Toward a Computational Archaeology of Fictional Space

Dennis Yi Tenen

New Literary History, Volume 49, Number 1, Winter 2018, pp. 119-147 (Article)

Published by Johns Hopkins University Press
DOI: https://doi.org/10.1353/nlh.2018.0005

Toward a Computational Archaeology of Fictional Space

Dennis Yi Tenen

Space is a hard thing to pin down. It identifies dimensional continuity and a topography, that is, a relationship between objects. It is also itself an object: a limit-defining quantity even in its most abstract sense. "O God, I could be bounded in a nutshell and count myself a king of infinite space—were it not that I have bad dreams," Hamlet says of his ambition and his dreams.[1] A human palm is a part of the body and a map. A mirror is a piece of furniture and a frame for reflection. Under extreme magnification, the head of a pin appears a vast and mountainous terrain, home to angels and bacterial detritus. The characterization of diegetic—let us call it also virtual and fictional—space presents further difficulties. A stretch of land in fiction also measures a stretch of the imagination. These units do not always have names or explicit boundaries. Vladimir and Estragon wait for Godot: "*A country road. A tree.*"[2] Two vectors are enough to situate the world. A road gives us the *x* and a tree the *y* axis: an infinity in a nutshell.

In this paper I propose to reconsider theories of diegetic space that rely on explicit framing (i.e., "two people walk into a room" or "in Spain"). Rather than looking for maps, I define space in terms of lexical categories denoting objects. The emphasis on objects leads to a method for literary archaeology, informed by cognitive theory and anthropology. If the universe is made of atoms, a fictional world is also made up of atomic relationships that form basic, stable configurations, or what I call *narratological primes*. I construct several such basic spatial buildings blocks here—*diegetic density* and *clutter distance*. Their application to a well-explored body of nineteenth-century novels challenges several long-standing historical intuitions related to the development of material culture in the nineteenth century.

A theoretical reconfiguration grounds a descriptive method, a model by which fictional space can be, if not fixed, then approximated. Further, following Lisa Samuels, Jerome McGann, Johanna Drucker, and others, I am interested in literary modeling and visualization as kinds of a transformative reading practice. The formal, computational meth-

New Literary History, 2018, 49: 119–147

ods I present here are not meant to prove anything. They are first and foremost exploratory and experimental tools. They occasion opportunities for close reading, and not just reading at scale. My methods are diagnostic, in that they identify areas of interest and unusual trends that require closer critical attention.

An undercurrent and starting place of this essay is therefore also a critique of a certain mode of quantitative literary analysis, which essentially advances a number of complex methodological procedures to arbitrary effects. Formalists old and new are perpetually and cyclically in danger of falling into the trap set by Stanley Fish in the 1970s. Fish cautioned that a relation can always be found between any number of low-level, formal features of a text and a given high-level account of its meaning. For example, the use of past participles or gerunds may rise and fall with the vagaries of literary style. Or it may be due to changing archive collection practices over time: sample bias. As Fish puts it: "there are always formal patterns" and "a relation will always be found."[3] This is astute. Methods inevitably produce results. The difficulty lies in filtering meaning from the noise. A whole subset of statistical methods—causal analysis—is dedicated to this problem, involving explanatory frameworks that chart a pathway between dependent and independent variables. Methods require theory. A tautological formula cannot in itself produce meaning: one must have a priori ideas about what is meaningful. Correlations become more convincing with the interpretation of their causes. Analytics, in other words, are hollow without poetics.

To get past Fish, we must first understand the difference between prediction and explanation. Karl Popper wrote that the aim of theoretical science is to find "explanatory theories . . . which describe certain structural properties of the world."[4] A theoretical interest in explanation is noninstrumental, insofar as it is "irreducible to the practical technological interest in the deduction of predictions."[5] A diagram of a storm system holds explanatory and not just predictive power. In fact, it does not describe any specific weather systems at all, past or future. Rather, a diagram teaches us something about the relationship between causes and effects: high pressure, condensation, wind, and rain. To understand how something works—to form a theory—further entails the possibility of effecting systemic change. To trace the pathology of a disease is also to imagine ways of inhibiting it. In this way, explanatory models reach beyond the apparent phenomenon, from what is to what might be: a fever lifted from a sick child. Models thus contain the remainder of the real, which is poiesis itself, creativity.[6] What is the point of theory? There is none, Popper answers. Theory is a storehouse of potential applications, knowledge for knowledge's sake.

Even a simple forecast places ideational constructs in relation to the empirical world. Predictive power and explanatory power reciprocate one another. To predict is also to imagine, albeit in a more directed, fully potentiated way. Pattern recognition suffices for this purpose. One can forecast the sun to rise each day based on past experience, knowing nothing about planetary mechanics. Patterns allow us to extrapolate from known contexts. Weather models will produce accurate predictions provided that our planet remains more or less the same.

Not all theories need be predictive, however. The interpretation of historical events involves an account of singular causes and effects that do not repeat in the same configuration. The interpretation of literary texts also hinges on deeply contextualized, affective, or embodied dynamics. Predictive power may or may not be necessary to understand how a text works, where explanatory power is. To know how texts work, in the echoing words of Percy Lubbock, Boris Eikhenbaum, and Susan Sontag, is the essence of poetics.

I offer these reflections on method en route to an argument about cultural analytics—the application of computational methods to the study of literature and culture—generally, and, more specifically, in an attempt to outline some strategies for explanatory modeling of space in literary texts. Some research questions, I maintain, are amenable to predictive analysis. Others require explanatory power. The two complimentary approaches involve different logics and modes of argumentation. It is important, however, to differentiate clearly between the two. The worst kind of error is one where predictive results are taken for explanatory ones: the sun will continue to rise because it has been rising regularly. The "because" is unwarranted. We must not mistake mere extrapolation for an account of deep causes and effects. The understanding of culture, to paraphrase Clifford Geertz, cannot be limited to pattern recognition. It is rather an interpretive effort, "in search of meaning."[7] As such, it requires other attributes besides statistical significance or convenience, such as simplicity, novelty, or persuasiveness. An explanatory, exploratory, interpretive model does not just extrapolate; it leads to insight.

Narratological Primes

Many common problems in computational text analysis are fundamentally problems of classification. To this we may attribute questions of genre: given two stacks of known texts labeled detective and romance fiction, what is the probability of an unclassified text belonging to one or the other category? Similarly, questions of authorship or period attribution can also be reduced to a categorization guessing game.

The explanatory approach to solving such problems involves building a model. The model may take into account a number of formal elements intrinsic to a text. In this way, we can reduce the complexity of the detective fiction genre to a set of commonly occurring themes involving gun violence, murder, or elements of investigation. Whatever the case may be, a critic in search of explanatory models will rely on a set of convincing, normative assumptions. These will have to be argued on a speculative, theoretical basis, settled prior to experimental observation.

If I told you, for example, that my understanding of detective fiction involved only novels of a certain length, you might reasonably object to the limitations of a model that ignores the short stories of Edgar Allan Poe.[8] We intuit that formal elements such as length do not always translate into epistemic categories, which are, at least in part, socially constructed. At the very least, the assumptions so formalized need to be revealed at the outset of analysis, discussed, and defended against expected objections. An explanatory model contains, among other attributes, an account of poetics, or how the thing is made and how its parts fit together. It is reductive to the extent that it simplifies the phenomenon to its most salient features. To build a formal model of this sort is to understand the domain mechanics: the causal linkages between words and themes, crime and violence, pulp fiction and modernism, guns and detectives, suspense and its resolution.

Alternatively, in reasoning about detective fiction naively, without prior knowledge, we could rely on existing genre classification as identified by authors themselves, library catalogs, booksellers, or publishing houses. A common strategy for predictive analysis involves training a classifier on an existing corpus, thereby extending known epistemic categories onto a new set of unlabeled texts. By these so-called "supervised" means, a researcher can ask readers to identify detective fiction in a collection of texts that includes a mix of genres. An algorithm can subsequently use a number—dozens or millions—of formal features to identify other documents that are similar, in some way, to the ones in the "detective" pile. That task is useful if we simply want to find "more texts like these." The classifier tells us little about the mechanics—the causal linkages—of detective fiction. Rather, it reveals a correlation between arbitrary discovered features and common epistemic categories. Some of these correlations may be meaningful: shorter sentences are associated with hard-boiled fiction. This association is "meaningful" because we understand its linkages. Other associations are less so—were we to find, for example, that greater lexical variety corresponds inversely to books published in paperback. The association between lexical variety and binding quality is confounded by numerous intermediate factors,

and could, under closer causal examination, dissolve into contingency.

Predictive methods for text categorization are judged by their measure of recall and precision.[9] They do not have to appear convincing, only effective. In the case of categorizing detective fiction, researchers may not care about what the category means, what social function it plays, or how it is constructed. At times, it is important merely to select the appropriate document efficiently: returning a high-enough number of relevant documents.[10] In these cases, formal features such as word frequency or a preponderance of pronouns are incidental to analysis. It is sufficient that they are indicative of a relationship, however spurious—human births and the number of nesting storks—between variables.

Supervised machine learning algorithms are therefore reductive in a different way: they identify compelling correlations between categories used by humans and abstract features that make sense only to a machine. They are, in a word, without a model. Witness, for example, the early development of self-guiding robots, which one researcher describes as a "model-less description of a problem space, hidden in the adaptive connection strength values of the neural net."[11] However, something about literary genres cannot be reduced to a set of formal features of the text, much less to "adaptive connections" hidden in metaphorically "neural" approximations of the human brain. Genres are also social and commercial categories. Formal features correlate to their social functions in a contingent, nondeterministic way. Some authors writing in the "detective" genre will explicitly violate its formal rules, for example, in order to differentiate themselves in the marketplace. Given a model-less understanding of conventional detective fiction, these outliers will simply be discarded as "imprecision," where they might be precisely the texts of concern to a literary historian.[12]

Explanatory models require a different approach. Where prediction often involves complex statistical tools—neural-networks, machine learning, Markov chains, semantic word vectors, and word adjacency networks, among other state-of-the-art techniques—explanation demands a convincing account of linkages between the constituent parts of the proposed model. I therefore advocate for what Daniel Nolan and other philosophers of science have called the virtue of "quantitative parsimony," particularly as it applies to the analysis of complex cultural systems such as literature.[13] For the purposes of literary analysis, quantitative parsimony implies a preference for modular, atomic models, which make use of the configurations I have termed *narratological primes*.[14] Complex intuitions about diegetic worlds—narrative structure—rely on a limited number of foundational building blocks, which, when articulated, can be used to construct more sophisticated analytical models with considerable explanatory potential.

In this paper, I introduce one such primal construct as a stepping stone to a discussion about narrative space. I am guided in my thinking by recent developments in the material anthropology of domestic space as well as by the work of a little-known, but important, early twentieth-century Russian philologist, Boris Iarkho, whose writings I will translate and introduce here. Iarkho is important to a contemporary audience interested in cultural analytics because the methods he developed anticipate the "bag-of-words" and "topic modeling" approaches widely used in modern computational text analysis. I will finally propose several metrics for measuring the density of diegetic space and subsequently use these metrics to explore the well-known distinction between urban and rural space in the realist novel.

Effect Space

Thematically, this essay takes up the challenge posed by David Elson, Nicholas Dames, and Kathleen McKeown in their influential study "Extracting Social Networks from Literary Fiction."[15] The study tested a literary-theoretical thesis positing a relation between the increasingly urbanized settings of the realist novel and the complexity of the social networks found within. According to prior theories, as the number of characters in a novel increases, we should expect to observe a decrease in the amount of dialogue. This decrease seems to arise, at first glance, as a consequence of urbanization, which alters the structure of social networks and therefore the interaction between characters. However, neither of these propositions were found to hold up by experimental results in the original study—its provocative conclusion.[16]

Is it possible for increased urbanization of literary space to entail other qualitative differences? Like those before me, I begin with an insight from Mikhail Bakhtin, who described the fabric of fictional worlds as the entwining of time and space, or chronotope—a place "fraught with time."[17] The warp and woof of fictional space-time entwines also in the real world, and is therefore also subject to the laws of the real, as opposed to an imagined, universe. In this way, a novel may span several continents in diegetic space while occupying only a few square inches on a reader's desk; a tale of a century might take up only a few hours of a reader's time.[18]

Theoretically, changes in the structure of time and space in the one realm should refract changes in the other. Fredric Jameson wrote, for example, about the "emergence of a new space and a new temporality" related to the "philosophical programme of secularization and modern-

ization" of the Enlightenment.[19] According to Jameson, the realist novel of the nineteenth century reflects new modes of cultural production, which mirror broader socioeconomical changes related to industrialization. Roland Barthes similarly mentioned the emergence of what he calls "narrative luxury," a kind of a superfluous proliferation of notable details that accompanies the effect of fictional realism.[20] Dames et al. quote Franco Moretti, who likewise wrote about a "shifting center of gravity" in the transition from the rural picaresque to the Bildungsroman.[21] According to Moretti, the move of fictional protagonists from rural to urban spaces resulted in a more complex public sphere, in which "the narrative system becomes complicated, unstable." "Quantity has produced a new form," Moretti concluded, although the empirical results suggest otherwise.[22]

The broad historical scope of these hypotheses warrants the proportionate deployment of quantitative "distant reading" methods. When Jameson and others posit the emergence of qualitatively new kinds of descriptions, they are making an argument at scale, which speaks to an aggregate rather than an individuated phenomenon. If the close reading of singular representative passages can approximate that historical pattern, the hypothesis of large-scale systemic change should also admit evidence in aggregate.

The methodological difficulty of modeling fictional space presents several interesting theoretical problems. We intuit that any account of high-level systemic changes in the quality of narrative space must rest on a quantity of low-level linguistic observations. Unlike real spaces, however, fictional spaces defy conventional notions of size or magnitude. For this reason, defining space in terms of explicit magnitudes, settings, or frames—as is often done in narratological theory—is insufficient for our purposes.[23] It is not enough also to rely on identifiable geography, as Moretti does in his *Atlas of the European Novel*. The explicit framing approach privileges sparse, structured, and geographically specified worlds—Jules Verne's *Around the World in 80 Days*—while failing to account for the richness of more localized, amorphous, or domestically dense narratives—Marcel Proust's *Swann's Way* or Franz Kafka's *The Metamorphosis*. How are we to survey such heterotopic imaginary expanses? The challenge lies in "the possibility of spatial modeling through concepts that are not in themselves spatial."[24]

To bridge quality and quantity—the subjective sense of fictional space and its objective properties—I turn to the concept of *Umwelt*, or perceptual environment, proposed by the biologist Jakob von Uexküll and later developed by Thomas Sebeok. Uexküll wrote: "Every subject spins out, like the spider's threads, its relations to certain qualities of

things and weaves them into a solid web, which carries its existence."[25] Consequently, one is able to imagine an animal's *Umwelt* by examining its physiological capabilities of perception. In a paradigmatic example, Uexküll discussed the sensory organs of the common wood tick, which is blind but able to perceive tactile "collisions," changes in temperature, and the presence of butyric acid. These simple "perceptual signs" [*Wirkzeichen*] comprise the animal's "effect space" [*Wirkraum*]. The stimuli "glow like signal lights in the darkness and serve as directional signs that lead the tick surely to its target."[26]

The light of perceptual signs illuminates fictional spaces as well. By isolating it, a reader can better characterize the topography of fiction, not in terms of measurements (such as miles or feet) or named entities (such as city or street names), but in subjective terms, as objects of attention. These radiate from the subject outward.

Consider, for example, the initial impressions of Gregor Samsa's monstrous transformation in *The Metamorphosis*: "He lay on his armour-like back, and if he lifted his head a little he could see his brown belly, slightly domed and divided by arches into stiff sections. The bedding was hardly able to cover it and seemed ready to slide off any moment. His many legs, pitifully thin compared with the size of the rest of him, waved about helplessly as he looked."[27] The description proceeds from the sense of proprioception, literally the "grasping of self," the awareness of one's own body. Gregor finds himself on his back: he lifts his head to observe his legs, the "domed arches," and the "stiff sections" of his abdomen. The discomfort of the situation is reinforced by a curious painting hanging above Gregor's work table, depicting "a lady fitted out with a fur hat and fur boa who sat upright, raising a heavy fur muff that covered the whole of her lower arm towards the viewer." The muff is particularly disturbing in that it obscures the human shape, elongating and rendering the woman's limbs monstrous under animal pelt. Gregor's corporeal discomfort is redirected toward the viewer of the picture and hence the reader, in a mimetic displacement of beetle appendages. Kafka further engages the senses in his suggestion of tactile fabric samples (spread on the table), the sound of rain hitting glass, a sense of dull pain, an itch, and a cold shudder. A cursory look at the story's initial perceptual surroundings reveals a semantic chain of nouns and adjectives related to the sense of a body in distress.

Diegetic expanses—Uexküll's "effect spaces"—stretch between "things that are important." Note that, for the purposes of a general survey, it is not necessary to distinguish between a narrator's and a character's point of view. The spotlights of narrative description pick out distinct objects, which the reader subsequently weaves together into a unified locality,

"filling in" the gaps between Gregor's bed, chest of drawers, table, window, and door. The size of his room is not limited to direct observation, "a proper human room although a little too small." It takes shape in the density of available things, the "directional signs" that bear the weight of narrative. Gregor wakes, turns, slides, falls, crawls, and stands amidst a crowded domestic space, filled with large, unwieldy things. The clutter reinforces the impression of spatial constraint.

Literary Archaeology

Thinking of space in terms of objects rather than dimensions facilitates an approach to the theory of the novel at once materialist and phenomenological. We can now revisit our initial intuitions about urbanization to posit a more robust account of the "empty feeling" that Jameson attributes to the "realist floor-plan."[28] My methods derive from two primary sources: the ethnoarchaeology of domestic space and the early "bag-of-words" experiments by the Russian formalist Iarkho.

In *The Meaning of Things*, an ethnography of domestic space published in 1981, Mihaly Csikszentmihalyi and Eugene Rochberg-Halton proposed taking stock of an "ecology" of things, which "reflects as well as *shapes* the pattern of the owner's self."[29] The authors surveyed physical objects based on interviews in the field. The resulting inventories included detailed counts of "special objects": their frequency of mention during interviews, demographic (social class, age, gender) differences in the distribution of objects mentioned, and the "number and percentages of meanings associated with acquisition categories."[30] In this way, among the superset of all things that surround people at home, researchers isolated those objects that their subjects considered particularly important. These inventories approximate Uexküll's "effect spaces," in that they identify those perceptual signs that capture a subject's attention.

In 2012, a group of UCLA anthropologists used similar methods to characterize "life at home in the twenty-first century." Going beyond questionnaires, researchers made extensive site visits, documenting a number of middle-class households to create what the authors call an "ethnoarchaeology of modern material culture."[31] The study relied on a "simple set of time-tested archaeological and observational methods to record and then critically analyze the domestic material world of U.S. households today."[32] The authors wrote: "Our research design called for mapping, intensive photography of virtually everything material in people's homes. . . . The information we recovered is systematic rather than anecdotal or confined to single cases so as to maximize its

explanatory power."[33] Of particular interest to the characterization of fictional space is the archaeological concept of "material saturation." Similar to Barthes, Jameson, and Moretti, the anthropologists began their observations with a historical hypothesis, writing that average European households of the nineteenth and twentieth centuries were "sparsely appointed" in comparison with the "mountains of possessions" often found in modern American homes: "Even the relative excesses of domestic property that were common during the Victorian period, when it was fashionable to add rugs, mirrors, paintings, and overstuffed chairs to crowded parlors, truly pale by comparison to the total possessions of average families today in the U.S."[34]

To quantify a measure of "tangible artifacts" in a household, the authors proposed a metric they call "material saturation." The protocol was described as follows: "Trained coders assigned every photographed object to an overarching category (such as furniture, media electronic, decorative item, or toy) and then directly counted (for most categories) or estimated (for abundant items such as books, CDs, or toys) the number of such items present, room by room. These counts are essential because they provide firm quantitative evidence of the material richness and diversity of modern American homes."[35] These metrics are compelling because they offer evidence for seemingly intractable, large-scale historical trends, such as the "rise of consumer culture," through a number of reliable microempirical observations, such as "visible possessions per room." Given the related difficulty of characterizing the "superfluous proliferation of notable details" in literature, as per Barthes, I propose a parallel approach for narratological analysis. Changes in material density of diegetic space can be used to test our intuitions about the novel as it relates to theories of realism, consumerism, or urbanization.

A detailed survey of a fictional world presents itself readily in narrative description. A literary archaeologist can inventory mentioned objects to create metrics of material density per location (a city or a room) or per textual unit (a chapter or an arbitrary number of words). The resulting metrics can be used to characterize narrative space systematically, both in close reading single novels (responding to narrative development across chapters, for example) or in distant reading across corpora. The formula is reductive to the extent that it attempts to isolate a signal sensitive to the underlying changes in material culture. An approximation of what counts for an "average" number of perceptual signs helps articulate a sense of "superfluous" detail and luxury by contrast. Derived metrics such as these present useful markers for further investigation, in the way a fever marks an important symptom of a complex medical etiology.

The articulation of "effect spaces" via material density presents the additional complexity associated with literary representation. To take an inventory of things in fictional worlds we must convert grammatical or lexical categories, such as sentence predicates or parts of speech, into semantic ones, such as objects and subjects. This is difficult to do convincingly, because meaning-making involves a manifold and nondeterministic chain of causes and effects. As the saying goes: "Sometimes you eat the bear and sometimes the bear eats you."

An archaeology needs a method for recovering things. For this purpose, I turn to the work of Iarkho, a little-known, but increasingly important early twentieth-century Russian formalist literary critic, classicist, medievalist, and member of the Moscow linguistic circle. Iarkho's quantitative experiments were all the more remarkable for being composed under severe repression from the Soviet regime in the years between 1919 and 1942. Writing in forced exile from a small Siberian town, Iarkho advanced an expansive philological program, considerably more grounded in descriptive statistics than the work of his contemporaries.

A number of his methodological proposals anticipate modern statistical techniques, made popular by computational means. The so called "bag-of-words" approach to text classification dates back to the rather technical problem in information science of retrieving a subset of documents relevant to a given search term. The widespread use of digital knowledge management systems—library catalogs and computerized indices—in the 1970s sparked general interest in "term-frequency" analysis. For example, a library patron wishing to retrieve documents related to the term "beverage" would also likely be interested in articles on "coffee" and "tea," even when these do not explicitly mention "beverage" in the body of the text.[36] For these purposes, a text can be treated as a loose collection of terms—in the words of Karen Spärck Jones, a pioneer of the technique, "a fine mixed bag."[37]

The growth of computational power and the increasing availability of digital materials have precipitated the development of complex classification models that use modern statistical methods such as Gibbs sampling and Markov chains, among other probabilistic approaches. Topic modeling in particular has been applied widely to the study of literary texts, following the insight that many epistemological problems of category-formation, such as genre or period attribution, can be reduced to topic classification. David Blei explains with characteristic clarity: "Topic modeling algorithms are statistical methods that analyze the words of the original texts to discover the themes that run through them, how those themes are connected to each other, and how they change over time."[38]

Iarkho's methods were less complicated than their modern counterparts, relying on straightforward, descriptive word frequency tabulations. They present more than historical interest, however, because they are tightly connected with longstanding philological practices, from which Iarkho derived his hermeneutics. For example, in a 1928 article in *Speculum*, Iarkho examined several ninth-century works by Sedulius Scottus, placing them in an unbroken tradition of similar texts by earlier Carolingian poets.[39] He supplemented his close readings with vocabulary charts, which cluster related concepts under categories of "plant kingdom" [*Pflanzenreich*], "animal kingdom" [*Tierreich*], and "emotions" [*Gefühle*].[40] The overlap in vocabularies evidenced literary influence.

In his methodological treatise "Comedies and Tragedies of Corneille: A Study on the Theory of Genre," Iarkho proposed a similar "bag-of-words" technique to differentiate between the comedies and tragedies of Pierre Corneille, the seventeenth-century French dramatist. Employing a method analogous to modern sentiment analysis, Iarkho made lists of affective terms, which he further reduced to a smaller number of broader semantic categories, such as "anger" or "happiness." Like a modern data scientist, Iarkho began by "tokenizing" and "stemming" his corpus, in a process by which texts are broken up into individual unique words ("tokens") and then into their related cognates, lemmas, or word stems. By these means the words "flame," "conflagration," and "inferno"—to give an English analogy—can be reduced to the same root canonical terms.[41] Iarkho then clustered more than 400 such stems into topic hypernyms such as "sexual love," "happiness," "fear," and "suffering." He used these "affective groupings" to find those that correlate strongly with "drama" and "comedy" categories.[42] Finally, after evaluating more than thirty distinct features by which the two genres could be differentiated, he picked out those that are most "typical," using them discuss the evolution of genre in historical context.[43]

Mikhail Gasparov, the (late) contemporary Russian philologist who has done much to rehabilitate Iarkho's legacy in Russia, has simplified the method further, calling it "immanent analysis, of the kind that does not violate the boundaries of a text."[44] In what essentially is a methodological statement derived from Iarkho, Gasparov described a "mechanical method of reading poetry" aimed at converting a reader's intuitions into observable features of the text, in a kind of a mechanically deconstructive reading. It begins with the assumption that literary motifs and plots [*siujety*] are constituted through "figures," defined as "affectively imagined things or persons, i.e. potentially every noun." Consequently, a motif is "any action, i.e. potentially every verb," where a *siujet* comprises "a sequence of related motifs."[45]

Gasparov used these definitions to "deform" two short lyrics by Pushkin, "Premonition" and "Anchar."[46] Like Iarkho, he separated nouns, adjectives, and verbs, which he then clustered into semantic categories. Several patterns emerged. For example, Gasparov found that abstract, affective vocabulary dominates the inward-looking "Premonition." Here, Pushkin creates a world "almost without dimensions," due to a lack of spatial nouns.[47] By contrast, in my own, Gasparov-inspired reading of the text, "Anchar" emerges as an exercise in Russian colonial imagination. Verse by verse, the poem expands in multiple dimensions: horizontally across the desert in the first verse, down to the roots of the poisonous Anchar tree in the second verse, from the inside to the outside of the tree in the third, up toward the sky in the fourth, and away from the tree in the fifth (Appendix I). The sixth stanza marks a change in perspectives. A ruler sends his vassal on an arduous journey toward the tree. The man returns with its poisonous resin in the seventh stanza. He then lies down and dies at the feet of his master, who in turn sends a multitude of arrows in all directions outward, "onto his neighbors in distant lands."[48]

The poem's diegetic space is never explicitly framed. Pushkin rather gives us a number of circulatory systems. The dominant grammatical device is one of prepositional traversal. The sense of a *Werkraum* (perceptual space) is derived from subjects or substances moving toward, through, away, or across. Space stretches in movement: the movement of water from up high down to the tree, of sap that rises up from the desert through the plant, of poison imparted to the rushing wind, of a vassal sprinting from the plant to his master, and of poisonous arrows that bring destruction from the master to his neighbors. The circulation of elements creates effect spaces on scales both large and small, open and closed, internal and external. The desert is as much of a setting of the poem as is the capillary system of the plant itself.

Diegetic Density and Clutter Distance

The combined use of these literary-theoretical and anthropological insights leads to a method for literary archaeology. I am now ready to transpose our initial intuitions about the changing nature of fictional space into a minimally viable, analytical approach suitable for computational analysis. First, following Uexküll, Csikszentmihalyi, and Arnold et al., I propose to represent diegetic space in terms of density of perceptual objects.

I further assume that whatever is meant by "perceptual objects" can be regularly approximated by corresponding grammatical and semantic categories, in this case expressed in the Stanford Typed Dependencies Representation (SD) scheme and the WordNet database.[49] In trawling for salvageable goods, my aim is not to recover everything, but to recover most. Recall—the completeness of the results—can be improved in further iterations of the algorithm.

I am initially interested in direct and indirect objects of a verb (*dobj* and *iobj* in the SD schema). The noun extraction pipeline involves additional steps to remove nonhuman nominal subjects, passive nominal subjects, and conjuncts, which I label according to the Stanford Typed Dependencies Manual.

In light of Gasparov's methods, we must also be interested in prepositional objects (*pobj*) of the sentence: language often structures space through prepositions.[50] When Gerard Manley Hopkins writes that "man's mounting spirit in his bone-house, mean house, dwells," he means to say something about the nominal subject, "man's mounting spirit," which relates to the prepositional object "bone-house, mean house." The preposition (*in*) places the noun phrases in spatial accord to one another. The spirit at once shares a space with its "bone-house" and is on the interior.[51]

The final list of nouns and noun phrases encompasses most possible objects and locations, while discarding grammatical relations such as adverb modifiers, parataxis, phrasal verbs, temporal modifiers, and other elements that do not usually mark material culture. Consider the following, more complex example, from Emily Brontë's *Wuthering Heights*, where, if you recall, Mr. Lockwood finds himself reluctantly accepting guest lodgings, described as follows:

> Too stupefied to be curious myself, I fastened my door and glanced round for the bed. The whole furniture consisted of a chair, a clothes-press, and a large oak case, with squares cut out near the top resembling coach windows. Having approached this structure I looked inside, and perceived it to be a singular sort of old-fashioned couch, very conveniently designed to obviate the necessity for every member of the family having a room to himself. In fact, it formed a little closet, and the ledge of a window, which it enclosed, served as a table. I slid back the panelled sides, got in with my light, pulled them together again, and felt secure against the vigilance of Heathcliff, and every one else.[52]

Following the described heuristic above, in the first pass, I extract a list of perceptual objects based on their grammatical category. In the subsequent pass, I filter the list semantically, relying on several "supersenses," also referred to as "hypernyms" and "lexicographer file names," made

available by the Stanford WordNet lexical database.[53] From the list of twenty-six available noun supersenses, which include *feeling*, *motive*, and *phenomenon*, I keep the ones likely related to perceptual objects, those having physical dimensions and those capable of anchoring space: *artifact*, *food*, *possession*, *object*, *substance*, *animal*, and *plant*. I discard nouns belonging to categories of *person*, *communication*, *cognition*, *state*, *time*, *attribute*, *process*, *phenomenon*, *motive*, *feeling*, *shape*, and *relation*. The resulting inventory gives a reasonable approximation of the effect space. In the case of Mr. Lockwood (in the paragraph quoted above) we obtain the following results:

Table 1. Diagnostic noun inventory based on a passage sample from Wuthering Heights.

Noun Phrase	Grammatical Category	Super-sense Label
door	dobj	noun.artifact
bed	pobj	noun.artifact
chair	pobj	noun.artifact
press	conj	noun.artifact
case	conj	noun.artifact
square	pobj	noun.location
window	pobj	noun.artifact
structure	dobj	noun.artifact
couch	pobj	noun.artifact
necessity	dobj	noun.object
room	dobj	noun.location
closet	dobj	noun.artifact
ledge	nsubj	noun.object
window	pobj	noun.artifact
table	pobj	noun.artifact
side	nsubj	noun.location
light	pobj	noun.artifact

The list is not perfect, containing words like "necessity," for example, which falls under the "noun.object" category in one of its many definitions contained in the lexical database. However, the preliminary results show that a noun-based model of effect space is robust enough to tolerate a measure of ambiguity, since it captures the majority of perceptual objects apparent in the close reading of the passage. Any bias implicit in the heuristic will be consistently distributed across the collection of

analyzed texts. More complex, machine-learning-based approaches may be used to further purge the inventory, although the added precision is not necessary at this stage. Nouns signifying location ("room" and "square") and named geographical entities ("Yorkshire" and "Wuthering Heights") are also programmatically extracted for later use at this time.

In the second stage of my analysis, I propose two major metrics derived from the observational results above. The first is clutter distance, roughly signifying "average words per thing." A clutter score of 100 words per thing would mean that the reader should expect to encounter, on average, a single perceptual thing every 100 words. The related average unique clutter distance metric, or *u-clutter*, provides a similar score per uniquely named thing. We can imagine, for example, a fictional setting that contains many objects, but few uniquely-named ones (a world full of pencils, for example). This would indicate conceptual paucity in conditions of material abundance.

The clutter metrics posit a relationship between "story space," measured in units of text, and "discourse space," measured in semantic units that denote perceptual objects.[54] Another way to think about clutter distance is to imagine highlighting all thing-related words on a page: the smaller the index of clutter, the fewer words there would be between each encountered thing. Visually, the page would look cluttered with highlights.

Table 2. Derived metrics for clutter distance and diegetic density.

Metric	Short Name	Units
clutter distance	*clutter*	words per thing
unique clutter distance	*u-clutter*	words per unique thing
diegetic density	*diedensity*	total things per unique location
unique density	*u-diedensity*	unique things per unique location

In contrast to clutter distance, the diegetic density metric is expressed in purely intra-diegetic, discursive terms, without reference to text dimensions. Diegetic density characterizes fictional space as the number of total things per uniquely named location, where a unique name includes both general nouns in the "location" category (such as "room" and "square") and named geographic entities (such as "Yorkshire" and "Wuthering Heights"), both extracted alongside perceptual objects in the

first stage of the analysis. Likewise, unique diegetic density summarizes the average number of uniquely named things per unique location. In the case of clutter distance, pages are cluttered; in the case of diegetic density, the fictional spaces are dense.

The above heuristics are not yet meaningful and should be viewed as methodological building blocks, not as theories in themselves. They can aid in the description of qualitative change in the texture of the narrative fabric within or between texts. In my initial applied experiments, they showed promise in distinguishing between "urban" and "rural" novels.

Urban Novels Reconsidered

For the purposes of testing, I extracted the perceptual inventory from the same sixty-four Victorian novels used for classification in the "Extracting Social Networks" project. This corpus contains more than 11,000,000 words and has the advantage of being collected and classified by an independent team. Assumptions that have gone into the density metrics are therefore isolated from the classification process, avoiding self-verification bias. The original team of researchers classified each novel as having a "rural" (0), "urban" (1), or "mixed" (2) setting, and as written in first-person (1) or third-person (3) perspective.

In the final stage of my analysis, I derived density and clutter metrics for each of the novels in the corpus. I then obtained simple correlation statistics between things and settings (urban and rural), and things and perspectives (first or third).

The results are instructive. The initial, exploratory visualizations suggest a relationship between diegetic density and urban setting and between *clutter distance* and perspective, as evidenced by the first and third plots in Figure 1.[55]

Box plots of the same data solidify the above intuitions more formally. Urban novels appear to be on average more dense than rural ones, in the sense that urban spaces contain more objects in total per unique location. Note that urban spaces do not contain more unique objects per unique location. This implies that rural spaces are no less semantically rich than urban ones. The possessions are rather more densely situated. Characters in urban fiction "bump" into or interact with their things more often. However, they generally contain a similar typology of things. As you can see from the third and fourth box plots in Figure 2, a reader of urban novels should expect to encounter roughly the same number of things per word as a reader of rural novels. Things, to put it in terms suggested by Seymour Chatman, are no more privileged in the "discourse space" of the urban novel. Rather the "story spaces" contained within are more dense and hence more constrained.

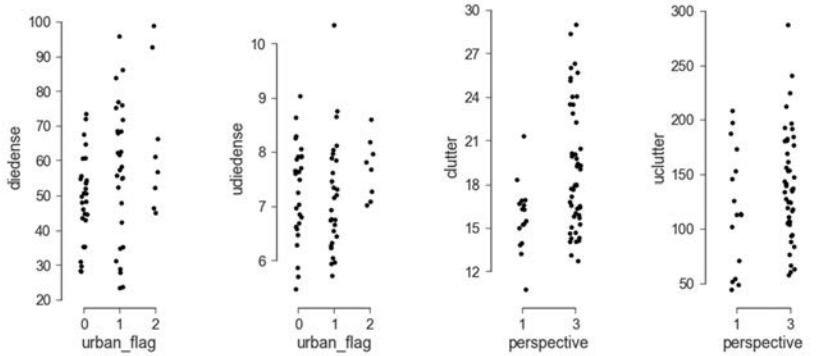

Fig. 1. Scatter plots exploring the relationship between diegetic density and the urban novel, and between clutter distance and perspective. Urban_flag represents the rural (0), urban (1), and mixed (2) categories. Perspective is grouped into first-person (1) and third-person (3) perspectives along the x-axis. Grouping differences in the first and third plots indicate some correlation.

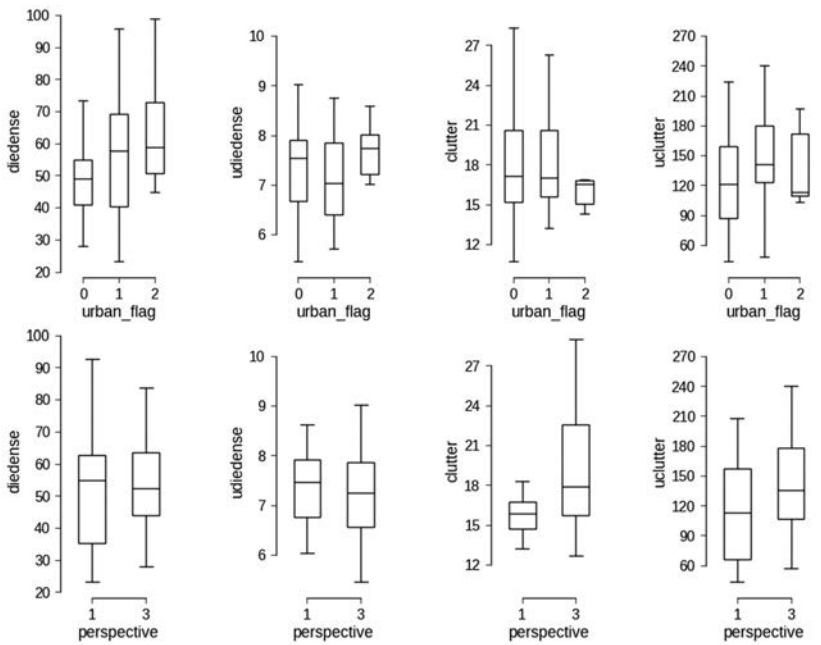

Fig. 2. Box plots further formalizing the correspondence between diegetic density and the urban novel, and between clutter distance and perspective. Urban_flag represents the rural (0), urban (1), and mixed (2) categories. Perspective is grouped into first-person (1) and third-person (3) perspectives, along the x-axis. A positive slope of an imaginary line between the boxes suggests a relationship.

Curiously, the opposite dynamic holds true for perspective, as seen in the second row of Figure 2. Diegetic density is not terribly responsive to perspective. Novels told in first- and third-person perspectives contain spaces that are, on average, populated with a similar amount of stuff. However, the reader of third-person-based novels will encounter significantly fewer things per word than the reader of first-person-based novels. The floorplans, in other words, are similarly populated, where the spotlight of the more singularly subjective narrative takes more frequent note of its material surroundings. First-person subjects project a denser, more individuated sense of space; they weave a tighter web on a page.

These visual observations can further be expressed in quantitative terms. Since our problem is one of correlation and not classification, I have used point-biserial correlation metrics, which produce robust results when dealing with association between continuous and categorical variables; in our case, the derived metrics of diegetic density and noun clutter, binary urban or rural settings, and first- or third-person perspectives.

The following table contains a list of possible relationships: their corresponding correlation coefficient and p-values, with most significant findings highlighted in bold.[56] The correlation coefficient of 0.313, on the scale from 0–1, implies a modest degree of relationship between diegetic density and the urban novel, where 0.8 and above can be characterized as strong and 0.1 and below as weak. These results further have a highly significant p-value (0.011), meaning that assuming the null-hypothesis is true (there is no correlation), the model would be unlikely to produce these or more extreme results by chance. The relationship between clutter distance and perspective is similarly compelling: a modest relationship with an even higher degree of certainty (0.006).[57]

The strength of these findings lies not, however, in low p-values or other metrics, but in the methods used to produce them. Everything potentially correlates with everything. These particular correlations give credence to the explanatory theory that connects high-order intuitions about "the changing nature of fictional space" to aggregate low-level observations based on grammatical and semantic categories. We intuit that this connection exists at the outset; critics routinely posit longue durée theses about the changing nature of material representation in literature, based on paradigmatic, "representative" examples. A more robust theory of fictional space can begin to chart the pathway by which such claims are possible, based on numerous observations. The absence of a relationship between diegetic density and the urban novel would, at the very least, force a reconsideration of the proposed explanatory framework. The metrics are diagnostic, in other words, not necessarily

Table 3. Likely degrees of relationship and significance.

METRIC	DEPENDENT VARIABLE	CORRELATION COEFFICIENT	P-VALUE
density	**urban-rural**	**0.313**	**0.011**
u-density	urban-rural	0.083	0.512
clutter	urban-rural	-0.110	0.388
u-clutter	urban-rural	0.125	0.324
density	perspective	-0.108	0.394
u-density	perspective	-0.118	0.352
clutter	**perspective**	**0.338**	**0.006**
u-clutter	perspective	0.134	0.291

confirmatory in a positivist, reductive way. They attain their full significance in parallel with other hermeneutical strategies, such as close and descriptive reading, historical contextualization, or intertextual analysis.[58]

Although in this paper I am primarily interested in a theory of fictional space and in the methods of characterizing it, the diagnostic metrics already suggest a few tantalizing possibilities for applied analysis. For example, the novels of Charles Dickens are by far the most diegetically dense texts in our collection, with *The Pickwick Papers*, *Bleak House*, and *David Copperfield* all scoring above 90 where the median is 54. This might not come as a surprise to attentive readers of Dickens, an author whose novels "come alive" with "furniture, textiles, watches, and handkerchiefs."[59] The Victorian novel, and Dickens in particular, "showers us with things," in what Elaine Freedgood (quoting Hippolyte Taine) describes as "metonymic madness," where the proliferation of things begins to "swamp" and overwhelm the author.[60] A detailed inventory of Dickensian worlds reveals an abundance of handkerchiefs, doors, beds, windows, chairs, coats, cases, boxes, boots, books, hats, and candles (all these in the top quartile of most commonly occurring objects in Dickens, according to my tabulations).

The other end of the density spectrum is perhaps more unexpected. The urban novels of Robert Louis Stevenson and Sir Arthur Conan Doyle occupy a place diametrically opposed to Dickens, at the bottom of the scale, with scores hovering around 20. In Doyle's detective fiction in particular, we expect an emphasis on material minutia. Instead, the descriptive passages such as the one that follows are more common.

We met next day as he had arranged, and inspected the rooms at No. 221B, Baker Street, of which he had spoken at our meeting. They consisted of a couple of comfortable bedrooms and a single large airy sitting-room, cheerfully furnished, and illuminated by two broad windows. So desirable in every way were the apartments, and so moderate did the terms seem when divided between us, that the bargain was concluded upon the spot, and we at once entered into possession. That very evening I moved my things round from the hotel, and on the following morning Sherlock Holmes followed me with several boxes and portmanteaux. For a day or two we were busily employed in unpacking and laying out our property to the best advantage. That done, we gradually began to settle down and to accommodate ourselves to our new surroundings.[61]

The passage is remarkable for its lack of material texture. Baker Street 221B consists of "bedrooms" and a "sitting-room," "cheerfully furnished." The pair "comes into possession" of the apartment on "moderate terms." Watson "moves his things in" and Holmes follows with "boxes and portmanteaux." The locality is vague, described in terms of general "property" and "surroundings." Passages like these, not the ones in which Holmes picks out careful clues, prevail in Doyle's fiction. Such paucity may be explained by the episodic concentration on the "important details," or clues in detective fiction, clustered around the scene of a crime.[62] Exploratory data analysis shows some episodic clustering, although not unusually so when compared, for example, to *Bleak House* by Dickens (Fig. 3).

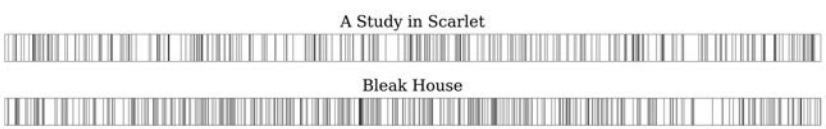

Fig. 3. Clutter distance plots exploring object clustering in Doyle's *A Study in Scarlet* and Dickens's *Bleak House*. Each bar represents a thing encountered in the linear progression of the novel. In both works, object description comes in batches, although the Dickensian page is more thickly populated.

The distinction between diegetic density, expressed purely in the terms of a fictional world, and clutter distance, expressed in terms that relate textual space (words on a page) to semantic units (nouns expressing things) also yields notable results. Novels by Jane Austen occupy three of the bottom four novels ranked by average clutter distance. The reader should expect to encounter an object every 29 words, where the mean is 18. Rural topography in Austen's novels is expansive, described in terms of "spaces," "lawns," "towns," "properties," "views," and "fields." Austen's narrative is concerned primarily with people, not with things.

Those looking for density in the social network sense would likely find it here. Austen's room descriptions focus on people. Consider the following excerpts, which crowd around entrances:

> The party entered the assembly-room, it consisted of only five altogether: Mr. Bingley, his two sisters, the husband of the eldest, and another young man.[63]

> But when the gentlemen entered, Jane was no longer the first object; Miss Bingley's eyes were instantly turned toward Darcy, and she had something to say to him before he had advanced many steps.[64]

> Till Elizabeth entered the drawing-room at Netherfield, and looked in vain for Mr. Wickham among the cluster of red coats there assembled, a doubt of his being present had never occurred to her.[65]

> Almost as soon as I entered the house I singled you out as the companion of my future life.[66]

These passages are not sparse in the sense of generic description found in Doyle. Rather, objects are rare in the text where the spotlight of narrative perception prefers to shine on and illuminate human subjects. Such medium-length observations—neither close nor distant—are exploratory. They lead to reading and rereading. I include them here as corroborative evidence for the immediate effectiveness of density statistics as tools for both synchronic and diachronic analysis.

Finally, material density primitives convey historical trends, not immediately apparent in the analysis of individual texts. In Figure 4, I plot diegetic density over time to produce a trend line using nonparametric locally weighted (LOWESS) regression.[67] The resulting figure contradicts scholarly expectation. The conventional story of the long nineteenth century is one of gradually intensifying commodification, in which capitalism "produced and sustained a culture of its own," and where descriptions of things begin to "overtake and dominate" other forms of literary description well into the twentieth century.[68] However, data show a marked decline in descriptive material culture (in terms of diegetic density), which begins mid-century and continues unabated toward the centennial boundary. These results are pronounced enough to merit further investigation. Although representative of the nineteenth-century canon, our sample size is small. Further work must be done to ensure accurate sampling in order to draw more definitive conclusions.

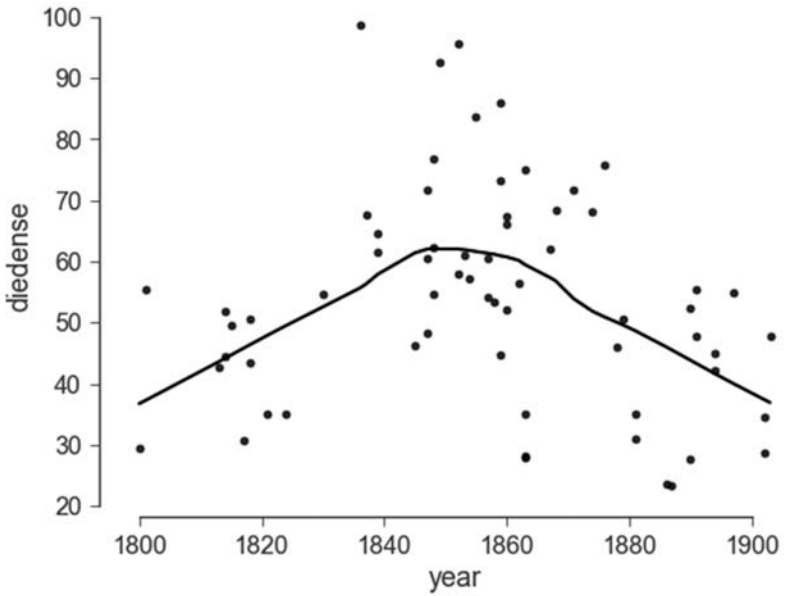

Fig. 4. A time-series that plots chronologically the diegetic density of nineteenth-century novels. Critical literature posits an upward slope, where we observe a mid-century decline.

Concluding Remarks

The primary theoretical contribution of this paper is in the reconceptualization of fictional space in terms of things rather than forms or framing devices. Space unfurls between objects. It takes almost 4,000 words for Gregor Samsa to move from his bed to the dresser. A whole perceptual universe is born and transpires in that span.

Space can be understood through things. An inventory of things contained in novels characterizes space and therefore comprises a method of literary-historical archaeology. The challenge for a literary archaeologist remains the modeling of pathways between things and their representations. A thick description of the diegetic effect space explicates the mechanisms by which objects in the real world are converted into linguistic, cognitive phenomena. A computational approach to recovering material artifacts in volume further extends local insight across time and corpora, corroborating longterm historical changes in material culture.

In light of these postulates I have proposed two descriptive metrics: diegetic density, defined in terms of things per unique location, and

clutter distance, defined in terms of words per thing. These narratological primes are atomic in that they form small, stable configurations that embody the chronotope. Through their use, literary-sociological models can become more parsimonious and therefore more evident, that is, open to critical scrutiny and refinement. The approach has led me to several preliminary but significant findings about space in the nineteenth-century novel.

COLUMBIA UNIVERSITY

Appendix I: Pushkin's "Anchar" in translation

TRANSLATION BY THE AUTHOR. ORIGINAL FROM PUSHKIN, 1959.

АНЧАР	ANCHAR
В пустыне чахлой и скупой	An Anchar tree stands stern.
На почве, зноем раскаленной,	A sentinel of empty, burning grounds,
Анчар, как грозный часовой,	it stands alone, amidst a stingy
Стоит - один во всей вселенной.	and emaciated plain.
Природа жаждущих степей	Born of a thirsty steppe,
Его в день гнева породила,	in days of wrath, its lifeless greenery,
И зелень мертвую ветвей	its roots and branches, imbibed
И корни ядом напоила.	the desert's poison.
Яд каплет сквозь его кору,	Its toxic sap melts in the midday heat
К полудню растопясь от зною,	and seeps through the tree's bark.
И застывает ввечеру	At night, the liquid stiffens back
Густой прозрачною смолою.	into a thick, transparent resin.
К нему и птица не летит,	Birds do not visit the Anchar,
И тигр нейдет: лишь вихорь черный	tigers won't come near, only a black wind
На древо смерти набежит -	will sometimes rush this deadly tree
И мчится прочь, уже тлетворный.	and race away polluted.
И если туча оросит,	And if a cloud were to lose its way
Блуждая, лист его дремучий,	to shed its dew onto the tree's dense foliage,
С его ветвей, уж ядовит,	a toxic rain would trickle down
Стекает дождь в песок горючий.	from the branches to the searing sand.
Но человека человек	A man sent forth a man
Послал к анчару властным взглядом,	with a commanding look. The last obeyed
И тот послушно в путь потек	to brook his course towards the tree,
И к утру возвратился с ядом.	returning with the poison in the morning.

Принес он смертную смолу	He brought the deadly resin
Да ветвь с увядшими листами,	and a wilted branch.
И пот по бледному челу	Cold sweat
Струился хладными ручьями;	Streamed from his pale face.
Принес - и ослабел и лег	He brought his gift, grew weak, and sank
Под сводом шалаша на лыки,	into the rugs under the great tent's dome.
И умер бедный раб у ног	So died the poor vassal at the feet
Непобедимого владыки.	of his unvanquished master.
А царь тем ядом напитал	The czar imbued compliant arrows
Свои послушливые стрелы	with that venom.
И с ними гибель разослал	He then dispatched destruction
К соседям в чуждые пределы	onto his neighbors in distant lands.

NOTES

I thank Nicholas Dames, Sierra Eckert, Jonathan Reeve, Milan Terlunen, and other members of Columbia University's Literary Modeling and Visualization lab for their comments on an early draft of this paper.

1 William Shakespeare, *Hamlet: Revised Edition*, ed. Ann Thompson and Neil Taylor (London: Bloomsbury Arden Shakespeare, 2016), 496.

2 Samuel Beckett, *Waiting for Godot: A Tragicomedy in Two Acts*, bilingual ed. (New York: Grove, 2010), 3.

3 Stanley Fish, "What Is Stylistics and Why Are They Saying Such Terrible Things about It?-Part II," *boundary 2* 8, no. 1 (1979): 132, 144. See also Fish, "Literature in the Reader: Affective Stylistics," *New Literary History* 2, no. 1 (1970): 123–62; and Fish, "What Is Stylistics and Why Are They Saying Such Terrible Things About It?" in *Is There a Text in This Class? The Authority of Interpretive Communities* (Cambridge, MA: Harvard Univ. Press, 1980), 21–68.

4 Karl R. Popper, *The Logic of Scientific Discovery* (New York: Harper & Row, 1968), 40.

5 Popper, *Scientific Discovery*, 40.

6 In his "Observations, Explanatory Power, and Simplicity," Richard Boyd writes about the excess of the scientific method as follows: "The standards for theory assessment . . . required by those features of scientific methodology are, at least apparently, so different from those set by the requirement that the predictions of theories must be sustained by observational tests that it is, initially at least, puzzling what they have to do with the rational scientific assessment of theories or with scientific objectivity." See Boyd, Philip Gasper, and J. D Trout, *The Philosophy of Science* (Cambridge, MA: MIT Press, 1999), 349.

7 Clifford Geertz, *The Interpretation of Cultures: Selected Essays* (New York: Basic Books, 1973), 5.

8 See, for example, Marshall McLuhan, "Footprints in the Sands of Crime," *The Sewanee Review* 54, no. 4 (1946): 617–34.

9 Recall and precision metrics for document selection were introduced by the Cranfield Research Project in the 1960s. See Jean Aitchison and Cyril Cleverdon, *A Report on a Test of the Index of Metallurgical Literature of Western Reserve University* (Cranfield, UK: College of Aeronautics, 1963). See also Yiming Yang, "An Evaluation of Statistical Approaches to Text Categorization," *Information Retrieval* 1, no. 1–2 (1999): 69–90.

10 Imagine classifying texts as "detective" or "romance" novels and getting many false positives (low precision), or writing an algorithm that was very precise but returned only a small fraction of possibly correct results (low recall).

11 J. Racz and A. Dubrawski, "Qualitative Pose Estimation Using an Artificial Neural Network," in *ICAR '95: Proceedings of the 7th International Conference on Advanced Robotics: September 20–22, 1995, Saint Feliu de Guíxols, Catalonia, Spain* (Barcelona: Universitat Politècnica de Catalunya, 1995).

12 In a recent article titled "Attributing the Authorship of the *Henry VI* Plays by Word Adjacency," the authors use predictive methods effectively and with care not to confuse prediction with explanation. Their methods involve a complex chain of assumptions, which, although problematic on their own, together produce verifiably reliable results. In conclusion, they write: "No one knows why methods that count frequencies of common words are able to distinguish authorship, and we offer no explanation for why our method of measuring their proximate adjacency is equally successful" (251). See Santiago Segarra, Mark Eisen, Gabriel Egan, and Alejandro Ribeiro, "Attributing the Authorship of the *Henry VI* Plays by Word Adjacency," *Shakespeare Quarterly* 67, no. 2 (2016): 232–56.

13 See Alan Baker, "Quantitative Parsimony and Explanatory Power," *The British Journal for the Philosophy of Science* 54, no. 2 (2003): 245–59; Daniel Nolan, "Quantitative Parsimony," *The British Journal for the Philosophy of Science* 48, no. 3 (1997): 329–43; Elliott Sober, "The Principle of Parsimony," *The British Journal for the Philosophy of Science* 32, no. 2 (1981): 145–56.

14 I borrow and extend the approach to narratology from Lexical Conceptual Structures analysis in cognitive linguistics. See Koichi Takeuchi, "Thesaurus with Predicate-Argument Structure to Provide Base Framework to Determine States, Actions, and Change-of-States," in *Computational and Cognitive Approaches to Narratology*, ed. Takashi Ogata and Taisuke Akimoto (IGI Global, 2016); Ray Jackendoff, *Semantic Structures* (Cambridge, MA: MIT Press, 1992); and Jackendoff, *The Architecture of the Language Faculty* (Cambridge, MA: MIT Press, 1997). Jackendoff's primitive noun categories include thing, event, state, action, place, path, property, and amount and correspond roughly to WordNet lexicographer file names. See his *Architecture of the Language Faculty*, 31 and *Semantic Structures*, 22. Jackendoff writes: "If there is an indefinitely large stock of possible lexical concepts, and the innate basis for acquiring them must be encoded in a finite brain, we are forced to conclude that the innate basis must consist of a set of generative principles—a group of primitives and principles of combination that collectively determine the set of lexical concepts. This implies in turn that most if not all lexical concepts are composite, that is, that they can be decomposed in terms of the primitives and principles of combination of the innate 'grammar of lexical concepts'" (*Semantic Structures*, 10–11).

15 David Elson, Nicholas Dames, and Kathleen McKeown, "Extracting Social Networks from Literary Fiction," *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics* (2010): 138–147.

16 Elson, Dames, and McKeown, "Social Networks," 141, 146.

17 Mikhail Bakhtin, "K Romanu Vospitaniia" [Toward a Theory of Bildungsroman], in *Teoriia Romana* [Theory of the Novel], vol. 3 of *Sobranie Sochinenii* [Collected Works] (Moscow: IMLI-RAN, 2012), 268. All translations are mine, unless otherwise noted.

18 On the distinction between what he calls "story-space" and "discourse-space," see Seymour Chatman, *Story and Discourse: Narrative Structure in Fiction and Film* (Ithaca, NY: Cornell Univ. Press, 2007), 96–101.

19 Fredric Jameson, "The Realist Floor Plan," in *Narrative/Theory*, ed. David H. Richter (White Plains, NY: Longman, 1996), 373.

20 Roland Barthes, "The Reality Effect" in *The Rustle of Language*, trans. Richard Howard (Berkeley: Univ. of California Press, 1989), 141.

21 Franco Moretti, *Atlas of the European Novel, 1800–1900* (London: Verso, 2015), 64; Elson, et al., "Social Networks," 140.

22 Moretti, *Atlas*, 68–69.

23 For traditional narratological approaches to the characterization of space see Chatman, *Story and Discourse*; Ruth Ronen, "Space in Fiction," *Poetics Today* 7, no. 3 (1986): 421–38; Gabriel Zoran, "Towards a Theory of Space in Narrative," *Poetics Today* 5, no. 2 (1984): 309–35.

24 Yuri Lotman, *Struktura Khudozhestvennogo Teksta [The Structure of Literary Text]* (Providence, RI: Brown Univ. Press, 1971), 67. Translation mine. For the full English translation, see Yuri Lotman, *The Structure of the Artistic Text*, trans. Gail Lenhoff and Ronald Vroon, Michigan Slavic Contributions, vol. 7 (Ann Arbor: Univ. of Michigan Press, 1977), 218.

25 Jakob von Uexküll, *A Foray into the Worlds of Animals and Humans*, trans. Joseph D. O'Neil (Minneapolis: Univ. of Minnesota Press, 2010), 53. See also Uexküll and Georg Kriszat, *Streifzüge durch die Umwelten von Tieren und Menschen* (Frankfurt am Main: Fischer-Taschenbuch-Verlag, 1983). Compare with Geertz, 5: "Believing, with Max Weber, that man is an animal suspended in webs of significance he himself has spun."

26 Uexküll, *A Foray into the Worlds*, 51, 53–63. The German *wirken* is used here both in the meaning of "appearance" and "effect."

27 German text from Franz Kafka, *Erzählungen*, ed. Max Brod (Frankfurt: Fischer, 1967). I chose to quote from widely-available online translations by David Wylie after evaluating the alternatives.

28 Jameson, "The Realist Floor Plan," 377: "What is imposed on the reading mind here is a training in uneven surfaces, in the abstract, empty feeling for the inequality of adjacent co-ordinates." Elsewhere on the same page, "the 'stumbling' over this uneven surface subliminally inscribes this empty form on the reading body itself . . . this tripping over the levels within the house is thus what Bakhtin calls a chronotope."

29 Mihaly Csikszentmihalyi and Eugene Rochberg-Halton, *The Meaning of Things: Domestic Symbols and the Self* (New York: Cambridge Univ. Press, 2012), 15.

30 See "Appendix D: Additional tables" in Csikszentmihalyi and Halton, *The Meaning of Things*, 278–289.

31 Jeanne E. Arnold, Anthony P. Graesch, Enzo Ragazzini, and Elinor Ochs, *Life at Home in the Twenty-First Century: 32 Families Open Their Doors* (Los Angeles: Cotsen Institute of Archaeology Press, 2013), 7.

32 Arnold, et al., *Life at Home*, 7. See also Naomi Woods, who writes of "artefact assemblages," "analysed and linked with broader events and processes that were occurring during the second half of the nineteenth-century in North Dunedin." Woods, "Artefacts and Neighbourhood Transformations: A Material Culture Study of Nineteenth-Century North Dunedin," *Australasian Historical Archaeology* 31 (2013): 61.

33 Arnold, et al., *Life at Home*, 9.

34 Arnold, et al., *Life at Home*, 23.

35 Arnold, et al., *Life at Home*, 24–25.

36 Karen Spärck Jones, "A Statistical Interpretation of Term Specificity and Its Application in Retrieval," *Journal of Documentation* 28, no. 1 (1972): 494.

37 Jones, "Term Specificity," 496. Jones also developed the concept of inverse document frequency (IDF).

38 David M. Blei, "Probabilistic Topic Models," *Communications of the ACM* 55, no. 4 (2012): 77–78.

39 Boris Iarkho [Jarcho], "Die Vorläufer des Golias," *Speculum* 3, no. 4 (1928): 554.

40 Iarkho, "Die Vorläufer," 552–554.

41 Iarkho, "Komedii i tragedii Kornel'a: Etud po teorii zhanra," in *Metodologia Tochnogo Literaturovedenia* [The Methodology of Literary Science], ed. M. I. Shapir, vol. 5, *Philologica russica et speculativa* (Moscow: Iaziki Sloviasnkih Kultur, 2006), 498.

42 Iarkho, "Komedii i tragedii Kornel'a," 499: "Based on these tables we can conclude that the author's comedies privilege erotic themes, merriment, and deception, where themes of fear, spiritual suffering, wrath, and valor dominate his tragedies."

43 Iarkho, "Komedii i tragedii Kornel'a," 547–549.
44 Mikhail Gasparov, "The Method of Analysis: 'Again the Clouds above Me,'" in *On Poetry*, vol. 2, *Izbrannye Trudy* [Selected Works] (Moscow: Iaziki Russkoi Kul'tury, 1997), 9. I am suggesting related English equivalents rather than translating directly.
45 Gasparov, "The Method of Analysis," 14.
46 It is productive to consider experiments by Gasparov and Iarkho in relation to "deformative criticism." See Jerome McGann, "Deformance and Interpretation (with Lisa Samuels)," in *Radiant Textuality: Literature After the World Wide Web* (New York: Palgrave Macmillan, 2001), 105–35.
47 Gasparov, "The Method of Analysis," 15.
48 I substitute my own reading of "Anchar" in preference to Gasparov's. See Appendix I for the English translation of the text. Original from Alexander Pushkin, "Anchar," in *Stihotvorenia 1823–1836* [Poems 1823–1836], vol. 2, *Sobranie sochinenii* [Collected Works] (Moscow: Gosudarstvennoe Izdatel'stvo, 1959–62), 229–30.
49 See Marie-Catherine de Marneffe and Christopher D. Manning, "The Stanford Typed Dependencies Representation," in *The Proceedings of COLING 08 Workshop on Cross-Framework and Cross-Domain Parser Evaluation* (Manchester, UK: Association for Computational Linguistics, 2008), 1–8.
50 See René Dirven, "Dividing up Physical and Mental Space into Conceptual Categories by Means of English Prepositions," in *The Semantics of Prepositions From Mental Processing to Natural Language Processing*, ed. Cornelia Zelinsky-Wibbelt (Boston: De Gruyter, 1993); Leonard Talmy, "How Language Structures Space," in *Spatial Orientation: Theory, Research, and Application*, ed. Herbert L. Pick and Linda P. Acredolo (Boston: Springer US, 1983), 225–82; and Joost Zwarts and Yoad Winter, "Vector Space Semantics: A Model-Theoretic Analysis of Locative Prepositions," *Journal of Logic, Language and Information* 9, no. 2 (2000): 169–211.
51 The discussion is informed by Talmy, "How Language Structures Space," 230.
52 Emily Brontë and Anne Brontë, *Wuthering Heights; and, Agnes Grey* (London: Smith, Elder, and Co., 1889), 33.
53 See Massimiliano Ciaramita and Mark Johnson, "Supersense Tagging of Unknown Nouns in WordNet," in *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing* (Stroudsburg, PA: Association for Computational Linguistics, 2003), 168–175; and de Marneffe and Manning, "Stanford Type Dependencies."
54 I am using Chatman's distinctions between "story" and "discourse." To be more precise, we are dealing with three categories of space and not just two: one in purely physical, textual terms, as character length or a number of words; the second in semantic units, in terms of the fictional, diegetic world; and the third in terms of the mixture between physical and diegetic space. It is that third type of "size" or "volume" units that I would properly designate as "narratological."
55 Text processing, code, and data visualization by the author, unless otherwise noted.
56 The distribution of density and clutter fulfills E. S. Pearson's assumptions, passing tests for normal distribution and homoscedasticity, justifying the regression approach selected here. See Pearson, R. B. D'Agostino, and K. O. Bowman, "Tests for Departure from Normality: Comparison of Powers," *Biometrika* 64, no. 2 (1977): 231–46; and Ralph B. D'Agostino, "An Omnibus Test of Normality for Moderate and Large Size Samples," *Biometrika* 58, no. 2 (1971): 341–48.
57 Since our data contains three categories (rural, urban, and mixed), I calculated point biserial correlations using a "dummy" variable to express a ternary category in binary terms. I then took the alternative approach of adducing the "mixed" flag to the "urban" column, effectively conflating the two. The second approach did not alter the results in any significant way. I therefore report the simpler, binary statistics. See Peter Y. Chen and

Paula M. Popovich, "Special Cases of Pearson's *r*: Point-Biserial Correlation, r_{pb}," in *Correlation: Parametric and Nonparametric Measures* (Thousand Oaks, CA: Sage Publications, 2002), 26–29; and S. Das Gupta, "Point Biserial Correlation Coefficient and Its Generalization," *Psychometrika* 25, no. 4 (1960): 393–408.

58 Thomas Richards, *The Commodity Culture of Victorian England: Advertising and Spectacle, 1851–1914* (Stanford, CA: Stanford Univ. Press, 1991), 2.

59 See Stephen Best and Sharon Marcus, "Surface Reading: An Introduction," *Representations* 108, no. 1 (2009): 1–21; Heather Love, "Close but Not Deep: Literary Ethics and the Descriptive Turn," *New Literary History* 41, no. 2 (2010): 371–91.

60 Elaine Freedgood, *The Ideas in Things: Fugitive Meaning in the Victorian Novel* (Chicago: Univ. of Chicago Press, 2010), 1, 105.

61 Arthur Conan Doyle, *A Study in Scarlet* (New York: Penguin, 2001), 15.

62 This line of reasoning was developed subsequent to helpful discussions at the Analysing Text conference hosted by the Alan Turing Institute at the British Library (10/13/2017) and at Yale's Theory and Media Studies Colloquium (9/22/2017). I extend my gratitude to Simon DeDeo, Marta Figlerowicz, and R. John Williams, among other participants.

63 Jane Austen, *Pride and Prejudice* (London: Macmillan, 1906), 7.

64 Austen, *Pride and Prejudice*, 51.

65 Austen, *Pride and Prejudice*, 84.

66 Austen, *Pride and Prejudice*, 99.

67 See William Cleveland, "LOWESS: A Program for Smoothing Scatterplots by Robust Locally Weighted Regression," *The American Statistician* 35, no. 1 (1981): 54; William Cleveland and Susan J. Devlin, "Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting," *Journal of the American Statistical Association* 83, no. 403 (1988): 596–610.

68 See Moretti, "Serious Century," in *The Novel: History, Geography, and Culture* (Princeton NJ: Princeton Univ. Press, 2007), 364–400; Liesl Olson, *Modernism and the Ordinary* (New York: Oxford Univ. Press, 2014), 18; and Richards, *Commodity Culture*, 1. See also Cynthia Sundberg Wall, *The Prose of Things: Transformations of Description in the Eighteenth Century* (Chicago: Univ. of Chicago Press, 2006), 10: "Furniture and fabric and object details of particularized rooms . . . dominate ninetieth-century novels and poetry"; Freedgood, *Ideas in Things*, 4: "There is more description of things as the nineteenth century gets going." See Andres Miller, *Novels Behind Glass: Commodity Culture and Victorian Narrative* (Cambridge, MA: Cambridge Univ. Press, 2008), 6: "Among the dominant concerns motivating mid-Victorian novelists was a penetrating anxiety . . . that their social and moral world was being reduced to a warehouse of goods and commodities."