PART I

GLOBAL HISTORIES OF DIGITAL HUMANITIES
This page intentionally left blank
There is an epistemological puzzle that is likely to become increasingly acute as the field of digital humanities grows more diverse and encompasses, more and more, knowledge that lies outside of the normative expectations that most existing knowledge infrastructures assume by default. Technosocial ensembles acting as apparatuses to create computationally inferred knowledge are power-laden, in the sense that the epistemological assumptions built into them have the power to shape the outcome of produced knowledge. The motivation for this chapter comes from my experience with a text analysis tool that visualizes the properties of a digitized text corpus in relation to trends in the usage of specific words in the corpus (Auvil et al., “Exploration of Billions of Words”), but the resulting insights can, I think, be generalized to apply to many kinds of digital knowledge infrastructure.

Invisibilization as an Epistemological Problem Accentuated by Technology

I have described in previous work (Bhattacharyya, “Words in a World of Scaling-up”) how a powerful state-of-the-art tool for text analysis and visualization nevertheless tends to undercount occurrences of material in non-Western languages, especially from the Global South, encountered in roman script in Western-language texts. Such words tend to suffer an “invisibilization.” This is because data that is epistemically heterogeneous can become illegible within a representational scheme that enforces standardization. As the normative assumptions of the scheme tend to privilege standardized data, data that is nonstandard in some sense registers as visible only when made to give up its variation and conform to the norm.

I showed in that earlier work that the problem arose from a scale effect, combined with the fact that when transliterated into roman script, morphological forms of non-Western-language words are representationally more heterogeneous than the normative epistemic assumption of the tool can easily allow for. The large size
of the corpus and the long-tailed frequency distribution of words in human language together mean that the large corpus will inevitably contain many words that occur only a small number of times; hence words occurring fewer times than some threshold value, such as non-Western-language words in roman script, will inevitably be ignored. Even where of specific interest, such words cannot be protected from such threshold-induced cutoffs, for to do so would require knowing, a priori, what those words might be. One could imagine compiling a list of such words by anticipation (for example, by consulting dictionaries of the language in question). However, non-Western-language words tend to have too much and too unpredictable orthographic variation when transliterated into the roman alphabet, making such an approach unviable.

I also argued that the problem was similar in some ways to that of translating world literature into Western languages. As already explained, when non-Western-language words are transliterated into the normative representational scheme of the knowledge apparatus (here the roman alphabet), they tend to show much more representational variation than such tools implicitly expect. Similarly, the nuances of a language cannot easily be encoded into specific, determinate words in another language, while extensive glossaries or footnotes might make the translation unworkable. This is a general conundrum beyond particular constellations of words, alphabets, and languages, and an instance of a general problem of invisibilization specially accentuated in digitally mediated inquiries. Nonnormative knowledge objects, such as those from the Global South, are particularly vulnerable to this risk. The greater the distance of the cultural object from the metropolitan center, the greater, as a rule, the extent of this nonconformity and greater, consequently, the chance of knowledge objects undergoing occlusion and invisibility. In this chapter, I extend this argument to show how cultural forces, through the sociotechnical ensemble to which they belong, determine what it is that digital processes and infrastructures make more (or less) visible.

**Cognitive Capitalism as Discursive Framework: Scale Effects and Network Effects**

While information itself is a nonmaterial abstraction, processes and systems, which combine small quanta of data to produce larger data ensembles, are intertwined with materiality. These systems, across which procedural operations are transacted to produce knowledge, have material effects because they are not purely informational. In the context of digital humanities, the epistemological dilemma I am describing has the potential to affect the legibility of cultural production from the Global South, not only because of the relations of power baked into our knowledge apparatuses but also, and perhaps more crucially, because of scale effects and network effects. Both knowledge discovery on a large scale and
knowledge discovery using networks can exacerbate the illegibility of forms of knowledge that are either intrinsically heterogeneous or simply expressed differently, often on a different scale.

A discursive framework is needed as a common language to talk about these issues. A challenge for digital humanities is that its close relationship to technical fields means that as Alan Liu (“Where Is Cultural Criticism in the Digital Humanities?”) has pointed out, criticism, as practiced elsewhere in the humanities, often takes a back seat in the digital humanities in the absence of an adequate framework for understanding and analysis. A framework for thinking about these epistemological issues must often be created by adapting theories developed for other purposes and domains.

I propose that the logics of hierarchical and nonhierarchical production and accumulation—whose function in the contemporary sphere of political economy is noted by the economic sociologists Luc Boltanski and Eve Chiapello (New Spirit of Capitalism)—provide a useful framework for theorizing some of the epistemological issues I describe. While Boltanski and Chiapello concern themselves with the analysis of value creation in society as a whole, I argue that their analysis can also be applied to value creation in technosocial apparatuses of knowledge, including information processing. Nonhierarchical value creation is a form of what the sociologist Yann Moulier Boutang describes as “cognitive capitalism” (Boutang, Cognitive Capitalism): it converts networks into value by multiplying the points of contact with human activity. The frequency-driven registering of data in knowledge apparatuses, such as the tool I mentioned, is an example of this: data registers as data only when it exceeds some threshold frequency of occurrence. For example, in the case above, a word is registered as such only by having relatively many points of contact with the corpus—that is, relatively many occurrences. Boltanski and Chiapello point out that the “new spirit of capitalism” is as interested, in principle, in favoring these network forms constituted by points of contact as it is in accumulation of the kind traditionally favored by capitalism.

Their discourse aligns itself, in principle, with the social critique of exploitation and emancipatory positions on social justice (Boltanski and Chiapello, New Spirit of Capitalism, 101). The invisibility created in these types of knowledge apparatuses is therefore not due to deliberate omission, nor brought about by everyday social practices, habits, and prejudices. Instead, it is a structural problem, somewhat similar to the kind Marisa Fuentes (Dispossessed Lives) describes as embedded in the dynamics of the knowledge-producing system itself. However, while in the situation described by Fuentes the structural problem arises from the content of the archive (the power and apparent authority of material authored by a dominant social group), in the digital knowledge apparatus the problem inheres in how normative assumptions unintentionally shape the design decisions themselves, making the problem primarily epistemic as opposed to merely structural.
Cognitive Capitalism Extended to Digital Knowledge Apparatus

There is, of course, considerable room for debate about the applicability of Boltanski and Chiapello’s framework to digital knowledge apparatus. For one thing, while Boltanski and Chiapello are primarily concerned with value creation in capitalism, the kind of inquiry that digital knowledge apparatuses support is usually concerned with value creation only in a symbolic sense. Are digital knowledge objects the kinds of objects to which notions such as accumulation legitimately apply? Are they constitutive of a political economy of production and circulation? Would attempts to apply such notions to them lead to category errors? Victoria Gritsenko and Vladimir Orlov (“Universal Labor”) have argued that in postindustrial society, universal or automated labor cannot be properly accommodated within the frame of commodity value; instead, it may need to be understood in terms of informational value. Information is a nonrival good, while material commodities are rival goods. Duplication of information does not cause the previous copy to go away; information can therefore be in multiple places at the same time, in a way that a material commodity cannot. Perhaps these objections can be countered if we think of digital inquiry as simply a special instance of the digital governance of things human. Information is produced by the transformation of material structures by the computer: as a measure of organization and complexity, it is, after all, a part of the world of material systems and processes. Viewed in this way, thinking of digital humanities with and through Boltanski and Chiapello’s framework seems justified.

There is, in fact, a significant history of thinking about the circulation and aggregation of digital knowledge objects in varied contexts. As Rob Kitchin and Tracey Lauriault (“Small Data, Data Infrastructures and Big Data”) point out, the production of knowledge, especially academic knowledge, has progressed over the past few centuries using what may be called “small data” studies, that is, inquiry based on very selectively sampled data generated to answer specific questions. We can think of this as constitutive, in the humanities, of the paradigmatic activity of close-reading small bodies of text. The use of computational techniques based on big data offers an alternative epistemology. Kitchin and Lauriault point out that one consequence of the onset of techniques for analyzing big data is that “small data” is made more like big data through the development of new data infrastructures that “pool, scale and link small data” into larger datasets, opening them up to inquiry using “big data” analytics.

Foucauldian critiques of governmentality also provide a useful point of reference for thinking of knowledge objects in such configurations. Partha Chatterjee, for instance, argues that knowledge objects might be conceived as constituting the concrete embodiment of a rational consciousness (Chatterjee, Nation and Its Fragments, 207). Building on Chatterjee’s insights, Kalpana Ram points out that the governmental practice of “carving up” knowledge into discrete, disjunct domains
works significantly to render “minor” knowledge objects invisible if they do not conform to the epistemological assumptions behind such compartmentalization (Ram, “Silences in Dominant Discourses,” 125). The knowledge objects, in her case, are “practices” in the sense used by Michel de Certeau (Practice of Everyday Life), with the practices occluded by the epistemic framework rendered “minor” through such occlusion, in contrast to the salient foregrounded practices that constitute and organize society’s normative institutions (Ram, “Silences in Dominant Discourses,” 48).

INVISIBILIZATION AS AN EFFECT OF THE DIGITAL ON THE SUBALTERN

In earlier work, I have drawn attention to the connection between such occlusion and Gayatri Chakravorty Spivak’s key insight that subaltern voices are rendered inaudible unless already translated into a register that conforms to epistemic normativities (Bhattacharyya, “Words in a World of Scaling-up”; Spivak, “Can the Subaltern Speak?”). There is groundwork already in place for thinking about epistemological questions in the context of knowledge apparatuses broadly understood, including issues of particular importance in the context of the Global South. This can make for a productive future engagement between digital humanities and the broad field of postcolonial studies and Global South studies.

However, while my example above pertains to text and corpora, these issues encompass the entire gamut of the digital humanities (and not just digital humanities in, or about, the Global South either—though, of course, it is in the South that they tend to have more impact). For instance, the spatially oriented subfields of digital humanities face a similar challenge. Shannon Mattern has examined how representations of spatial data can produce this kind of invisibility or illegibility (Mattern, “Gaps in the Map”): geographical information systems, a key technology for spatial data, are often viewed by humanists as offering a reductionist epistemology that forces data into categories, as noted, for example, by David Bodenhamer (Bodenhamer, 24). Maps cannot easily represent counterfactuals like “buts,” “ifs,” and “however,” or metaphors. For that matter, anything not amenable to “an Enlightenment logic in which everything can be surveyed and pinned down”—anything, for example, that is only “partially” locatable in time and space—cannot be adequately represented visually in maps without undergoing some kind of epistemic violence (Pile and Thrift, “Introduction,” 1). Mattern has also drawn attention to another kind of invisibility to which digitally produced artifacts are particularly prone: the invisiblization of context. Referring to 3D-printed replicas of archeological artifacts from elsewhere that are erected in metropolitan centers, she observes how the relative ease with which such replicas can be produced tends to divest it of its “context and conditions of production” (Mattern, Code and Clay, Data and Dirt). Irit Rogoff notes a similar invisibility in physical artifacts, which is created by curatorial practice: “the good intentions of recognition,” she writes, “become a substitute . . . for
detailed analysis” by the public when curatorial activity hews the objects to existing norms and fails to challenge the audience to think (Rogoff, “Looking Away,” 1).

In such cases, the network effect, measuring utility reductively by the number of contact points, can drown out the factor of what those contacts actually are: mere numbers, often standing in decontextualized splendor in a technocratic landscape. Yet another pathological consequence of the network effect is the salience of incorrect yet authorized information, when authorization is determined by a network. The network effect, amplified by a scale effect, can then spread the incorrect information at runaway speed, or sometimes even create artificial “truths” that are reified iterations of claims about nonexistent “facts.” The journalist Jack Nicas (“As Google Maps Renames Neighborhoods”) has noted recently how the “rebranding,” intentional or otherwise, of particular neighborhoods by giving them new names, carried out by technologically knowledgeable users who submit the changes to Google Maps, often leads quickly to the previous names falling into disuse. Since Google creates and updates its maps from third-party data, the network effect ensures a kind of positive feedback loop. It is the familiar story of long-tailed data distributions growing even more long-tailed over time in the political economy of data circulation; also of the winner-takes-more phenomenon, exacerbating the inequality of power and influence—in this case, the power to rename. Alongside the network effect, the scale effect, resulting from Google’s massive user base, reinforces this outcome of data circulation and drowns out previous mechanisms of circulation, such as word-of-mouth dissemination of place names, that could have served as a corrective.

Significantly, Nicas suggests that for places in the Global South, the replacement of old names by new artificial ones can be particularly egregious. In India, for example, submissions to Google are vetted by contractors “with little local knowledge.” Although Nicas does not offer any explanation, this state of affairs can perhaps be attributed to the way the technocratic digital imagination often errs, as I pointed out earlier, in thinking that a one-size-fits-all approach will work everywhere. In reality, the kinds of social actors who possess local knowledge in the global metropole may be very different from those who do so in the global periphery: one size does not fit all.

Data from the Global South Is More Heterogeneous and in Unpredictable Ways

In a recent instance from India, epistemic normativity assumed by the technocratic digital imagination unwittingly led to massive disruption. In 2016, the government carried out a “demonetization” by decree, whereby currency notes of higher denominations ceased overnight to be legal tender. This was presented as an attempt to crack down on cash money circulating in the parallel economy operated by tax evaders. For a limited period, people could exchange a small number of these high-value
currency notes for new, redesigned ones, the hope being that those who held stashes of “black money” would be left with large amounts of unusable cash. There was a simultaneous push toward more electronic transactions, which would ease the pressures on the banks.

In point of fact, demonetization led to severe hardship for all too many honest poor people. This was because, as the economist Esther Duflo observes, the “fad of the moment” was converted into policy with little thought for implementation, “without having the gears in place” (Kishore and Bhattacharya, “We Might Never Know the Real Pain of Note Ban”). The government of India had expected a rapid uptake of digital, “cashless” electronic transactions by the population to tide over the scarcity of cash. However, the government had ignored the fact that much of the population participated in an informal economy where transactions were operated through social networks: they lacked the necessary tools and facilities to make a sudden transition to a formal economy and largely cashless transactions. Moreover, there was no effective mechanism to measure the gross domestic product (GDP) generated by this informal economy, since that GDP was routinely calculated by simply indexing it to the GDP of the formal economy—that is, calculated normatively. The government simply used those figures to argue that there was no significant hardship, whereas in truth, “we might never [even] know the exact magnitude of loss.”

In other words, members of the heterogeneous informal sector not only suffered from the normative assumptions of digital technocracy, as their network-based economic participation was subsumed by the state within the formal economy; on top of that, their suffering was rendered illegible or invisible. Tony Joseph observed at the time how this created a significant loss of jobs, as the formal sector drove out the informal sector. While the transformation of the informal sector into the formal sector should have been a gradual, phased process, allowing enough time for the different actors in the economy to adjust and evolve, “fast-forwarding this without safety nets” in an economy that “hasn’t taken care to provide its citizens with enough education and skills” was disastrous (Joseph, “Understanding Demonetisation”). This is reflective of a pattern found in marginalized entities everywhere. As Tim Unwin (“Digital Economies at Global Margins”) reminds us, for example, making people use the internet can widen the “digital divide” between digital haves and have-nots, setting back, relatively speaking, those who do not have the skills to leverage its opportunities.

There is an even more insidious problem here. The crucial point is that, even if the government had sought to provide safety nets, the attempt would have failed, since in an informal-sector economy like India’s, with all too much context-dependent heterogeneity, it would have been impossible to anticipate all the situations requiring such safety nets. We can see that this situation is analogous to that of the text-analysis tool I described earlier, where affirmative steps to make visible a set of words from non-Western languages, transliterated in the roman alphabet,
within the knowledge generated by the tool could not possibly succeed, as such words could not be isolated and identified a priori, on account of their propensity to occur in various nonstandard, heterogeneous spellings.

**Digital Humanities as Analogous to “Modernist” Developmentalist Ideology**

Even if it were possible, for argument’s sake, to carry out such an a priori enumeration in some situations, it might still not suffice, as such special provisions run counter to the normative assumptions of hegemonic epistemologies such as those of neoliberal capitalism. One of these normative assumptions is deregulation: the requirement that all special provisions for protection be undone or minimized in order to make all transactions smooth. A second is the notion that specific local conditions or particularities are merely a market distortion impeding a “natural” hierarchy of winners and losers. This, quite logically, leads to a one-size-fits-all approach: special context-based provisions are anathema to the neoliberal technocracy. This becomes particularly problematic in the Global South, whose people and places—and, increasingly, data and knowledge objects—are most in need of such protection owing to the uneven development of capitalism.

As I mentioned at the start, an approach eschewing the one-size-fits-all approach of the knowledge apparatus can serve to correct the problem. But algorithms tend to work most smoothly and efficiently when the data on which they act are most uniform. For the textual tool I described at the start, a possible approach would have been, as I mentioned, to assign different thresholds for certain words to the greatest feasible extent. However, such special provisions would interfere with the smooth functioning of the tool. Similarly, in a neoliberal market, such special arrangements would be considered market distortions interfering with the smooth, unregulated functioning of the market.

Extending the metaphor further, we could think of the context-free visibilization of individual words from digitized text as similar to what James Scott (Seeing Like a State) has described as the “high modernist” ideology of development in the twentieth century: a decontextualizing, top-down approach that enforces standardization, thereby making illegible those complex forms of local knowledge that are highly enmeshed in their context but ineffective when decontextualized into a standardized knowledge archive. Through demonetization and cashless transactions, the government sought the top-down standardization and integration of the transactional economy, but the move rendered the plight of small-time, informal-sector workers illegible and their cash unusable, unless the money was first exchanged, in painstaking steps, in formal institutions like banks. Likewise, in the case of the text-analysis tool I described, the words, heterogeneously spelled when transliterated into roman script, become invisible without an initial painstaking operation to replace them with transliterations that standardized the spelling.
Remediation, Technological and Otherwise

What are the possible ways of correcting epistemically produced invisibility? In the case of text, some interesting work is being done in computational linguistics to address the issue of variation and heterogeneity—for example, by Jacob Eisenstein and his collaborators on socially linked variation (Yang and Eisenstein, “Overcoming Language Variation”). Although the aim of such work tends to be utilitarian—namely, robustness in the face of sociolinguistic variation (Eisenstein, “What to Do About Bad Language”), taking variation as a problem to be “solved”—the approach holds much promise for incorporating expectations of variation and heterogeneity into the set of normative assumptions that knowledge apparatuses implicitly entertain.

But these are technological solutions. What can digital humanists, who are rooted primarily in the humanities, do to address such invisibility? While I think that exploring the questions might be more interesting than trying to propose solutions, here are some thoughts on the subject, in relation to the digital tool for text which has been the motivation for much of my thinking in this area. In undergraduate classes in the humanities, where I have used this tool, a promising corrective has emerged in what might be called “persistent annotation”: a way for the student users to annotate the invisibilities/illegibilities as and when they discover them—for example, in the form of a lasting written record passed on from one iteration of a course to another. A more sophisticated form of the solution could incorporate that record, in the form of a user-contributable manifest, into the tool itself—for instance, by including a visible pointer to such a manifest within the interface for the tool. However, for my relatively limited case, consisting of students in a course that is offered multiple times, something as simple as a document carried over and renewed from semester to semester across the content-management system for the class could suffice as a manifest. This is roughly similar, in principle, to the way users of Wikipedia can make edits (or, in a closer analogy, editing suggestions) that leave an audit trail of accountability.

Bernard Stiegler’s term for these technologies of digital humanities is “tertiary retention” (Stiegler, States of Shock). Drawing upon Jacques Derrida, Stiegler thinks that these tools bypass the problem of the finite capacity of human capabilities by means of exteriorizing those capacities, in the same way that writing does for speech. In another Derridean reference, Stiegler deconstructively presents these technologies as pharmacological, in the double meaning of “pharmakon” as both toxin and remedy. Using a crowdsourced—that is, exteriorized—technique of the kind I suggest, which depends on tertiary retention, to correct the problem of invisibility created by the epistemically dominant application of tertiary retention in the first place, may be thought of as a practice of “positive” pharmacology in Stiegler’s sense (Stiegler, States of Shock, 155–62).

The important general point to draw from here is that more attention needs to be given simply to understanding, studying, and theorizing our infrastructures
for producing knowledge, probing them attentively to see if the specific challenges posed by data in the humanities coming from the Global South are adequately met. This can only happen when, to borrow a phrase from Geoffrey Bowker and his colleagues, we move from treating infrastructure as a substrate to treating infrastructure as a substance in its own right (Bowker et al., “Toward Information Infrastructure Studies,” 99). As digital humanities tools become more ubiquitous and crowdsourced, and get taken up more and more by users in the Global South, such crowdsourced correctives to epistemically induced invisibility are likely to grow in importance.

Bibliography


