CAMBRIDGE
UNIVERSITY PRESS

REVIEW ESSAY

Digital Humanities and the Discontents of Meaning

Michael A. Fuller* (ID)

University of California, Irvine
*Corresponding author. E-mail:mafuller@uci.edu

The digital humanities offer more than just a set of tools. The application of software to assist in the analysis of large collections of data does not just expand the volume of material we can incorporate in our work, it also expands how we in the humanities understand the nature of meaning. The recent scholarly turn to the expanded modes of analysis made possible by DH is not just the "latest new thing" but gives the scholarly community a way to articulate and respond to long-standing doubts about the epistemological grounding in the practice of the humanities. Even more importantly, I believe that this broadening of inquiry afforded by DH is intrinsic to the humanistic project itself. In this essay, I seek in particular to connect the implicit conceptual substructure behind the architectural logic of the digital humanities to key strains of hermeneutic thought that have established a basis for exploring the question of how we are to understand the vast, variegated world of historical human experience that is the object of our humanistic inquiries across disciplines.

The Retreat of Meaning

How do the digital humanities provide an epistemological model for thinking about the human? I would argue that the digital humanities live within the impasse of the failure of language that has long shadowed humanistic study, but offer compelling ways of thinking about meaning within that impasse. This reorientation becomes clearer when we consider the digital humanities within the context of the epistemological predicament of the humanities in the past half century.

The failure of language—the simple truism that words are just words, signifiers not securely attached to things—has long presented an epistemological problem in the humanistic interpretation of texts. In the past fifty years in particular, scholars have intensely debated the question that if all we have are words—and if our mode of thinking about these words is through systems of yet more words—how can we reach beyond words to the things themselves? Of course not all our understanding of the past relies on words: we have the substantial resources of artifacts from fragments of textile to pottery and metal castings, to human bones, to extensive archeological sites. Still, many aspects of the past—and in particular, the intellectual, affective, and aesthetic dimensions—remain inaccessible without relying on textual corpora. Without a theory of reference to connect

CrossMark

texts firmly to the world, the meanings we can justifiably draw from textual evidence become suspect. The story of the unmooring of language is well-known: structuralism gave us a synchronic account of meaning in language. Its logic of mutual differentiation among signifiers offered important insights into the working of texts, but it also led to the vision of language as a closed, endless mirroring of words. Substantive links between signifiers and a signified world outside of language dissolved into an infinite deferral of meaning.

Although the problem of the retreat of reference has deep roots in the Western philosophical tradition, with the vanishing of reference to ground meaning, we have been forced to confront the question: if the world does not shape the structure of language, then what does? One answer derives from the isolated subjectivity of the individual reader. Words and texts mean to me what they mean to me, without access to larger structures other than those I supply. This is how we read unreflectingly most of the time, and its connoisseurship and inner world of responses can be very satisfying. This certainly is how most of my students read, and they are happy within an asserted subjectivity of meaning which admits no further analysis. However, a second set of responses has evolved. What is often collectively deemed the "hermeneutics of suspicion" unravels this isolated subjectivity of meaning and posits the individual as a node within systems of difference shaped by the structuring of power. Compelling critiques of institutional racism and sexism and the ways in which they are embedded in and structure our shared discourses offer ample evidence to support a vision of the ideological structuring of meaning. Critical theory both in literary studies and in the social sciences offers modes of revealing the duplicity and evasions of texts and showing that texts that claim to assert truths can be revealed as relying on concealed predications. While such analyses remain a crucial check on complacency in scholarly as well as broader public discourse, by their nature their reflections remain within and demonstrate in myriad ways the failures of language rather than offering deeper knowledge of the human. Aware of the limits of critique, scholars in critical theory increasingly have come to echo Bruno Latour's query in "Why Has Critique Run Out of Steam? From Matters of Fact to Matters of Concern."[1] Indeed, the movement toward "postcritique" discussed by Rita Felski and others is an important index of the discontents of meaning and of a desire in humanistic disciplines to develop perspectives and methods that move beyond the impasses of failures of language.[2]

The Challenge of Scientific Insights into the Human

While scholars in the humanities have been sharpening their modes of critical analysis, a second, less visible trend outside of the humanities has been rapidly gaining strength in the pragmatic study of language, perception, and cognition that treats the structuring of human experience as an object of scientific inquiry. The success of neural networks and ever more pervasive forms of artificial intelligence recently has caught the public imagination and seemingly rendered scholarship in the humanities yet more irrelevant to the exploration of human meaning. Thus, humanistic scholarship—either in the traditionalist mode of individual sensibility or in the contemporary mode of social critique —has little standing to speak to the larger patterns and deeper meanings of human

---

[1]Bruno Latour, "Why Has Critique Run Out of Steam? From Matters of Fact to Matters of Concern," *Critical Inquiry* 30.2 (2005), 225–48.

[2]For example, see Rita Felski, *The Limits of Critique* (Chicago: University of Chicago Press, 2015) and Elizabeth S. Anker and Rita Felski, eds., *Critique and Postcritique* (Durham: Duke University Press, 2017).

experience. Other disciplines—now largely in the sciences—are stepping in to provide insights into the human. Geoffrey Harpham, the former director of the National Humanities Center, lamented in 2006 : "One of the most striking features of contemporary intellectual life is the fact that questions formerly reserved for the humanities are today being approached by scientists in various disciplines such as cognitive science, cognitive neuroscience, robotics, artificial life, behavioral genetics and evolutionary biology."[3] This story of the crisis in the humanities has been presented—and lamented— many times and in innumerable variations in recent years and is in no way new. However, as I suggested earlier, it is a story that may have a happy—even if unlooked for—ending in which DH plays a part. Harpham continues in his essay, "Science and the Theft of Humanity," which I quote at length:

> Humanists, who have been only partially aware of the work being done by scientists and other nonhumanists on their own most fundamental concepts, must try to overcome their disciplinary and temperamental resistances and welcome these developments as offering a new grounding for their own work. They must commit themselves to be not just spectators marveling at new miracles, but coinvestigators of these miracles, synthesizing, weighing, judging and translating into the vernacular so that new ideas can enter public discourse.
>
> They—we—must understand that while scientists are indeed poaching our concepts, poaching in general is one of the ways in which disciplines are reinvigorated, and this particular act of thievery is nothing less than the primary driver of the transformation of knowledge today. For their part, those investigating the human condition from a nonhumanistic perspective must accept the contributions of humanists, who have a deep and abiding stake in all knowledge related to the question of the human.
>
> We stand today at a critical juncture not just in the history of disciplines but of human self-understanding, one that presents remarkable and unprecedented opportunities for thinkers of all descriptions. A rich, deep and extended conversation between humanists and scientists on the question of the human could have implications well beyond the academy. It could result in the rejuvenation of many disciplines, and even in a reconfiguration of disciplines themselves—in short, a new golden age.

I share both Harpham's optimism and his call for us humanists to look to the sciences to provide at least partial grounding for our work.

 In particular, engagement with the neuroscience of memory, emotion, language, and selfhood can deepen humanistic reflection on the patterns of human experience. However, there is a yet broader and more profound conceptual shift at work, of which whatever neuroscience can tell us is just a part. Words and texts are traces of human action and accordingly participate in the broader patterns of life as humans live it. The sciences can help us with crucially important creaturely dimensions of experience—help us understand the biological mechanisms of memory, affect, and language production—but situating and understanding texts within the human world built upon

---

[3]Geoffrey Harpham, "Science and the Theft of Humanity," *American Scientist* No. 4 (July–August 2006), 296–98, cited in Geoffrey Rockwell and Stéfan Sinclair, *Hermeneutica: Computer-Assisted Interpretation in the Humanities* (Cambridge, MA: MIT Press, 2016), 20.

these basic processes return us to the humanistic discipline of hermeneutics, of which, I argue below, the digital humanities are a modern technological incarnation. The scholars who shaped modern Western hermeneutics, born in the aftermath of Kant's Copernican revolution, confronted the problem of understanding the religious, philosophical, and literary legacy of the past without access to timeless essences. Instead, they had to develop theories and methods to allow them to extract understanding from the totality of the evidence at hand. We confront the same problem—with similar hopes—in the practice of the digital humanities. Thus I turn next to describe the antifoundational approach to language in DH, connect it to Wittgenstein's proposal of linguistic meaning defined through usage, and then trace Wittgenstein's model for understanding back to the hermeneutic lineage through Wilhelm Dilthey to Friedrich Schleiermacher. Having considered theoretical models for the broad integration of data from lived experience in the hermeneutic tradition, I return to propose that these integrative models implicitly shape the emerging approaches to the digital humanities and in fact complement the new approaches to thinking about human perception, memory, emotion, selfhood and meaning being developed through research in neuroscience and evolutionary biology.

The Digital Humanities

The digital humanities play a critical role in the gradual opening up of the humanities to the broader interpretation of the human because they embody and articulate a different understanding of the nature of meaning. I begin by returning to the issue of textual meaning, in particular to the central problem of how words mean: even if we stay for the moment in Saussure's structuring of signifiers through mutual differentiation, the paradigms of the digital humanities do not find language either condemned to the infinite regress of critical theory or an order built upon ideology.

Topic modeling, one of the most familiar techniques in the digital humanities, provides a clear example of the modeling of meaning in DH. Explaining the concept presents a challenge because the very phrase "topic modeling" all too easily misleads those who are without the technical knowledge of what these "topics" are and the mathematics by which they are derived. Without that frame, people seem to assume that the "words" in topic-modeling systems rely in some way on the semantic structure of language. Instead, in topic modeling, words as signifiers are not only cut off from any possible signified content but also from the entire system of mutual differentiation that defines the signifiers of a language. I believe that a brief introduction to the basic elements of linear algebra upon which topic modeling is built will go a long way to help clarify the logic of meaning as defined within the set of conceptual structures associated with topic modeling. In topic modeling, one begins with a collection of texts. In the most common approach, the order of the words does not matter, and each document is considered simply an unordered "bag of words." Moreover, the words in the texts are meaningless tokens, just strings of bytes. The goal of topic modeling then is to build a system of mutual differentiation relying only on the collection of the bag of words in the corpus to be analyzed.

This mathematized version of meaning and structure is unfamiliar to most humanists, but grasping it is a key to seeing how the digital humanities synthesize the vast corpora of data into a new world of empirically organized human connections. My aim at this juncture is to explain the basic mathematics in topic modeling to demystify the structure of meaning defined by topic modeling and related paradigms.

Texts as Matrices of Meaning

The first version of a structured representation of words within the particular domain of selected texts is simply a large matrix with the dimensions determined by the number of documents and the total number of different words in that collection of documents.[4] The value for each element in the matrix is the frequency (f) of a given word (w_i) in a given document (d_j). Thus with D documents and N different words, we have a matrix:

$$DocumentWords = \begin{bmatrix} f(d_1 w_1) & \cdots & f(d_i w_1) & \cdots & f(d_D w_1) \\ f(d_1 w_j) & \cdots & f(d_i w_j) & \cdots & f(d_D w_j) \\ f(d_1 w_N) & \cdots & f(d_i w_N) & \cdots & f(d_D w_N) \end{bmatrix}$$

$$document_i = (f(d_i w_1), \ldots f(d_i w_j), \ldots f(d_i w_N)) \text{ [column vector]}$$

$$word_j = (f(d_i w_j), \ldots f(d_i w_j), \ldots f(d_D w_j)) \text{ [row vector]}$$

Any particular document is defined as an array of words, and any word is defined as its frequency in the array of documents. If we had a thousand documents with ten thousand different words, we would have a matrix with 10 million entries (1,000 documents x 10,000 words). In topic modeling, one looks for a way to change this very large matrix into the product of two smaller matrices such that there are K number of topics (t_k) where each document can now be described as an array of topics, and each topic can be described as an array of words:

$$DocumentTopic = \begin{bmatrix} f(d_1 t_1) & \cdots & f(d_i t_1) & \cdots & f(d_D t_1) \\ f(d_1 t_j) & \cdots & f(d_i t_j) & \cdots & f(d_D t_j) \\ f(d_1 t_K) & \cdots & f(d_i t_K) & \cdots & f(d_D t_K) \end{bmatrix}$$

---

[4]A matrix is a mathematical formalism from the discipline of linear algebra. An m x n matrix is a set of numbers arranged in m rows, with each row having a set of n numbers. The n values all line up to produce n columns, each with m values. Thus the following is a 2 x 3 matrix:

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

It has 2 rows, (1,2,3) and (4,5,6), and three columns, (1,4), (2,5), and (3,6). The sets of numbers for the rows and columns in turn define vectors, which define positions in multidimensional spaces. The row-vectors in the example are 3-dimensional, while the column-vectors are 2-dimensional. An m x n matrix defines a specific way of mapping n-dimensional vectors into m-dimensional space (i.e., turn n-dimensional vectors into m-dimensional vectors).

$$document_i = (f(d_i t_1), \dots f(d_i t_j), \dots f(d_i t_K)) \text{ [column vector]}$$

$$topic(document)_j = (f(d_1 t_j), \dots f(d_i t_j), \dots f(d_D t_j)) \text{ [row vector]}$$

$$TopicWord = \begin{bmatrix} f(t_1 w_1) & \cdots & f(t_i w_1) & \cdots & f(t_K w_1) \\ \vdots & & \vdots & & \vdots \\ f(t_1 w_j) & \cdots & f(t_i w_j) & \cdots & f(t_K w_j) \\ \vdots & & \vdots & & \vdots \\ f(t_1 w_N) & \cdots & f(t_i w_N) & \cdots & f(t_K w_N) \end{bmatrix}$$

$$topic(words)_i = (f(t_i w_1), \dots f(t_i w_j), \dots f(t_i w_N)) \text{ [column vector]}$$

$$word_j = (f(t_i w_j), \dots f(t_i w_j), \dots f(t_K w_j)) \text{ [row vector]}$$

where

$$DocumentWords \approx DocumentTopics \times TopicsWords$$

The advantage of this factoring of the matrix into two separate matrices is that if one has 100 topics, then the *DocumentTopics* matrix has 100,000 elements (100 topics x 1,000 documents) and the *TopicsWords* matrix has 1,000,000 entries (100 topics x 10,000 words), for a grand total of 1,100,000 entries. Finding a matrix of 100 topics allows us to condense the initial data by a factor of about 10.

So far this looks like a mathematical trick, but what does a "topic" then mean, given the math that defines it? The words in the set of documents are not randomly distributed. They have an internal logic within the collection of documents, and "topics" capture the regularities in the appearance of the words. Words cluster together, and the topics represent those clusters. People working with topic modeling stress that the mathematical tools that find the two topic matrices are "semantically naïve": they know nothing about the meaning of the words; they just manipulate them. Topic modeling, however, does construct a new version of meaning for the tokens (words) in the system. While Mallet, the standard package used in the humanities for topic modeling, uses LDA (Latent Dirichlet Allocation) which is based on Bayesian probability, another, simpler approach uses a form of non-negative matrix factoring called PLSA (Probabilistic Latent Semantic Analysis).[5]

The Semantics of Matrices

Where does the semantics—the assignment of meaning—come in? Recall that the *columns* in the new matrix define the topics as arrays of words:

$$TopicWords = \begin{bmatrix} f(t_1w_1) & \cdots & f(t_iw_1) & \cdots & f(t_Kw_1) \\ \vdots & & \vdots & & \vdots \\ f(t_1w_j) & \cdots & f(t_iw_j) & \cdots & f(t_Kw_j) \\ \vdots & & \vdots & & \vdots \\ f(t_1w_N) & \cdots & f(t_iw_N) & \cdots & f(t_Kw_N) \end{bmatrix}$$

$topic(words)_i = (f(t_iw_1), \ldots f(t_iw_j), \ldots f(t_iw_N))$ [column vector]

At the same time, however, the *rows* in the matrix define the words through the topics in which they participate:

$$word_j = (f(t_iw_j), \ldots f(t_iw_j), \ldots f(t_Kw_j)) [\text{row vector}]$$

Since the topics themselves are mathematical constructs, defining the meaning of words as a vector of the weights they contribute to defining the array of topics may seem extremely abstract. However, we then take a next step of comparing the similarity of the words as defined by their role in the topics. The simplest approach is to take the normalized dot product of the two word vectors:[6]

$$similarity(\vec{w}_i, \vec{w}_j) = \frac{\vec{w}_i \cdot \vec{w}_j}{\|\vec{w}_i\| \, \|\vec{w}_j\|}$$

If the two word vectors are identical, the similarity = 1, and if they have no overlap at all, the value is 0. One then can use these similarity values to generate a *hierarchical clustering analysis* visualized as a tree graph (dendrogram) with branches that split in ever finer groupings to represent the clustering of words that share similarities in meaning.

The point to stress here is that this clustering of words by similarity is relative to a *specific collection of texts*. A different set of texts would produce a different clustering. This move from a collection of texts as bags of words to the combination of (1) the texts defined as arrays of topics and (2) the topics defined as arrays of words, and then to a hierarchical clustering of the words based on their similarity is a form of *distributional semantics* in which the meaning of words are defined through the patterns of their usage within a corpus. Even though the words remain a system of signifiers, this analytic approach from the digital humanities (originating in linguistics) takes us very far away from the enclosed world of poststructuralist analysis and much closer to the structuring of meaning in a textual corpus. That is, the algorithms here, seemingly a set of

---

[6]The dot product of two vectors is a single number (a scalar) calculated by taking the sum of each value in the first vector multiplied by the corresponding value in the second vector, i.e. $a \cdot b = \sum_{i=1}^{m} a_i b_i$. The norm (the length) of a vector is the square root of the dot product of the vector times itself, i.e., $\|a\| = \sqrt{a \cdot a}$. A vector defined as $\frac{a}{\|a\|}$ is a unit-vector (i.e., with a length of 1) in the direction of a.

mathematical functions, in fact embody and articulate a model for meaning defined through usage.

Meaning as Usage and the Hermeneutics of "Forms of Life": Wittgenstein and Dilthey

Some authors relate this meaning-as-usage to Ludwig Wittgenstein's famous dictum "For a *large* class of cases of the employment of the word 'meaning'—though not for all—this word can be explained in this way: the meaning of a word is its use in the language."[7] Thus Wittgenstein, like distributional semantics, assigns meaning according to usage. Wittgenstein arrived at his view of language when he confronted the failure of the more substantive model of traditional philosophy. In the model of logic inherited from Frege but with roots leading back to Plato, words refer to objects, and the truth of propositions relies on the correctness of the relationships they describe in the world. Wittgenstein rejected this appeal to reference to ground meaning and instead came to argue that the meaning of language comes simply and modestly from how humans use language. Wittgenstein's account, in other words, is *anti-foundational*: it rejects the possibility that human access to objects can serve as the foundation of knowledge and of language. Wittgenstein, encountering the failure of reference to provide meaning, developed an empirical response based on actual experience. He proposed language-games as the locus of meaning. Facing the same failure of reference, we now have turned to the digital humanities' ability to survey the vast corpora that document the human use of language.

Like previous scholars, however, I suggest that in assigning meaning to usage Wittgenstein was echoing—and perhaps drawing on—an earlier, broader tradition of interpretation in German hermeneutics from Schleiermacher to Dilthey. This hermeneutic tradition is of great significance to our understanding of the digital humanities as modes of exploring meaning.[8] The specific link between Wittgenstein and Dilthey in particular is in the concept of "forms of life" that Wittgenstein introduces—but does not expand on—in the *Philosophical Investigations*, the central work of his late career. Wittgenstein sees communication as possible because people are participating in the same language-games, but the question arises of how it is possible that people are playing the language-game in the same way, since rules cannot possibly specify all the variations allowable in a language-game. Agreement becomes possible, Wittgenstein argues, because people share a "form of life:"

> Here the term "language-*game*" is meant to bring into prominence the fact that the *speaking* of language is part of an activity, or a form of life.[9]
>
> "So are you saying that human agreement decides what is true and what is false?"—It is what human beings *say* that is true and false; and they agree in the *language* they use. That is not agreement in opinions but in forms of life.[10]

---

[7]Ludwig Wittgenstein, *Philosophical Investigations*, translated by G.E.M. Anscombe (Oxford: Basil Blackwell, 1976), 21.

[8]See Karl-Otto Apel, "*Wittgenstein* and the Problem of Hermeneutic Understanding," in *Ludwig Wittgenstein—Critical Assessments*, Vol. 4, ed. Stuart Shanker (London: Routledge, 1986) and Nicholas F. Gier, *Wittgenstein and Phenomenology: A Comparative Study of the Later Wittgenstein, Husserl, Heidegger, and Merleau-Ponty* (Albany: SUNY Press, 1981).

[9]Wittgenstein, *Philosophical Investigations*, 11.

[10]Wittgenstein, *Philosophical Investigations*, 88.

This grounding of agreement in shared forms of life resembles Wilhelm Dilthey's stress on the role of "objective mind" in providing a basis for mutual understanding: The citation here is long but important:

> I have shown how significant the objective mind is for the possibility of knowledge in the human studies. By this I mean the manifold forms in which what individuals hold in common have objectified themselves in the world of the senses. In this objective mind, the past is a permanently enduring present for us. Its realm extends from the style of life and the forms of social intercourse to the system of purposes which society has created for itself and to custom, law, state, religion, art, science and philosophy. For even the work of genius represents ideas, feelings and ideals commonly held in an age and environment. From this world of objective mind the self receives sustenance from earliest childhood. It is the medium in which the understanding of other persons and their life-expressions takes place: For everything in which the mind has objectified itself contains something held in common by the I and the Thou. Every square planted with trees, every room in which seats are arranged, is intelligible to us from our infancy because human planning, arranging and valuing—common to all of us—have assigned a place to every square and every object in the room. The child grows up within the order and customs of the family which it shares with other members and its mother's orders are accepted in this context. Before it learns to talk, it is already wholly immersed in that common medium. It learns to understand the gestures and facial expressions, movements and exclamations, words and sentences, only because it encounters them always in the same form and in the same relation to what they mean and express. Thus the individual orientates himself in the world of objective mind.
>
> This has an important consequence for the process of understanding. Individuals do not usually apprehend life-expressions in isolation but against a background of knowledge about common features and a relation to some mental content.[11]

Dilthey's argument is that we humans manifest our internal intentions in our actions and change the phenomenal world based on them. Although those intentions are not knowable in themselves, the world into which we are born and in which we live is shaped by the long history of intentional human structuring, and we learn to speak, act and think through the mediation of these humanly shaped forms. These are precisely Wittgenstein's forms of life. Because we share these forms, we understand one another. And if we seek to understand people from a different time or place, we need to understand the context of "objective mind" through which they thought and wrote. This is the hermeneutic project for Dilthey.

Dilthey's approach of seeking the totality of the mind-built world in which a person lived—a world of explicit traces in the sensory realm that mediate intentionality—was, like Wittgenstein's, a response to skepticism and the failure of meaning. As Dilthey explained:

---

[11]Wilhelm Dilthey, *Draft for a Critique of Historical Reason*, translated in Kurt Mueller-Vollmer, ed., *The Hermeneutics Reader: Texts of the German Tradition from the Enlightenment to the Present* (New York: Continuum Press, 1985), 155.

Today hermeneutics enters a context in which the human studies acquire a new, important task. It has always defended the certainty of understanding against historical skepticism and wilful subjectivity; first when it contested allegorical interpretation, again when it justified the great Protestant doctrine of the intrinsic comprehensibility of the Bible against the scepticism of the Council of Trent, and then when, in the face of all doubts, it provided theoretical foundations for the confident progress of philology and history by Schlegel, Schleiermacher and Boeckh. Now we must relate hermeneutics to the epistemological task of showing the possibility of historical knowledge and finding the means for acquiring it.[12]

Focusing on what is given, "the expression of what is expressed" in Dilthey's words, and grasping its meaning in a disciplined re-living through one's own experience makes understanding possible:

In such understanding, the realm of individuals, embracing men and their creations, opens up. The unique contribution of understanding in the human studies lies in this; the objective mind and the power of the individual together determine the mind-constructed world. History rests on the understanding of these two.[13]

Dilthey's goals here are very broad: he seeks to understand all of human experience through careful reflection on the manifest sedimentation of human intentions "from the style of life and the forms of social intercourse to the system of purposes which society has created for itself and to custom, law, state, religion, art, science and philosophy." More crucially, he asserts that *only* by such a broad reflection can one hope to understand another person or another time.

Dilthey's model for humanistic interpretation extends to aspects of human social organization that go far beyond the humanities and sees human production as drawing its material and meaning from all forms of the "objective mind." This understanding of human production—and thus of the humanities—as part of a broader matrix of meaning provides the underpinnings for the project of the digital humanities to search out regularities from the world of materials in which the objects of our inquiries are embedded. However, Dilthey's hermeneutics is so abstract that it does not offer concrete methodological models. For humanists, the model of Friedrich Schleiermacher, whose work on textual interpretation Dilthey extended and generalized, is more directly relevant.

Schleiermacher and Textual Hermeneutics

Schleiermacher, considered the father of both modern hermeneutics and modern Protestantism, was primarily a theologian, but he was deeply interested in the problem of understanding. The pressing task for him was to be sure that he understood the New Testament correctly, but he also was an innovative and acclaimed translator of Plato whose translations are still used today in Germany.[14] The story of the reasons behind his approach to hermeneutics gets very complicated very quickly, but it is nonetheless

---

[12]Dilthey, *Draft for a Critique of Historical Reason*, 162.

[13]Dilthey, *Draft for a Critique of Historical Reason*, 158

[14]For the connection between Schleiermacher's efforts as a translator and his hermeneutic theorizing, see, for example, Theo Hermans, "Schleiermacher and Plato, Translation and Hermeneutics," in *Friedrich Schleiermacher and the Question of Translation*, eds. Larisa Cercel and Adriana Serban (Berlin: De Gruyter, 2015), 77–106.

worth exploring, for the epistemological problematic that drove Schleiermacher's approach to understanding and the solutions he proposed for finding knowledge within that problematic are directly relevant to the situation of the humanities today and the role of digital humanities in reasserting the claims of humanistic knowledge.

Schleiermacher was part of the group of early German Romantic writers who were endeavoring to find ways to respond to Immanuel Kant's critical philosophy. Kant argued that not only do we not have access to objects in the world; we do not have inner access to our own self as the ground for experience. All we can know is within a phenomenal realm that is shaped in *a priori* ways by categories of perception we bring to the world to make experience possible. Among the major categories Kant proposed were subject-and-object, time-and-space, and cause-and-effect. We have no right to assume that these categories are actually part of the world, but we cannot experience the world without them. The post-Kantian early Romantic writers essentially worked within this epistemological critique that called metaphysically grounded foundational knowledge into question. Schleiermacher actually went Kant one better. While Kant asserted the necessity of the particular *a priori* categories of his analysis, Schleiermacher considered them to be as shaped by the same constraints of time and place as all other provisional human knowledge. Thus, while Schleiermacher was a theologian, he was a *post-Kantian* theologian whose approach to the religious understanding of the New Testament complemented his hermeneutic approach to texts. For Schleiermacher, the religious element in human experience lay in the capacity to have intuitions about unity that preceded any conceptual understanding of what that unity might be. This capacity is without any additional specific content. Thus, to understand any particular form of religious practice, one cannot bring any presumed content to one's observations and instead must rely on an understanding of the logic of practice within the particular community. The question, then, is how one understands another human community, once one sets aside access to universal truths. In particular, how is one to understand the Christianity as given in the New Testament without recourse to received dogma? Schleiermacher's hermeneutics provided his answer. For Schleiermacher, hermeneutics—the art of understanding—had two components. He asserted:

> 5. As every utterance has a dual relationship to the totality of the language and the whole thought of its originator, then all understanding also consists of the two moments: of understanding the utterance as derived from language, and as a fact in the thinker.[15]

Schleiermacher elaborated on these two moments:

> 5.3. According to this, each person, on the one hand, is a location in which a given language forms itself in an individual manner; on the other, their discourse can only be understood via the totality of the language. But then the person is also a spirit which continually develops, and their discourse is only one act of this spirit in connection with the other acts.[16]

---

[15]Friedrich Schleiermacher, *Hermeneutics and Criticism and Other Writings*, translated and edited by Andrew Bowie (Cambridge: Cambridge University Press, 1998), 8.

[16]Schleiermacher, *Hermeneutics and Criticism*, 8–9.

Schleiermacher is not naïve here. He knows that understanding, as he presents it, requires a totality that is impossible. He argues instead that one makes progressively greater sense of the fragments one *does* know through the sorts of intuitions without rules that Kant defined as aesthetic judgments. Thus, Schleiermacher further asserted:

> 9. *Explication is an art*
>
> 1. Each side on its own [is an art]. For in every case there is construction of something finitely determinate from the infinite [and] indeterminate. Language is infinite because every element is determinable in a particular manner via the rest of the elements. But this is just as much the case in relation to the psychological side. For every intuition of an individual is infinite. And the effects on people from the outside world are also something which gradually diminishes to the point of the infinitely distant. Such a construction cannot be given by rules which would carry the certainty of their application within themselves.
>
> 2. For the grammatical side to be completed on its own there would have to be a complete knowledge of the language, in the other case [the psychological] a complete knowledge of the person. As there can never be either of these, one must move from one to the other, and no rules can be given for how this is to be done.[17]

The Hermeneutics of the Digital Humanities

With this pair of impossible complementary tasks—the demand for complete grammatical and psychological understanding—defining the hermeneutic endeavor, we at last arrive at our destination. I assert that the project of the digital humanities in our own day, with our own epistemological problematic, is a continuation of Schleiermacher's infinite project of understanding. Schleiermacher, working within strong prohibitions against foundational knowledge of either the self or the world, turned to synthesizing the patterns of the details of what we *can* know of the middle realm, the world as given to human experience.

In reading texts, Schleiermacher required a difficult, disciplined synthesis of broad and deep knowledge and intuitions binding the disparate forms of information into that moment of synthesis. His approach here entirely recasts the distinction between close reading and the various methodologies in the digital humanities that come under the rubric of distant reading.[18] Each text is abuzz with patterns—patterns of language usage as well as the patterns of historical and social interaction that shaped the author in its writing. The modes of distant reading powerfully search through textual corpora on a scale that humans cannot hope to match, and provide a background of linguistic behavior for the texts we read. At the same time, as Schleiermacher pointed out, the particular texts we engage are also moments in human experience. On the one hand, the authors writing them are embedded in the structures of their society, culture, and language, and on the other, their writings diverge from the givenness of these structures and reflect particular intentions at a particular time and place. We need sensitivity to discern these divergences. Thus close reading remains vital, but given the growing availability of distant readings, the demands placed on close reading change. Close

---

[17]Schleiermacher, *Hermeneutics and Criticism*, 11.

[18]For a useful account of distant reading, see S. Jänicke, G. Franzini, M. F. Cheema, and G. Scheuermann, "On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges," in *Eurographics Conference on Visualization*, eds. R. Borgo, F. Ganovelli, and I. Viola (EuroVis) (2015), www.informatik. uni-leipzig.de/~stjaenicke/Survey.pdf.

reading opens out when we see the text as the *locus of synthesis for all the myriad patterns disclosed by distant readings* of which the text is a part (the modern version of Schleiermacher's project).

Consider some of the options offered by distant reading. We already have seen how distributional semantics can provide hierarchical clustering for the language in a corpus of texts. We can apply this approach to particular genres of history, literature, and religious and philosophical discourse within a given period to see if there are distinctive patterns of usage for each genre of which we need to be aware as we read. We can compare usage within a genre across time, with obvious examples like the scholarly notes (*biji* 筆記) from the Tang to the Southern Song.

Stylometry for Classical Chinese is still a work in progress, but we can sharpen its techniques as we explore the rise and complex dispersions of genres. Here I would cite Paul Vierthaler's exploration of various subgenres of Chinese fiction in "Fiction and History: Polarity and Stylistic Gradience in Late Imperial Chinese Literature."[19] Similar approaches can be applied to the whole range of informal writings in the Song or between early and later *biji* as the genre develops. Having aggregated data, we can see the distinctiveness of particular texts as identified by their divergence from the collective metrics.

Another important development in distant reading is the effort to identify intertextuality, as in the work on Latin texts by Walter Scheirer and others in their essay "The sense of a connection: Automatic tracing of Intertextuality by meaning."[20] (I confess I am astonished at how well their approach works given the state of the tools they bring to it.) This question of intertextuality is vitally important for the reading of Chinese literati texts, in particular, since so much of the connoisseurship in the close reading of Chinese poetry and prose from the Song dynasty onward is in the identification of allusive reference.[21] In the tradition, this search for allusions appears to be an effort to assert mastery and control the meaning of texts, but it relies on a rather arbitrary methodology. I will be very interested to see what large-scale intertextuality studies turn up and how those results will complicate close reading.

What I present here is just a very partial list of the role of distant reading in giving us important information about word-usage, genre, and intertextuality. There is much more that I have not seen and yet more in the offing, where scholars are still playing with the possibilities of current tools and learning new approaches to tagging texts that then can be used to greatly extend the power of the basic techniques we now have. These are exciting times, and it is my hope that we never will be able to look at texts the same way again. As we read and interpret, they will be deeper and more demanding because of the work of the digital humanities.

---

[19]Paul Vierthaler, "Fiction and History: Polarity and Stylistic Gradience in Late Imperial Chinese Literature," *Cultural Analytics* May 23, 2016, DOI: 10.22148/16.003.

[20]Walter Scheirer, Christopher Forstall, and Neil Coffee, "The Sense of a Connection: Automatic Tracing of Intertextuality by Meaning," *Digital Scholarship in the Humanities* 31.1, (2016), 204–17.

[21]Donald Sturgeon, as part of his important collection of digital texts, has created a platform for looking at text reuse that will be an important tool for exploring intertextuality in pre-modern Chinese texts. See his discussion of the tool in Donald Sturgeon, "Digital Approaches to Text Reuse in the Early Chinese Corpus," *Journal of Chinese Literature and Culture (JCLC)* 5.2 (2018), 186–213. From this special issue of *JCLC* devoted to digital humanities, also see Yi-long Huang and Bingyu Zheng, "New Frontiers of Electronic Textual Research in the Humanities: Investigating Classical Allusions in Chinese Poetry through Digital Methods," *JCLC* 5.2 (2018), 411–37.

The China Biographical Database (CBDB) as Hermeneutic "Technical" Interpretation

The digital techniques I have mentioned so far all center on what Schleiermacher called the "grammatical" mode of hermeneutics. In contrast, Schleiermacher's "technical" moment focuses on the question of "how the author arrived at the thought from which the whole developed, i.e. what relationship does it have to his whole life and how does the moment of emergence relate to all other life-moments of the author."[22] Thus the central challenge of the psychological component of hermeneutics is how to understand the larger life patterns within which an individual (or an era) lives.

The goal of presenting a systematic analysis of the "human sphere" informing life in premodern China drives the China Biographical Database project, where I am the chief data architect. Our goal is to systematically collect what data we can on the key structures shaping social experience in pre-modern China. We are acutely aware of data that we *cannot* collect and the limits of what we offer. For example, parish records in England allow historians to inventory people's worldly goods, but there are no corresponding sources of information for China. The information we have also is strongly biased toward the elite stratum that shows up in histories, including local gazetteers. Careful sifting of Buddhist and Daoist records allows us to know something about the lives of monks, but we know next to nothing about merchants, farmers, and artisans throughout Chinese history. These lacunae seriously distort what we can know of social experience in pre-modern China. Fortunately, it turns out that *most* of the extant authors from pre-modern China were from the elite stratum about which we can say a good deal.

The extant historical record allows us to track a range of important social and institutional systems that structured elite life in pre-modern China. Kinship relations, of course, come first, and then social relations, and locality. The examination system and one's place within the imperial bureaucracy also loomed large in the lives of many of the authors we read. When Harvard inherited the database from Robert Hartwell, all of these components already were included in his data structures. I cleaned these up a bit, but the only significant addition I made to the types of data was that of "social institutions," since we discovered that such entities as private academies and temples were institutions around which members of the elite stratum formed communities to achieve collective goals. The major innovation I brought to the functionality of the database was my realization that we could exploit the hybrid nature of the system on which CBDB ran. Hartwell created the initial database in dBase, an old database programming language. When I reconfigured the database in FoxPro, a close cousin of dBase, I realized that we could exploit the initial kinship and social relationship information we had for individuals by recursively searching through them. That is, we start with the kinship information for an individual, and then we add all the kinship information for that person's kin, and then we add the kinship information for all the newly discovered people, and so on, until the kinship distances reach a limit set by the end user. This sort of recursive search is relatively easy to set up in a procedural programming language like FoxPro or VBASIC, which Microsoft Access uses as its back-end programming language. It is much harder to build into SQL, Structured Query Language. In any case, I set the system to loop through *social relations* data in the same way to build social networks that could be exported to Social Network Analysis packages like Gephi. And in addition, we allowed the system to mix and match: to

---

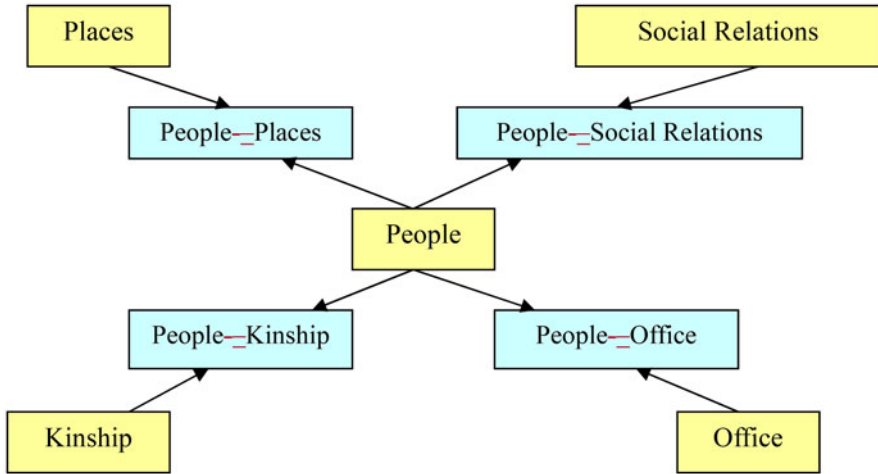[22]Schleiermacher, *Hermeneutics and Criticism*, 107.

Figure 1. Simplified Diagram of CBDB Relational Database Structure

pull in all the social relations of kin, all the kin of people in one's social network, and every other possible combination of kinship and social relationship.

Although Hartwell designed his database to allow him to look at the connection between office-holding and kinship, I don't think he quite realized what an extraordinarily powerful database he had developed. Schleiermacher asserted the relevance of the totality of social structures impinging on the individual. Because CBDB has tables representing the components of kinship, social network, social status, office-holding, and locality in an individual's life, it allows a scholar to explore their interactions in a corollary to Schleiermacher's proposed methodology. That is, a simplified version of the CBDB structure looks like Figure 1.

That is, people are at the center, and through them one can link all the additional components of social organization. We can ask questions like "Was the role of medical officer hereditary, that is, were medical officers the sons or nephews of medical officers, and did the families of medical officers marry their children to one another?" (Figure 2).

We can ask yet more complicated questions. Were officials from Fujian more likely to develop local kinship networks than were officials from Zhejiang? Did patterns differ depending on rank, and did the patterns change over time? This adds the dimension of locality (Figure 3).

When I was reading the writings of Liu Kezhuang 劉克莊 (1187–1269) from Fujian, this question was of considerable importance. Indeed, in my recent book on Southern Song poetry, being aware of the interactions among locality, kinship, social networks, participation in the examination system, and office holding compelled me to rethink the usual understanding of the "Rivers and Lakes" poets and realize that the usual story was wrong. There clearly were large networks of men from important local lineages who participated in the examination system as proof of their elite status but who had little hope to actually succeed and little interest in serving more than would be required to confirm their tax exemption. Instead, they traveled from patron to patron talking and writing and, frequently enough, joining in protest against the current imperial administration. Reading their poetry in this context allows one to develop a more nuanced understanding of their work.
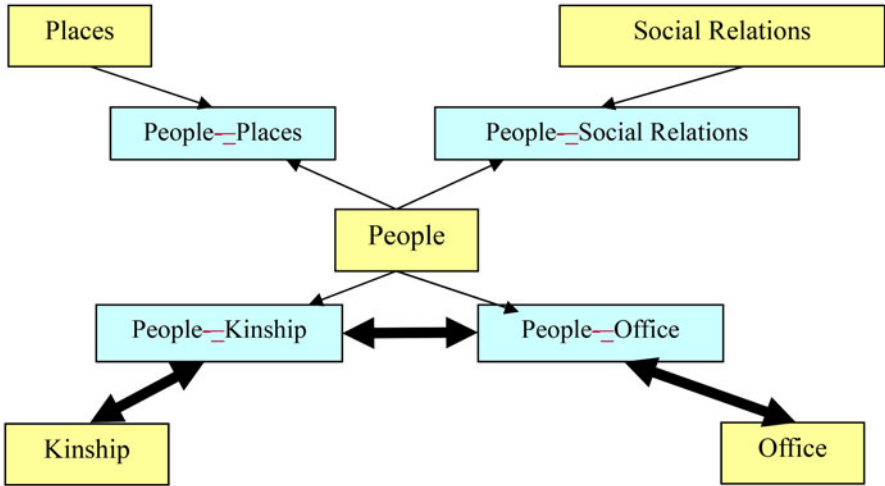
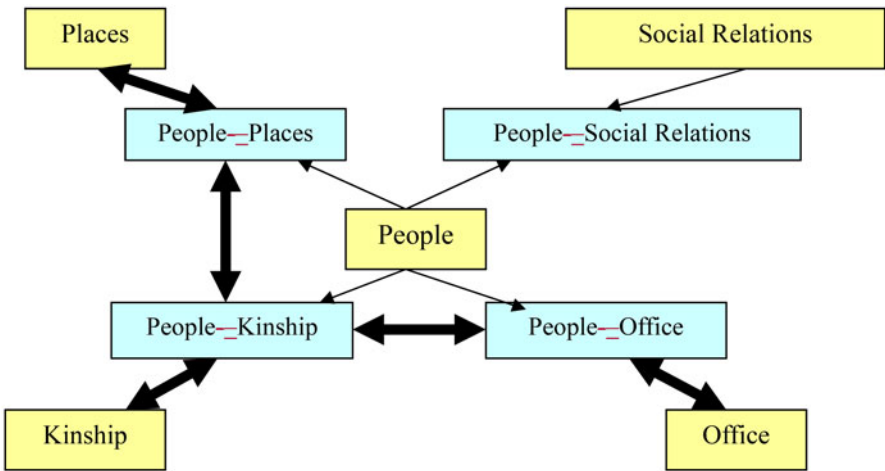**Figure 2.** Querying the Relation of Office and Kinship Networks

Figure 3. Querying the Relation of Place, Office and Kinship Networks

The Digital Humanities and the Connectedness of Meaning in Human Experience

Schleiermacher and Dilthey were circumspect and methodical in their efforts to allow contemporary readers to understand the lived significance of texts of the past. The postulates they needed to justify their methods were—keeping with their Kantian model— fairly minimal: people in the past were biologically similar to people today, and people acted with motives.[23] An understanding of all else that is built upon the basic hardware

---

[23]Dilthey's model for historical knowledge requires the postulate that people act with motives, either as 意, intentions, or as 情, feelings rather than randomly. This assumption is a corollary to Kant's postulate of the purposiveness of Nature as a totality behind the aesthetic judgments that are the first step toward

—from what we now consider the history of affect, to religious and philosophical systems, to literary history—must be constructed from a careful, reflective consideration of the texts (and other artifacts) that survive. And there was no absolute certainty in the conclusions they could draw, given the pastness of the past and the limitations of our sensibilities and the data that remains. This is precisely our current situation, except that the new digital technologies allow us to extend the range of our data and our ability to organize it in ways that they could not have imagined.

The sort of empirical results for individuals and for larger aggregations of people that are derived from network analyses and other forms of statistical analysis are not a reduction of people and texts to numbers but a hermeneutically compelling way to discover the large-scale patterns of the social world that informed people's lives. These results provide contexts for reading and thinking, and they put demands on reading. My understanding of the writings of the large community of "Rivers and Lakes" poets of early thirteenth-century China became profoundly different once I incorporated the myriad factors—from the evolving nature of local elites, to the examination system and the role of printing, to the details of the *Daoxue* networks (and their arguments)—that shaped the historical context for poetry at the time. The model of CBDB and access to large repositories of digital texts were both crucial to reconceiving how the texts I studied were connected and where their meaning resides: within the texts, certainly, but texts, the digital humanities shows us, are not just singularities; they radiate outward and are traces of moments of experience at the intersections of complex multidimensional patterns that the digital humanities can partially—if never fully—restore.

We live within epistemological impasses, but with the help of the digital humanities, we, like Schleiermacher, Dilthey, and Wittgenstein, are turning to the large world of human experience to see what we can learn about who we are. We as humanists have much to contribute to this project that we share with scholars pursuing other forms of inquiry. We have a great future if we open ourselves to the challenges of this shared endeavor and learn to see the methodologies for large-scale analysis as integral to our inquiry into the human.

---

empirical. Intentions, however, apply to the "life-world" of human experience rather than to the larger phenomenal realm.