point and no periphery, it is an endless functioning of interdependencies" (Werner Hüllen, *Networks and Knowledge in "Roget's Thesaurus"* [New York, 2009], pp. 140–41).

32. See Masterman, *Language, Cohesion, Form*, ed. Wilks (New York, 2005).

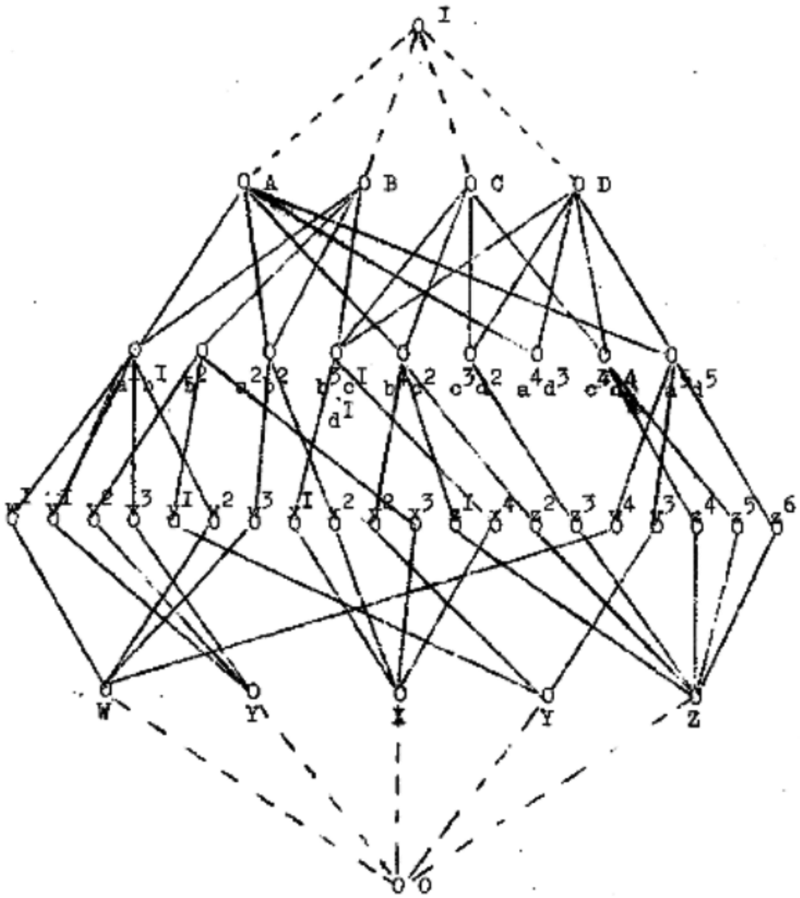

FIGURE 1. The lattice-shaped semantic network. From Masterman, "What is a Thesaurus?"

active, research continued in information retrieval, where the practical problems were both more urgent and easier to solve. The central problem of information retrieval is to identify documents—be they books in a library, scientific papers, legal records, or, later, websites—that are most relevant to a search query. Much like machine translation, information retrieval confronts the problem of ambiguity.[33] When entered as part of a

33. Masterman and Spärck Jones recognized this similarity early on; see Masterman and Karen Spärck Jones, "The Analogy between Mechanical Translation and Library Retrieval," *Proceedings of the International Conference on Scientific Information* (Washington, D.C., 1959), pp. 917–35. Spärck Jones adapted the lattice network for this purpose in later articles and in her book-length study, *Automatic Keyword Classification for Information Retrieval* (London, 1971).

search query, polysemous terms undermined what IR researchers called the "precision" of returned results (because the query would return many irrelevant documents) while also limiting the system's recall (because relevant documents that used slightly different vocabulary would be missed).[34] The challenge was to group documents into larger categories that corresponded meaningfully to what users had in mind when entering a query.

It was in information retrieval that the distributional hypothesis matured into a full-fledged theory of semantic space. Two scholars often credited with this development are Hans Peter Luhn and Gerard Salton. Luhn worked at the IBM Research Center in the 1950s, where he developed business intelligence systems that sorted through large catalogues of documents. He proposed using a Keyword-in-Context (KWIC) technique for sorting titles, creating de facto subject categories based on any word. Like Masterman's thesaurus, the system Luhn imagined would deal with polysemy by building a dictionary of notions that grouped related keywords under common, fixed headings.[35]

Salton's approach was to use keywords to identify groups of similar documents, then build the index from there. As early as 1963 he suggested associating documents by creating an "incidence matrix" of their shared keywords.[36] In 1975, Salton, with Andrew Wong and Chungshu Yang, proposed "A Vector Space Model for Automatic Indexing," in which they put forward a mathematical technique for organizing documents into thematic clusters.[37] Their basic premise was to treat word counts as dimensions in space. If you imagine a collection of documents indexed by the fre-

34. For a discussion of precision and recall as research problems in information retrieval, see Spärck Jones, *Automatic Keyword Classification for Information Retrieval*, pp. 27–44.

35. Luhn describes these ideas in a series of essays published in the 1950s. See, in particular, H. P. Luhn, "A New Method of Recording and Searching Information," *American Documentation* 4, no. 1 (1953): 14–16. Writing in 1987, Gerard Salton describes the significance of this early work:

> It was suggested, in particular, that instead of assigning complex subject indicators extracted from controlled vocabulary schedules, single term descriptors, or key words, could be used that would be assigned to the documents without context or role specification. These single terms could then be combined in the search formulations to produce complex phrase specifications and chains of synonyms. . . . The coordinate keyword indexing methods eventually became the standard used in all automatic information retrieval environments. [Gerard Salton, "Historical Note: The Past Thirty Years in Information Retrieval," *JASIST* 38 (Sept. 1987): 376]

36. See Salton, "Associative Document Retrieval Techniques Using Bibliographic Information," *JACM* 10 (Oct. 1963): 440–57.

37. See Salton, A. Wong, and C. S. Yang, "A Vector Space Model for Automatic Indexing," *Communications of the ACM* 18 (Nov. 1975): 613–20. Further applications of the model were developed over the next fifteen years, so it is perhaps misleading (if common) to cite

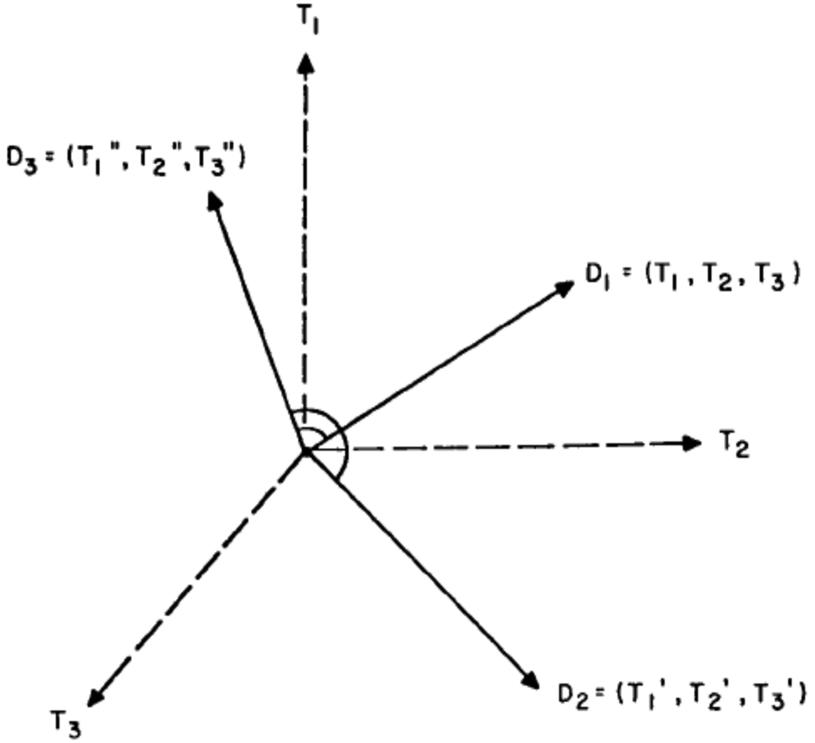

FIGURE 2. The Vector Space Model shows the distribution of documents in mathematical space. Each of these three dimensions (T1, T2, and T3) corresponds to the frequency of three key terms in the documents, which then can be plotted as points in vector space. In practice, vector spaces have many dimensions and so are difficult to visualize, except when projected (reduced) to a two-dimensional frame, as in figure 3. See Salton, Wong, and Yang, "A Vector Space Model for Automatic Indexing."

quencies of just three terms, those word frequencies can be plotted in a three-dimensional Cartesian frame. Documents that use similar words will appear near each other in this space and so can be grouped into automatically generated clusters (fig. 2). The key insight here is to move away from hierarchically defined structures of meaning (fig. 3). Rather than organize titles into predefined subjects or categories, Salton proposed using word co-occurrence (here, at the document level) to project search results into an imaginary field. Salton's original goal was to automatically generate subject headings, but it was quickly realized that this technique allowed for

the 1975 essay as the theory's authoritative defining statement. For the history of the reception of this paper and the theory's subsequent evolution, see David Dubin, "The Most Influential Paper Gerard Salton Never Wrote," *Library Trends* 52 (Spring 2004): 748–64.

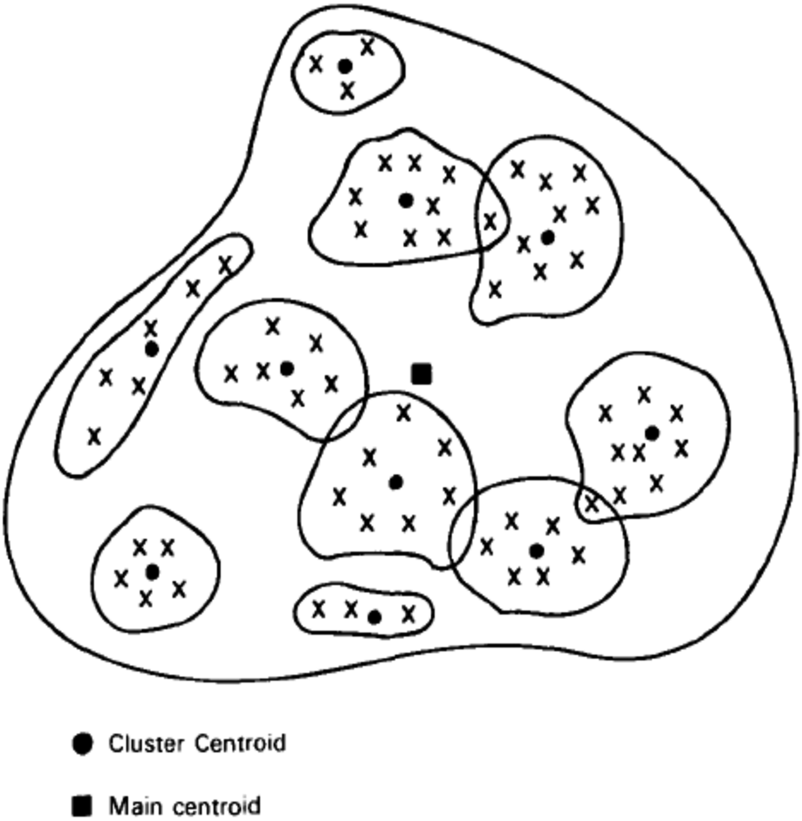

FIGURE 3. The Vector Space Model.

much greater flexibility. Whatever the search query, it was now possible to rank documents based on their location in space. Any time you search a catalog and the results are sorted by relevance, odds are you're using a system built on this principle.

Vector semantics uses spaces like these to model word meaning.[38] With the availability of large corpora and increased computing power in the 1990s, it became possible to create profiles of every word, measuring how often they appear in each other's contexts. Consequently, words them-

38. In 1992 Hinrich Schütze introduced vector semantics by drawing a parallel to information retrieval, citing Salton's and Michael J. McGill, *Introduction to Modern Information Retrieval* (New York, 1983) and Scott Deerwester, et. al, "Indexing by Latent Semantic Analysis," *Journal of the American Society for Information Science* 41, no. 6 (1990): 391–407; see Hinrich Schütze, "Dimensions of Meaning" in *Proceedings of the 1992 ACA/IEEE Conference on Supercomputing* (Minneapolis, 1992), pp. 787–96.

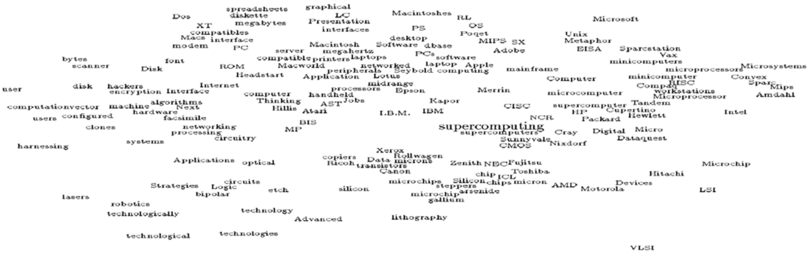

FIGURE 4. "The semantic field of *supercomputing* in sublexical space" (Schütze, "Dimensions of Meaning," p. 791): When projected onto a two-dimensional frame, Word Space looks like what we now call a word cloud. However, unlike a word cloud, position in the graph reflects each word's underlying semantic profile in relation to the others, usually when measured over a large corpus.

selves, not just documents, could be projected into what Hinrich Schütze called "Word Space" (fig. 4).[39] Harris long ago proposed that word distribution correlates to meaning; the challenge then was to discriminate among words' putatively discrete definitions. Using dictionaries or thesauri made sense because it was presumed that words had fixed meanings that could be made to fit more or less neatly under subheadings. Words were polysemous but not excessively so. Vector-space models dispense with this assumption. According to Schütze, "Vector similarity is the only information present in Word Space: semantically related words are close, unrelated words are distant."[40] In his 2006 study, Magnus Sahlgren extended this idea to what he called the "geometric metaphor of meaning": "*Meanings are locations in a semantic space, and semantic similarity is proximity between the locations*."[41] Word-sense disambiguation works by finding clusters in a word's semantic space, then measuring how closely any individual use of the word sits near each cluster.

In Word Space, definitions and categories are replaced with similarities and proximities, opening a critical vocabulary for describing ambiguities of infinitely fine gradation that stretch in infinitely many directions. Inso-

39. See Schütze, "Word Space," *Advances in Neural Information Processing Systems 5* (San Francisco, 1993), pp. 895–902.

40. Ibid., p. 896. Schütze differentiates the vector-space technique from dictionary- and thesaurus-based methods: "Any clustering into classes introduces artificial boundaries that cut off words from part of their semantic neighborhood. . . . Conversely, a multidimensional space does not make such an arbitrary classification necessary" (pp. 896–97).

41. Magnus Sahlgren, *The Word-Space Model: Using Analysis to Represent Syntagmatic and Paradigmatic Relations between Words in High-Dimensional Vector Spaces* (PhD. diss. Stockholm University, 2006), p. 19, eprints.sics.se/437/1/TheWordSpaceModel.pdf

far as vector semantics is a theory of meaning, it is a theory of ambiguity, pushing strongly against the impulse to draw clear boundaries that isolate words into discrete concepts. In most applications, word-sense disambiguation serves a purpose the mirror opposite of Empson's; instead of breaking a statement apart to trace its various possible connections, computational semantics are designed to choose a single meaning and return that meaning to the user. For software engineers, the goal is to design systems that accurately guess what people mean to say; that is, to reduce language to the level of human communication where words are more commercially valuable (as in, for example, filtering programs that target advertisements to users online). However, in order to achieve this, research in information retrieval and machine translation needed to develop a way of modeling ambiguity that was itself unambiguous. The result was a mathematical theory of finite space that measures lexical relationships across thousands of dimensions, where meaning as such is radically indeterminate.

In the next section I'll turn to the nuts and bolts of vector-space modeling and explain how it might be used to support Empson-style interpretation. Given Empson's use of the *OED*, scholars might be tempted to use computer-based dictionaries and thesauri, like those first designed by Masterman.[42] However, when analyzing historical literature, two considerations must be kept in mind. The semantic model must be sensitive to historical meanings, just as the *OED* was for Empson. As a practical matter, there exists no computer-based historical thesaurus or dictionary that is free and open to the public.[43] Building a vector-space model from scratch using historical corpora is actually easier and simpler than working with a structured lexicon. The more important consideration, though, is theoretical. Vector-space models are more closely aligned with Empson's notion of ambiguity. Empson always pushed against the dictionary he made use of. Meanings weren't discrete objects for him but makeshift focal points across which the ambiguities of a poem could be viewed. However, as Lynda Mugglestone has emphasized, the primary intellectual labor of the *OED* involved "condensing meaning into the economical and unambiguous ex-

42. The WordNet thesaurus, for example, contains entries for more than 150,000 modern English words and has provided data for many important studies. See *WordNet: An Electronic Lexical Database*, ed. Christiane Fellbaum (Cambridge, Mass., 1998). For a discussion of computational lexicography more generally, see Wilks, Slator, and Guthrie, who describe the history of computational lexicography and its relation to semantic philosophy in *Electric Words*.

43. *The Historical Thesaurus of English* project, based primarily on the *Oxford English Dictionary*, contains entries for over 800,000 words. Editors express willingness to license the data for academic research. See *The Historical Thesaurus of English*, historicalthesaurus.arts.gla.ac.uk

pression which was required."[44] By contrast, vector-space models build meanings from the corpus up—every use of a term is recorded and embedded in the semantic profile of a word.[45] Just as Empson imagined a horizonless range of interpretive possibility in the reader's mind, vector-space models are built to be sensitive to every historical use of a term—at least, to every use of a term as represented in the corpus. If words have surprising connections in vector space that seem to violate conventions of grammar and common sense, that's because words regularly violate grammar and common sense. The best way to know them is to compile and compare their uses. This, at least, is a premise that vector semantics share with Empson, whose perverse attention to diction deformed poetic statements, not to paraphrase their meaning, but to trace lexical associations through the conceptual space of language.

3. Word Space: A Mathematical Theory of Ambiguity

The distributional hypothesis says that words with similar meanings will tend to appear in similar contexts. The easiest (and, for information retrieval, often the best) way to measure collocation is to count words that appear in the same texts. In a term-document matrix, word counts are aligned as rows in a large table, with a column for each document. Figure 5 represents a subset of a matrix built from the works of Shakespeare, showing frequencies of the terms *battle*, *child*, *war*, and *wife* over five plays (fig. 5). Their distribution can be modeled by imagining those values as coordinates in vector space, just as in Salton's original model (fig. 6). Words suggesting political conflict predominate in history plays like *Henry V* and *Richard II*, while terms of domesticity appear more heavily in comedies like *The Merry Wives of Windsor* and *The Taming of the Shrew*. (As a story about a group of soldiers who engage in courtship while on holiday, *Much Ado about Nothing* provides a liminal case in this crude division.) The similarity between any two documents or any two words can be measured as a func-

44. Lynda Mugglestone, *Lost for Words: The Hidden History of the "Oxford English Dictionary"* (New Haven, Conn., 2005), p. 57.

45. The *OED*'s chief editor James A. H. Murray was aware of this possibility, in theory, but saw too many practical obstacles. He once remarked: "Quotations *will* tell the full meaning of a word, *if one has enough of them*" (James A. H. Murray, "The Evolution of English Lexicography," *International Journal of Lexicography* 6 [July 1993]: 118). Murray here responds to Charles Richardson's *A New Dictionary of the English Language: Combining Explanation with Etymology and Illustrated by Quotations from the Best Authorities* (1844), which had perfunctory definitions and several quotations for each entry. Murray continues: "while Richardson's notion was correct in theory, mundane conditions of space and time rendered it humanly impracticable" (p. 118).

	Henry V	Merry Wives	Much Ado	Richard II	Taming of the Shrew
battle	11	0	0	12	1
child	2	6	10	3	3
war	18	0	3	14	3
wife	4	39	6	3	29

FIGURE 5. A term-document matrix drawn from the works of Shakespeare. In this example, the word *battle* is represented by the row of numbers (11, 0, 0, 12, 1) and the play *Henry V* is represented by the column (11, 2, 18, 4.)

tion of the distance between their angles. On the right, *battle* and *war* actually sit relatively far apart as points, but their vector lines trace very similar trajectories. The matrix doesn't know that *battle* and *war* have related meanings, but it does show that they follow similar paths through the corpus. In the same way, the matrix doesn't know that genre categories like comedy and history exist, but it does find lexical variations that sort this miniature archive into clearly delineated groups.

When measured at the document level, collocation thus follows the intellectual tradition of information retrieval by gathering documents into clusters and grouping words into subject headings, or "topics."[46] However, to identify more finely grained relationships among words requires measuring collocation more closely. In "Distributional Structure," Harris imagined breaking a corpus into sentences, but most studies use a simpler context-window method, which measures bare proximity (using a range of, say, five words, or ten, or fifty) regardless of grammar. Essentially, this method involves building a concordance, counting the context words under each heading, and then placing those values into large tables, sometimes called term-term matrices or context-word matrices, in which each word is represented as a large sequence of numbers showing its co-occurrence with the others. Whereas figure 6 above displayed the distribution of terms and texts over just two dimensions, in practice context-word matrices are built with hundreds, even thousands of variables (that is, columns or dimensions). Nonetheless, the basic geometry of meaning continues to apply, so it's still possible to calculate similarities among the trajectories each word follows through the corpus. To create a profile of any given word's various

46. See David M. Blei, "Probabilistic Topic Models," *Communications of the ACM* 55 (Apr. 2012): 77–84. Topic modeling is a popular variation of this more general information-retrieval task.

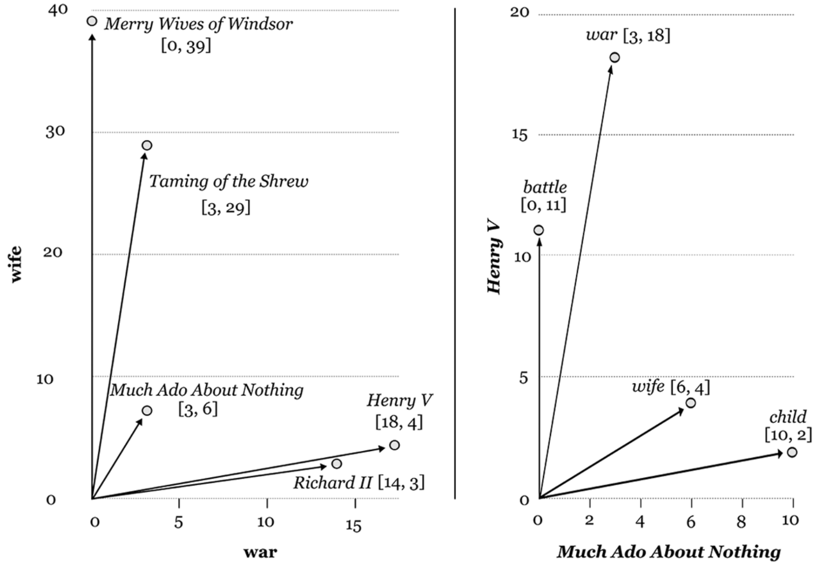

FIGURE 6. Values from a term-document matrix can be plotted onto a two-dimensional frame. *Left*: over *war* and *wife*, the plays separate into genres. *Right*: over *Much Ado* and *Henry V*, the terms separate into topics.

uses requires two basic steps: first, finding the words that tend to appear in similar contexts, then second, comparing those words to each other and thereby sorting them into conceptual groups.

Consider, for example, the word *foot*. A semantic profile of this word was compiled using data drawn from the full-text documents of *Early English Books Online*, published in 2015 by the Text Creation Partnership, and limited to include only the 18,351 documents dated 1640 to 1699. From those documents I compiled a list of 1,751 keywords (excluding about 300 stop-words made up of prepositions, pronouns, and various oddities caused by the transcription process). I built a concordance for each keyword using a context window of five words, then placed the resulting word counts into a large matrix with 1,751 columns (for each keyword) and 28,235 rows (for each collocate term, excluding extremely rare words). Each row thus represents a word in the corpus as a vector of numbers, showing its rate of co-occurrence with each keyword. The result is a large vector-space model of seventeenth-century printed English.

Figure 7 shows the thirty words with collocation profiles most similar to *foot*, grouped using K-means clustering and projected onto a two-dimensional graph (fig. 7). In the seventeenth century, *foot* was invoked primarily in

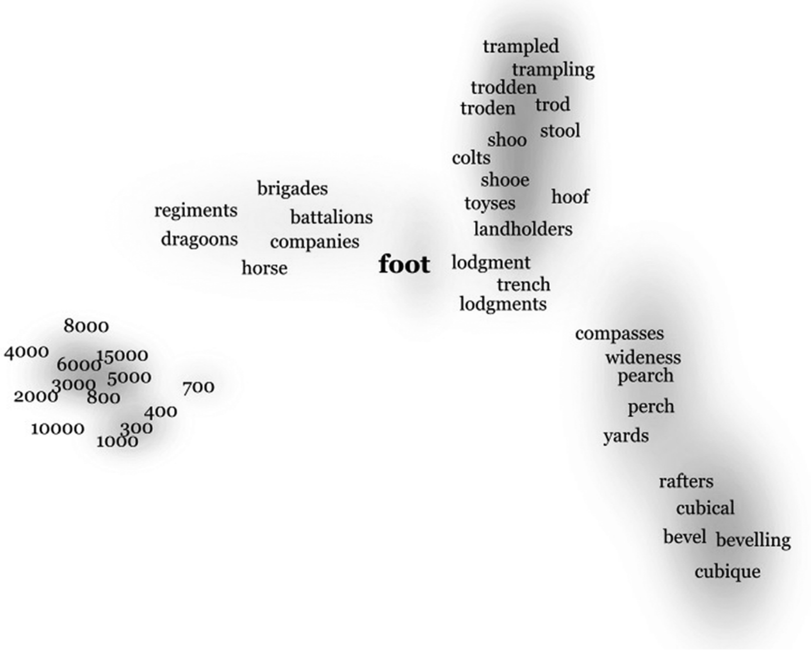

FIGURE 7. The semantic neighborhood of "foot." This graph shows about fifty of the words most similar to *foot* in the seventeenth century. These words were grouped using K-means clustering; contour lines were added to help visualize the semantic region occupied by each cluster.

three contexts. Most commonly, it served in military discourse as a synecdoche for infantry, and so it scores highly similar to words like *regiments*, *brigades*, *battalions*, *dragoons*, and *companies*, while the cluster of number words near the bottom left suggest the sizes of these companies. Numbers were also likely invoked when *foot* was used as a unit of measurement, where it corresponded closely to *yards* and carpentry (*bevel*, *beveling*, and *rafters*). As a part of the body, *foot* invoked certain kinds of movement (*trod*, *trodden*, as well as the more aggressive *trampled* and *trampling*) that were associated, not with the human body, but with the *hoof* of a *horse*. Dividing *foot*'s uses in this way provides a straightforward example of word-sense disambiguation; the term had different meanings, and those meanings can be teased apart through simple statistical extrapolation.

If we imagine a search engine trained on seventeenth-century data, the task would be to guess which sense of *foot* a user was interested in and to return documents that correspond to that meaning. The system must differentiate between the word in general and its use in any particular context. The simplest and most common technique is called vector composition,

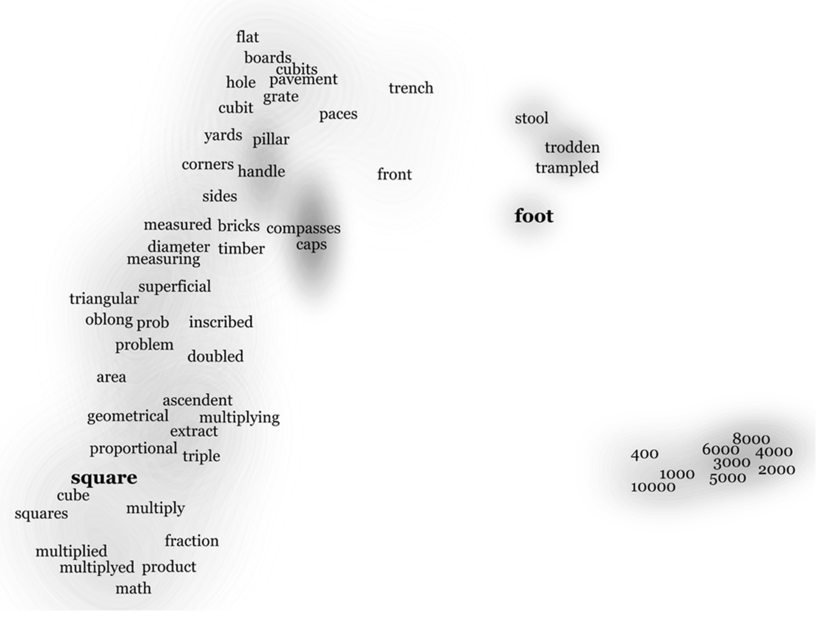

FIGURE 8. The semantic neighborhood of "square foot" = *square* + *foot*. When two vectors are added together, the angle of the resulting vector bisects its component vectors, exposing the semantic field that rests between them.

which means simply to add words together.[47] If a user were to enter the search query "square foot," the system could break it apart into two words and add together their corresponding word vectors into one composite sequence (fig. 8). The effect is the same as if you took the mathematical average of the two words. The angle of this aggregate vector's trajectory through Word Space cuts through the middle of each of its components, exposing the semantic area they share while stripping away less relevant collocates and sometimes teasing out new connotations. The composite vector *square foot* veers far away from *foot*'s military associations, focusing on terms of measurement and mathematics. Dominic Widdows explains, "Continuous methods enable us to model not only atoms of meaning such as words, but the space or void in between these words. Whole passages of text are mapped to points of their own in this void, without changing the underlying shape of the space around them."[48]

47. For an introduction to this technique, see Katrin Erk, "Vector Space Models of Word Meaning and Phrase Meaning: A Survey," *Language and Linguistics Compass* 6 (Oct. 2012): 635–53.

48. Dominic Widdows, *Geometry and Meaning* (Chicago, 2004), p. 164.

Such visualizations can assist disambiguation in the style of Empson. They provide intuitive, thought-provoking lines of comparison that show how words relate to each other in Word Space, even if the relations they visualize sometimes resist logical description. Empson, it's worth noting, anticipated something like this form. He remarked that his method promotes "a sort of doctrinaire sluttishness; one is tempted to set down a muddle in the hope that it will convey the meaning more immediately" (*S*, p. 236). Indeed, one of the many pleasures of vector-space models—as well as of data visualization in general—is the pleasure of succumbing to precisely this temptation.

4. Milton's Satan in Vector Space

Before turning to a literary example, it's worth pausing to review the argument laid out so far. Ambiguity presented Empson with a motivating framework for describing many possible interpretations of a poem. By recognizing that words gained meaning through use across different contexts, then assuming that those uses somehow inhered to words across all contexts, he could examine their latent implications by comparing individual poetic statements to the language as a whole (as modeled by the *OED*). For the budding field of computational linguistics, ambiguity posed an intriguing challenge. The goal was to automatically measure and adjudicate among words' possible connotations. In machine translation, this involved differentiating among word senses to choose the appropriate analogue in a target language. In information retrieval, word-sense discrimination was a necessary first step for choosing the best documents to return in a search query. Inspired by George Zipf's and Shannon's work, early corpus linguists assumed that the best mathematics for studying meaning was information theory, but over the next several decades researchers learned to rely on finite dimensional representations of vector space.[49] The key insight these scholars shared was to view words as part of a geometric system. Every word entails latent connections with the others, and these connections are what make possible their meaningful recombination in human discourse, whether that recombination happens in a graph, a search query, or a poem.

49. Developments in linear algebra coincided with those in literary studies and linguistics and necessarily preceded their application in information retrieval. In his foreword to Widdows's *Geometry and Meaning*, Pentti Kanerva describes linear algebra as "the proper math" that is "appropriate for describing concepts and meaning" (Pentti Kanerva, foreword to Widdows, *Geometry and Meaning*, pp. x, xi).

This theory of ambiguity implies a new ontology of the text that was intuitively grasped by Empson. Every word in every text is a kind of node connecting that text to a larger universe of language, where the word appears in many variations that supervene over every particular use. Vector semantics embraces his theory's most radical connotations while lending it greater precision and analytic power. Once a Word Space is modeled, any text becomes an ordered set of points from that space. A text is no longer structured by formal markers of genre, but neither is it reduced to an isolated sequence of words; instead, every word participates in a complex network of relations that collectively make up lexical space as such. The text is reimagined as an ordered subset of that space.

To get a picture of what I mean, consider for example the following oft-cited lines from book 9 of John Milton's *Paradise Lost*. In this passage, Satan in serpent form has intruded on the Garden of Eden and, confronting Eve, is suddenly overawed by her innocence and beauty:

> Such Pleasure took the Serpent to behold
> This Flourie Plat, the sweet recess of *Eve*
> Thus earlie, thus alone; her Heav'nly forme
> Angelic, but more soft, and Feminine,
> Her graceful Innocence, her every Aire
> Of gesture or lest action overawd
> His Malice, and with rapine sweet bereav'd
> His fierceness of the fierce intent it brought:
> That space the Evil one abstracted stood
> From his own evil, and for the time remaind
> Stupidly good, of enmity disarm'd,
> Of guile, of hate, of envie, of revenge;
> But the hot Hell that alwayes in him burnes,
> Though in mid Heav'n, soon ended his delight,
> And tortures him now more, the more he sees
> Of pleasure not for him ordain'd: then soon
> Fierce hate he recollects, and all his thoughts
> Of mischief, gratulating, thus excites.[50]

This passage brings together a number of important Miltonic themes: among them the problem of evil, the operation of the senses and the passions and their effects on the will, and the relationship between intentional actions and the spatiotemporal conditions of action's possibility.

50. John Milton, *Paradise Lost* (London, 1667), tei.it.ox.ac.uk/tcp/Texts-HTML/free/A50/A50919.html#index.xml-body.1_div.9

These lines have been the subject of extensive Empson-style commentary, even though Empson himself never reflected on them at length. He mentions them only in passing in *Milton's God* (1961), which was much more devoted to his critique of Christianity than his semantics. (Empson argues that, because Satan is "still capable of being struck 'stupidly good,'" "one must also feel horror at the God who has deliberately reduced him to such a condition.")[51] However, Brooks calls special attention to these lines in "Milton and the New Criticism" (1951), in which he attempts to reconcile Milton to a paradigm that valorized John Donne. *Paradise Lost*, he writes, is "bound up with functional metaphor, with dramatic tension, and with the fusion of thought and emotion."[52] For Brooks, this encounter shows paradoxically that when Satan "is abstracted from his own evil, he is not only good (as separated from evil), but stupidly so, for he has so thoroughly given himself up to evil that he is now abstracted from part of himself, and can be 'good' only in a dazed and stupid manner. . . . Such goodness is not an active virtue at all."[53] Writing in 1967, Wayne Shumaker noted how the lines describe "an emptying, a draining of both moral and physical energy which leaves Satan powerless and gaping."[54] Annabel Patterson sees here a reflection on the nature of evil itself, arguing that Milton "raises in stark form the central epistemological issue as to whether evil really should be conceived in the aggregate, as an abstraction, as distinct from discrete events and actions."[55] Stephen Fallon, by contrast, argues that in Milton "evil actions . . . are essentially good actions misdirected."[56] Shannon Miller and Adam Piette have emphasized how Satan's desiring gaze creates a moment "of rapture, of a suspension in time, and most significantly of inaction,"[57] during which Eve's beauty "occasions an *abstracting* of the gaze: [Satan's] intent can no longer be itself."[58] Most recently, Tom

51. Empson, *Milton's God* (London, 1961), p. 70.

52. Cleanth Brooks, "Milton and the New Criticism," *Sewanee Review* 59 (Winter 1951): 3.

53. Ibid., p. 14.

54. Wayne Shumaker, *Unpremeditated Verse: Feeling and Perception in "Paradise Lost"* (Princeton, N.J., 1967), p. 184.

55. Annabel Patterson, "Milton and the Problem of Evil: a Preemptive Modernism?" in *Uncircumscribed Mind: Reading Milton Deeply*, ed. Charles W. Durham and Kristin A. Pruitt (Selinsgrove, Penn., 2008), p. 40. Blakey Vermeule extends this line of thinking by reflecting on Satan's automaticity in "The New Unconscious: A Literary Guided Tour" in *The Oxford Handbook of Cognitive Literary Studies*, ed. Lisa Zunshine (New York, 2015), pp. 463–82.

56. Stephen M. Fallon, *Milton among the Philosophers: Poetry and Materialism in Seventeenth-Century England* (Ithaca, N.Y., 1991), p. 187.

57. Shannon Miller, *Engendering the Fall: John Milton and Seventeenth-Century Women Writers* (Philadelphia, Penn., 2008), p. 64.

58. Adam Piette, "Beckett's Eve: *Ill Seen Ill Said* and the Miltonic Attendance Motif," in *After Satan: Essays in Honour of Neil Forsyth*, ed. Kirsten Stirling and Martine Hennard Dutheil de la Rochère (Newcastle upon Tyne, 2010), p. 169.

MacFaul has synthesized these insights into a general account of Milton's moral time space: "The word 'space' here refers to a spot of time, but it also implies that the suspension of evil has to do with being embedded in good space; the time taken over the simile enacts that suspension into stupor."[59]

Vector semantics intervenes in the critical reading process by providing more finely grained models of individual words and phrases and by showing how meanings are transformed in context. I will focus particularly on the ninth line quoted above: "That space the Evil one abstracted stood." In this line, the words *that*, *the*, and *one* are all stop-words, leaving the analysis to focus on four terms: *space*, *evil*, *abstracted*, and *stood*. Much like the word *foot*, the words *space* and *abstracted* are divided into fairly different senses, while the words *evil* and *stood* occupy more thoroughly integrated conceptual spaces (figs. 9–12): *space* segments fairly neatly into two subgroups that emphasize spatial extension (*empty*, *infinity*, *vacuum*) and time interval (*years*, *fortnight*, *lasted*); *evil* affiliates closely with intentional action (*doers*, *doing*, *deeds*), synonyms in both noun and adjective form, as well as with verbs that suggest a dangerous relation between grammatical subjects and objects (*avoid*, *shun*, *temptation*, *befall*); *abstracted* connects to highly specialized vocabularies for ontology (*entity*, *essence*, *existence*), epistemology (*imagination*, *sensation*), and physics (*selfmotion*); *stood* was most similar to modifiers associated with wonder (*amazed*, *astonished*) and conflict (*defiance*, *obstinately, undaunted*), as well as to immobile physical objects and architectures (*statue*, *pillars*, *porch*, *altar*).

When considered as a principle of information retrieval, a line of poetry operates essentially like a search query that selects items from a database, each of which contains a vector profile that can be compared over the whole. The associations mapped in figure 9 show the words with profiles most similar to those in the subset. Whereas Empson understood such connections as a latent possibility in the reading mind—as "alternative reactions"—vector spaces represent those connections using collocation measurements. Readerly possibilities are described anew as statistical extrapolations of historically concrete text data.[60]

Just like a search query, a line of poetry can be composed into an aggregate entity; in this case, $line\_9 = space + evil + abstracted + stood$. Figure 13 shows the words most similar to this composite vector (fig. 13). When

59. Tom MacFaul, *Shakespeare and the Natural World* (New York, 2015), p. 47.

60. Empson later recognized this theoretical commensurability. In a 1971 letter, he remarks that collocations "often give rise to new meanings of the word, so cannot be kept isolated from its meanings" (Empson, letter to Jha, p. 510). In a letter the next day he was even more emphatic: "a standard collocation often becomes a meaning of a word" (Empson, letter to Norris, 21 June 1971, in *Selected Letters of William Empson*, p. 511).

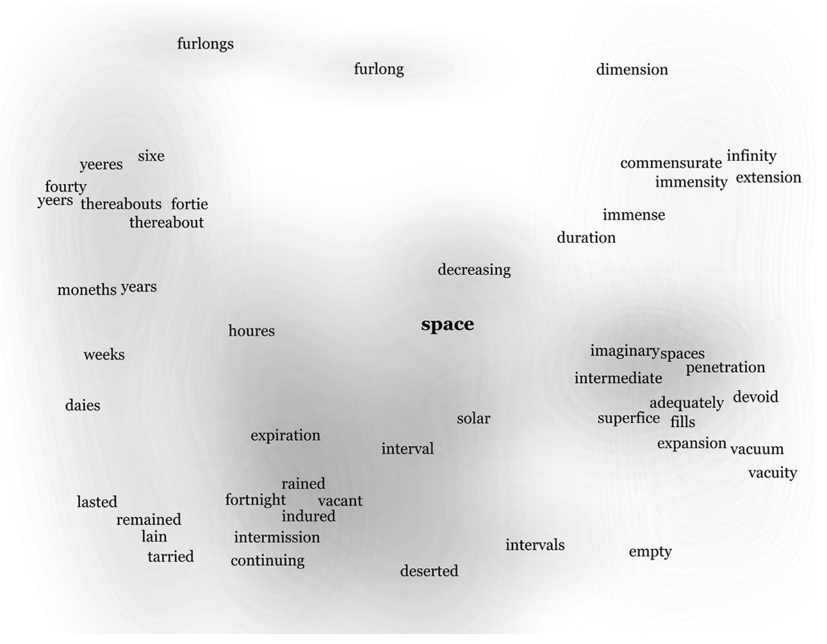

FIGURE 9. Lexical neighborhood of the word *space*.

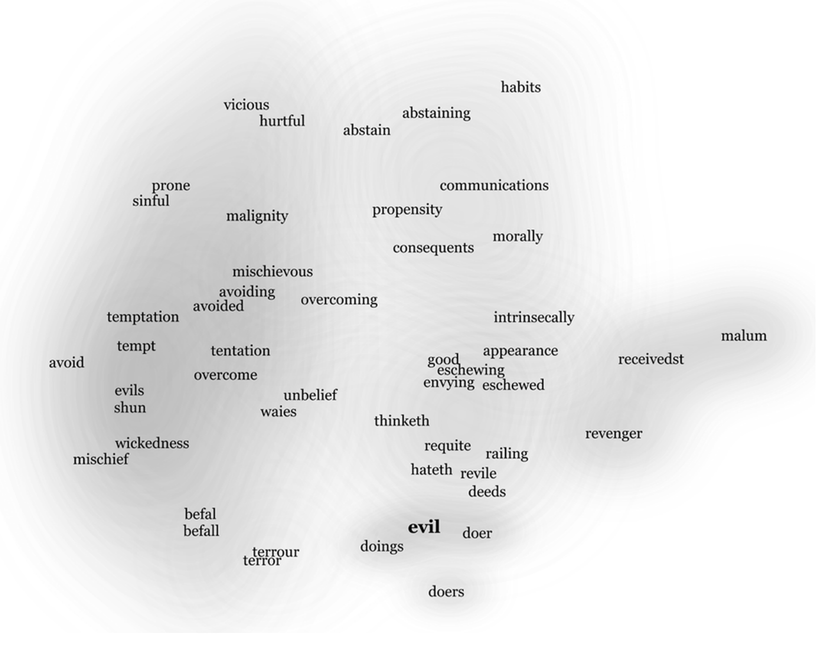

FIGURE 10. Lexical neighborhood of the word *evil*.

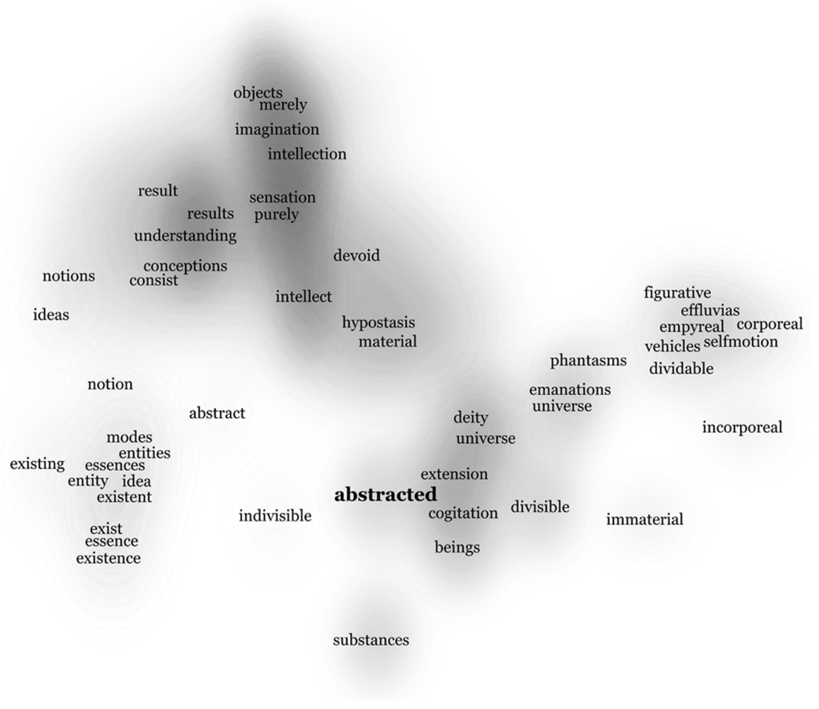

FIGURE 11. Lexical neighborhood of the word *abstracted*.

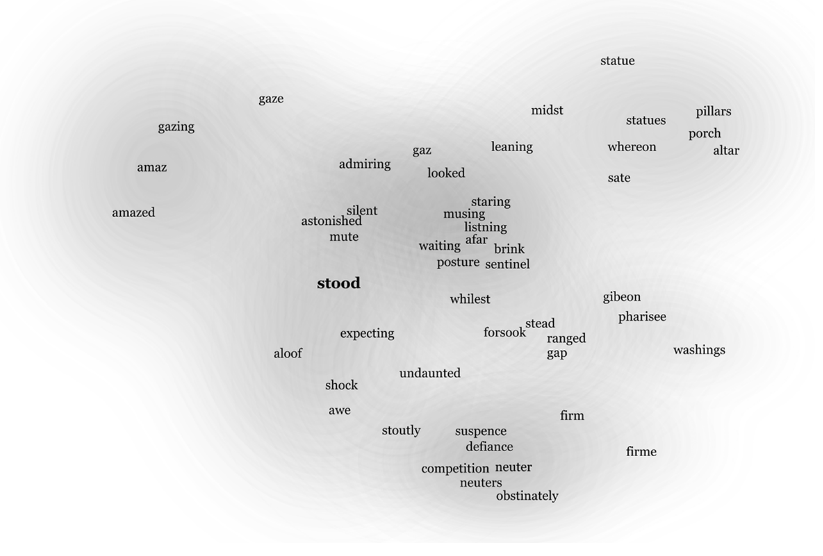

FIGURE 12. Lexical neighborhood of the word *stood*.

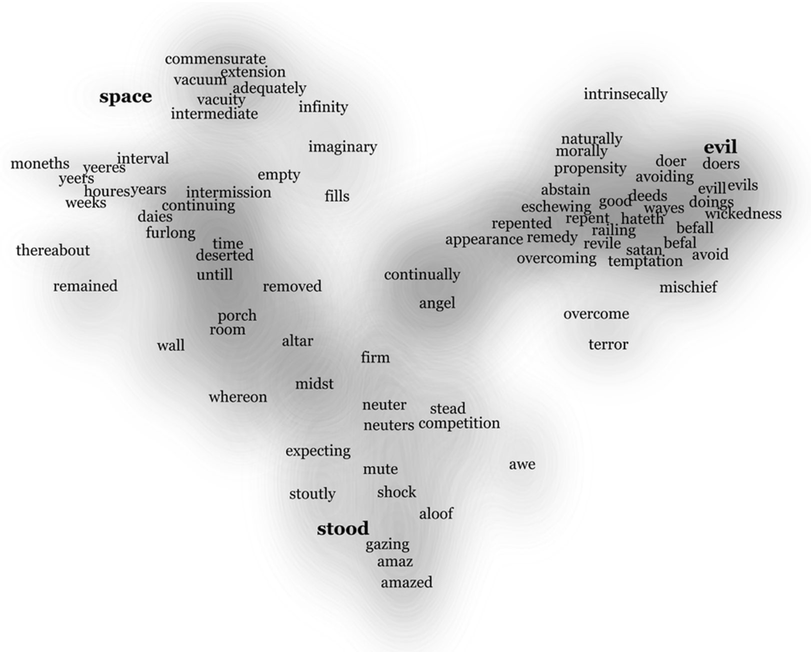

FIGURE 13. The semantic neighborhood of the composite vector, *space* + *evil* + *abstracted* + *stood*.

combined into an aggregate quasi word, a number of meanings associated with the individual terms fall away. The specialized vocabulary that surrounds *abstracted* dissolves into the whole, and when projected onto two dimensions the phrase assumes a triangular shape. The temporal and spatial connotations of *space* remain tightly juxtaposed, and they connect to *stood* through concepts of physicality (*porch*, *room*, *wall*, *whereon*). Wonder and astonishment surround *stood* closely, while its conflict-laden associations mostly disappear. The word *evil* retains its coherence and its emphasis on action while gaining new connotations: *satan*, most tellingly, but also words like *intrinsically*, *naturally*, and *morally*, which suggest the juxtaposition of *evil* with *space* and *stood*, put conceptual pressure on *evil* itself, as if an evil that stands in space is already subject to a kind of ambient doubt about its very nature, especially when contrasted to the figure of the *angel* who now sits quite literally in the midst of the line's semantic field. The model thus captures well the conceptual dynamic that Milton is playing with here, contrasting an active evil with the passive experience of the gaze. The space of evil is a space of action, of doings and deeds, of eschewing and overcoming, while Satan at this moment occupies a very different kind of

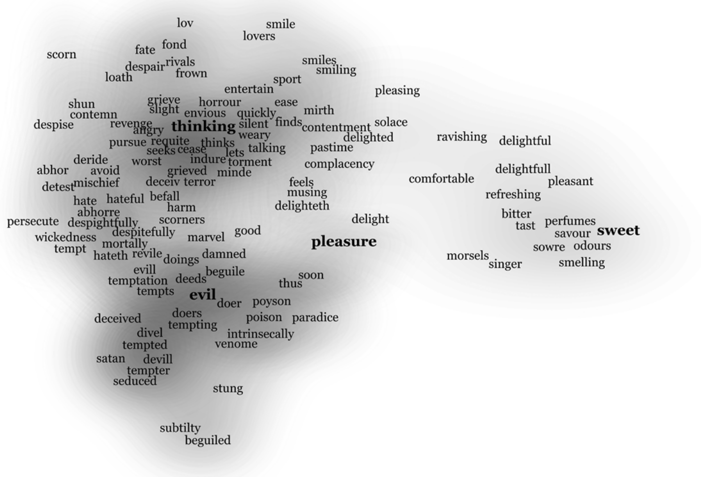

FIGURE 14. The semantic neighborhood of the entire passage.

space, one characterized by abstract spatiotemporal dimensionality more appropriate to angels, an emptiness visible only as an absence of action and volition.

It's possible to scale out further, taking a composite vector of the entire passage (fig. 14). When taken as a whole, the lines describe a kind of evil thinking, suddenly confronted by an unexpected pleasure of sensation. The words that orbit *thinking* form rings of association that descend toward *evil* on one side and connect loosely, impotently toward *solace* on the other.

We might say that the space Satan briefly occupies is outside history, or that it provides history's conditions of possibility and is therefore conceivable only by abstracting history away. We might say something similar about vector space, where variety permeates across many dimensions that are irreconcilable to the human experience of actuality in time.

5. Conclusion

I began this essay by promising to introduce vector semantics to an audience of humanists. In doing so I have tried to avoid dreary polemic of the kind that typically surrounds the digital humanities, most of which fetishizes computer technology in ways that obscure the scholarly origins of the software tools so haphazardly employed. As I hope is clear, the primary obstacle humanities computing faces is neither technological nor political but disciplinary and formal. Mathematicians, computer scientists, and software engineers have a lot of smart things to say about the nature of meaning, space, and time. We need to learn how to learn from them, though, because how they write is, at the level of style and genre, so far outside the bounds of humanistic scholarship that it's difficult to recognize as theory. Like Satan at the foot of Eve, the digital humanities stand stupidly good. Hence the proliferation of uninspiring calls to import cultural criticism to the field, as if familiar modalities of thinking will get things moving (or reveal more than they mystify). Perhaps the same accusation could be leveled at my use of Empson in this essay. Readers might very well respond, "Forget about Empson. . . . Tell me more about Margaret Masterman!" If so, that would be a sign that the conversation is starting to get serious.