Narrative and Database: Natural Symbionts

N. KATHERINE HAYLES

AH, THE POWER OF METAPHORS—ESPECIALLY those that propagate with viral intensity through a discursive realm. At issue here is Lev Manovich's characterization of narrative and database in *The Language of New Media* as "natural enemies" (228), a phrase Ed Folsom rehearses in his generous and enlightening discussion of *The Walt Whitman Archive*. The metaphor resonates throughout Folsom's essay in phrases such as "the attack of database on narrative," culminating in his figure of database's spread as a viral pandemic that "threatens to displace narrative, to infect and deconstruct narrative endlessly, to make it retreat behind the database or dissolve back into it." In this imagined combat between narrative and database, database plays the role of the Ebola virus whose voracious spread narrative is helpless to resist. The inevitable triumph of database over narrative had already been forecast in Manovich's observation that "databases occupy a significant, if not the largest, territory of the new media landscape." Indeed, so powerful and pervasive are databases for Manovich that he finds it "surprising" narratives continue to exist at all in new media (228). In Manovich's view, the most likely explanation of narrative's persistence is the tendency in new media to want to tell a story, a regression he identifies with cinema. Even this, he suggests, is being eradicated by experimental filmmakers such as Peter Greenaway (237–39).

Rather than natural enemies, narrative and database are more appropriately seen as natural symbionts. Symbionts are organisms of different species that have a mutually beneficial relation. For example, a bird picks off bugs that torment a water buffalo, making the beast's existence more comfortable; the water buffalo provides the bird with tasty meals. Because database can construct relational juxtapositions but is helpless to interpret or explain them, it needs narrative to make its results meaningful. Narrative, for its part, needs database in the computationally intensive culture of the new millennium to enhance its cultural authority and test the generality of its insights. If narrative often dissolves into database, as Folsom suggests, database catalyzes and indeed demands narrative's reappearance as soon as meaning and interpretation are required. The dance (or, as I prefer to call it, the complex ecology) of narrative and database originates in their different ontologies, purposes, and histories. To understand more precisely the interactions between these two cultural forms, let us consider these characteristics.

As Manovich observes, database parses the world from the viewpoint of large-scale data collection and management. For the late twentieth and early twenty-first centuries, this means seeing the world in terms that the computer can understand. By far the most pervasive form of database is the relational, which has almost entirely replaced the older hierarchical, tree, and network models and continues to hold sway over the newer object-oriented models. In a relational database, the

N. KATHERINE HAYLES, Hillis Professor of Literature and Distinguished Professor in the departments of English and design / media arts at the University of California, Los Angeles, teaches and writes on the relations among science, technology, and literature in the late twentieth and twenty-first centuries. Her books include *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics* (U of Chicago P, 1999), which won the René Wellek Prize for the best book in literary theory for 1998–99, and *Writing Machines* (MIT P, 2002), which won the 2003 Susanne K. Langer Award for Outstanding Scholarship in the Ecology of Symbolic Form. Her most recent book is *My Mother Was a Computer: Digital Subjects and Literary Texts* (U of Chicago P, 2005), and her new project is *Electronic Literature: New Horizons for the Literary* (U of Notre Dame P, forthcoming Feb. 2008).

https://doi.org/10.1632/S0030812900168580 Published online by Cambridge University Press

data are parsed into tables consisting of rows and columns, where the column heading, or attribute, indicates some aspect of the table's topic. Ideally, each table contains data pertaining to only one "theme" or central data concept. One table, for example, might contain data about authors, where the attributes might be last name, first name, birth date, death date, book titles, and so on; another might have publishers' data, also parsed according to attributes; another, books. Relations are constructed among data elements in the tables according to set-theoretic operations, such as "insert," "delete," "select," and especially "join," the command that allows data from different tables to be combined. Common elements allow correlations between tables to be made; for example, Whitman would appear in the authors table as an author and in the books table correlated with the titles he published; the publishers table would correlate with the books table through common elements and through these elements back to the authors table. Working through these kinds of correlations, set-theoretic operations also allow new tables to be constructed from existing ones. Different interfaces can be designed according to the particular needs of users. Behind the interface, whatever its form, is a database-management system that employs set-theoretic notation to query the database and manipulate the response through SQL and related languages (*SQL* is commonly expanded as *Structured Query Language* and pronounced "sequel").

The great strength of database, of course, is the ability to order vast data arrays and make them available for different kinds of queries. Two fundamental aspects typically characterize relational databases. One, indicated above, is their construction of relations between attributes and tables. The other is a well-constructed database's self-containment or, as the technical literature calls it, self-description. A database is said to be self-describing because its user does not need to go outside the database to see what it contains. As David Kroenke and David Auer put it in *Database Concepts*, the "structure of the database is contained within the database itself," so that the database's contents can be determined just by looking inside it (13). Its self-describing nature is apparent in SQL commands. For the database mentioned above containing information about authors, books, and publishers, for example, a typical SQL command might take the generalized form "SELECT AUTHOR .AuthorName, BOOK.BookTitle, BOOK .BookDate, BOOK.Publisher, PUBLISHER .Location," where the table names are capitalized in full (as are SQL commands) and the data elements are categorized according to the attributes, with a period separating table name from attribute. The database's self-description is crucial to being able to query it with set-theoretic operations, which require a formally closed logical system on which to operate. This is also why databases fit so well in computers; like databases, computers employ formal logic as defined by the logic gates that underlie all executable commands.

The self-describing nature of database provides a strong contrast with narrative, which always contains more than indicated by a table of contents or a list of chapter contents. Databases can, of course, also extend outward when they are linked and queried as a network—for example, in data-mining and text-mining techniques—but they do not lose the formal properties of closure that make them self-describing artifacts. Nevertheless, the technologies of linking databases have proved to be remarkably powerful, and the relations revealed by set-theoretic operations on networks of linked databases can have stunning implications. For example, data- and text-mining techniques allowed the epidemiology researchers Don Swanson and N. R. Smalheiser to hypothesize causes for rare diseases that hitherto had resisted analysis because they occurred infrequently at widely separated locales.[1] Even in this case, however, the mean-

https://doi.org/10.1632/S0030812900168580 Published online by Cambridge University Press

ing of the relations posited by the database remains outside the realm of data techniques. What it means that Whitman, say, used a certain word 298 times in *Leaves of Grass* while using another word only three times requires interpretation—and interpretation, almost inevitably, invokes narrative to achieve dramatic impact and significance. Many data analysts and statisticians are keenly aware of this symbiosis between narrative and data. John W. Tukey, in his classic textbook *Exploratory Data Analysis*, for example, explains that the data analyst "has to learn . . . how to expose himself to what his data are willing—or even anxious—to tell him," following up the lesson by later asking the student what story each dataset tells (21, 101).

Database and narrative, their interdependence notwithstanding, remain different species, like bird and water buffalo. Databases must parse information according to the logical categories that order and list the different data elements. Indeterminate data—data that are not known or that elude the boundaries of the preestablished categories—must either be represented through a null value or not be represented at all. Even though some relational databases allow for the entry of null values, such values work in set-theoretic operations as a contaminant, since any operation containing a null value will give the same as its result, as multiplying any number by zero yields zero. Null values can thus quickly spread through a database, rendering everything they touch indeterminate. Moreover, database operations say nothing about how data are to be collected or which data should qualify for collection, nor do they indicate how the data should be parsed and categorized. Such decisions greatly influence the viability, usefulness, and operational integrity of databases. Thomas Connolly and Carolyn Begg in *Database Systems* estimate that for corporate database software development projects, eighty to ninety percent do not meet their performance goals, eighty percent are delivered late and over budget, and forty percent fail or are abandoned (270). Anticipating such problems, database textbooks routinely advise students to obscure suboptimal performance by keeping the database design confidential and confining discussions with the paying client to what the interface should look like and how it should work.

The indeterminacy that databases find difficult to tolerate marks another way in which narrative differs from database. Narratives gesture toward the inexplicable, the unspeakable, the ineffable, whereas databases rely on enumeration, requiring explicit articulation of attributes and data values.[2] While the concatenation of relations might be suggestive, as Folsom remarks in discussing the new kinds of knowledge that the Whitman databases can generate, databases in themselves can only speak that which can explicitly be spoken. Narratives, by contrast, invite in the unknown, taking us to the brink signified by Henry James's figure in the carpet, Kurtz's "The horror, the horror," Gatsby's green light at pier's end, Kerouac's beatitude, Pynchon's crying of lot 49. Alan Liu, discussing the possibilities for this kind of gesture in a post-industrial, information-intensive era, connects it with "the ethos of the unknown" and finds it expressed in selected artworks as a "data pour," an overflowing, uncontainable excess that he links with transcendence (esp. 81).

Whereas database reflects the computer's ontology and operates with optimum efficiency in set-theoretic operations based on formal logic, narrative is an ancient linguistic technology almost as old as the human species. As such, narrative modes are deeply influenced by the evolutionary needs of human beings negotiating unpredictable three-dimensional environments populated by diverse autonomous agents. As Mark Turner has argued in *The Literary Mind: The Origins of Thought and Language*, stories are central in the development of human cognition. Whereas database allows large amounts of information to be sorted, cataloged, and queried, narrative models how

https://doi.org/10.1632/S0030812900168580 Published online by Cambridge University Press

minds think and how the world works, projects in which temporality and inference play rich and complex roles. Extending Paul Ricoeur's work on temporality and Gérard Genette's on narrative modalities, Mieke Bal analyzes narrative as requiring, at a minimum, an actor and narrator and consisting of three distinct levels, text, story, and fabula, each with its own chronology (6). To this we can add Brian Richardson's emphasis in *Unlikely Stories: Causality and the Nature of Modern Narrative* on causality and inference in narrative.[3]

Why should narrative emphasize these aspects rather than others? Bound to the linear sequentiality of language, narrative complicates it through temporal enfoldings of story (or, as Genette prefers to call it, discourse) and fabula, reflecting the complexities of acting when knowledge is incomplete and the true situation may be revealed in an order different from the one logical reconstruction requires. Narrator and actor inscribe the situation of a subject constantly negotiating with agents who have their own agendas and desires, while causality and inference represent the reasoning required to suture different temporal trajectories, motives, and actions into an explanatory frame. These structures imply that the primary purpose of narrative is to search for meaning, making narrative an essential technology for human beings, who can arguably be defined as meaning-seeking animals.

Bound to the linear order of language through syntax, narrative is a temporal technology, as the complex syncopations between story and fabula demonstrate. The order in which events are narrated is crucial, and temporal considerations are central to narratology, as Ricoeur's work, among others', illustrates. Datasets and databases, by contrast, lend themselves readily to spatial displays, from the two-dimensional tables typical of relational databases to the more complex n-dimensional arrays and spatial forms that statisticians and data analysts use to understand the stories that data tell.

Manovich touches on this contrast when he perceptively observes that for narrative, the syntagmatic order of linear unfolding is actually present on the page, while the paradigmatic possibilities of alternative word choices are only virtually present. For databases, the reverse is true: the paradigmatic possibilities are actually present in the columns and the rows, while the syntagmatic progress of choices concatenated into linear sequences by SQL commands is only virtually present. I would add to this observation that time and space, the qualities Kant identified as intrinsic to human sensory-cognitive faculties, inevitably coexist. While one may momentarily be dominant in a given situation, the other is always implicit, a natural symbiont whose existence is inextricably entwined with that of its partner. It should be no surprise, then, that narrative and database align themselves with these partners or that they too exist in symbiosis with each other.

Given this entwinement, is it plausible to imagine, as Manovich and Folsom imply at various points, that database will replace narrative to the extent that narrative fades from the scene? A wealth of evidence points in the other direction: narrative is essential to the human lifeworld. Jerome Bruner, in his book significantly entitled *Acts of Meaning*, cites studies indicating that mothers tell their children some form of narrative several times each hour to guide their actions and explain how the world works (81–84). We take narrative in with mother's milk and practice it many times every day of our lives—and not only in high-culture forms such as print novels. Newspapers, gossip, math story problems, television dramas, radio talk shows, and a host of other communications are permeated by narrative. Wherever one looks, narratives surface, as ubiquitous in everyday culture as dust mites.

What has changed in the informative-intensive milieu of the twenty-first century is the position narrative occupies in the culture. Whereas in the classical Greek and Roman

https://doi.org/10.1632/S0030812900168580 Published online by Cambridge University Press

era narrative was accepted as an adequate explanation for large-scale events—the creation of the world, the dynamics of wind and fire, of earth and water—global explanations are now typically rooted in data analysis. If we want to understand the effects of global warming or whether the economy is headed for a recession, we likely would not be content with anecdotes about buttercups appearing earlier than usual in the backyard or Aunt Agnes's son not finding a job. Data, the databases that collect, parse, and store them, and the database-management systems that concatenate and query them are essential for understanding large-scale phenomena. At the global level, databases are essential. However, narrative enters even in the interpretation of the relations revealed by database queries. When Alan Greenspan testified before Congress, he typically did not recount data alone. Rather, he told a story, and it was the story, not the data by themselves, that propagated through the news media because it encapsulated in easily comprehensible form the meaning exposed by data collection and analysis.

In contrast to global dynamics, narrative at the local level remains pervasive, albeit increasingly infused by data. As Folsom indicates, in the face of the overwhelming quantities of data that database-management systems now put at our fingertips, no one narrative is likely to dominate as *the* explanation, for the interpretive possibilities proliferate exponentially as databases increase. In this respect, the advent of the Internet, especially the World Wide Web, has been decisive. Never before in the history of the human species has so much information been so easily available to so many. The constant expansion of new data accounts for an important advantage that relational databases have over narratives, for new data elements can be added to existing databases without disrupting their order. Unlike older computer database models in which memory pointers were attached directly to data elements, relational databases allow the order of the rows and columns to vary without affecting the system's ability to locate the proper elements in memory. This flexibility allows databases to expand without limitation (subject, of course, to the amount of memory storage allocated to the database). Narrative in this respect operates quite differently. Sensitively dependent on the order in which information is revealed, narrative cannot in general accommodate the addition of new elements without, in effect, telling a different story. Databases tend toward inclusivity, narratives toward selectivity. Harry Mathews explores this property of narrative in *The Journalist: A Novel*, where the unnamed protagonist, intent on making a list of everything that happens in his life, thinks of more and more items, with the predictable result that the list quickly tends toward chaos as the interpolations proliferate. The story of this character's life cannot stabilize, because the information that constitutes it continues to grow exponentially, until both list and subject collapse.

That novels like *The Journalist* should be written in the late twentieth century speaks to the challenges that database poses to narrative in the age of information. No doubt phenomena like this explain why Manovich would characterize database and narrative as "natural enemies" and why thoughtful scholars like Folsom would propagate the metaphor. Nevertheless, the same dynamic also explains why the expansion of database is a powerful force constantly spawning new narratives. The flip side of narrative's inability to tell *the* story is the proliferation of narratives as they transform to accommodate new data and mutate to probe what lies beyond the expanding infosphere. No longer singular, narratives remain the necessary others to database's ontology, the perspectives that invest the formal logic of database operations with human meanings and that gesture toward the unknown hovering beyond the brink of what can be classified and enumerated.

https://doi.org/10.1632/S0030812900168580 Published online by Cambridge University Press

Notes

1. See, for example, Swanson and Smalheiser, "Interactive System" and "Assessing."

2. The exception is the null value, which has its own problems, as discussed above.

3. Discussing narrative, Bruner also emphasizes the importance of causality, identifying crucial components as agency, sequential order, sensitivity to the canonical (or context), and narrative perspective (77).

Works Cited

Bal, Mieke. *Narratology: Introduction to the Theory of Narrative*. 2nd ed. Toronto: U of Toronto P, 1998.

Bruner, Jerome. *Acts of Meaning: Four Lectures on Mind and Culture*. Cambridge: Harvard UP, 2002.

Connolly, Thomas, and Carolyn Begg. *Database Systems*. New York: Harlow; Essex: Pearson Educ., 2002.

Genette, Gérard. *Narrative Discourse: An Essay in Method*. Ithaca: Cornell UP, 1983.

Kroenke, David M., and David J. Auer. *Database Concepts*. 3rd ed. New York: Prentice, 2007.

Liu, Alan. "Transcendental Data: Toward a Cultural History and Aesthetics of the New Encoded Discourse." *Critical Inquiry* 31 (2004): 49–84.

Manovich, Lev. *The Language of New Media*. Cambridge: MIT P, 2001.

Mathews, Harry. *The Journalist: A Novel*. Boston: Godine, 1994.

Richardson, Brian. *Unlikely Stories: Causality and the Nature of Modern Narrative*. Newark: U of Delaware P, 1997.

Ricoeur, Paul. *Time and Narrative*. Vol. 1. Chicago: U of Chicago P, 1990.

Swanson, Don R., and N. R. Smalheiser. "Assessing a Gap in the Biomedical Literature: Magnesium Deficiency and Neurologic Disease." *Neuroscience Research Communications* 14 (1994): 1–9.

———. "An Interactive System for Finding Complementary Literatures: A Stimulus to Scientific Discovery." *Artificial Intelligence* 91 (1997): 183–203.

Tukey, John W. *Exploratory Data Analysis*. Reading: Addison, 1977.

Turner, Mark. *The Literary Mind: The Origins of Thought and Language*. New York: Oxford UP, 1998.

Reply

ED FOLSOM

Ah, the power of metaphors indeed! To describe the relation between narrative and database, N. Katherine Hayles offers an astute alternative to Lev Manovich's "natural enemies" metaphor: she suggests "natural symbionts," a metaphor I plan to appropriate and use from now on. Her claim that "database catalyzes and indeed demands narrative's reappearance as soon as meaning and interpretation are required" incisively articulates what she calls the "dance" of narrative and database. I've thought of the relation as an endless battle (once narrative begins to win, database rallies, and vice versa), but Hayles's metaphor more efficaciously captures what she rightly characterizes as "the complex ecology" of these two modes of organizing and accessing the represented world.

And, as Hayles makes clear, the metaphors are essential. The term *database* itself is a metaphor, a base onto which we put things that are given (data). The word is less than fifty years old and has mutated in meaning over the decades. Few of us (certainly not I) can approach a database without an array of metaphoric terms that make it seem something it is not. Years ago, when I used to hit a key on my old typewriter, I could follow and even explain the mechanical process that struck an inked ribbon with a typebar to impress a letter on a page. Now, when I hit a key on my computer keyboard, my knowledge of the process that makes a letter appear on my screen is hazy, to say the least, not to mention the process that transfers it to paper. How this sentence I'm writing gets preserved on my USB stick and in what form is a mystery to me. Without the metaphoric apparatus that

https://doi.org/10.1632/S0030812900168580 Published online by Cambridge University Press