1

Beyond retrieval? Computer science and the humanities

Plenary lecture for the CATCH Midterm Event
Den Haag, Netherlands
30 November 2007

> A main source of our failure to understand is that we do not *command a clear view* of the use of our words.—Our grammar is lacking in this sort of perspicuity. A perspicuous representation produces just that understanding which consists in 'seeing connexions'.
>
> Ludwig Wittgenstein, *Philosophical Investigations* 122

> We look back along the wall: could we not pull it down, has it always been there? As we scan its windings over hills and vales back in history we behold a land far, far away… where the wall flattens and disappears, and the path was not yet split, but was only *one*.
>
> Erwin Schrödinger, *Nature and the Greeks* (1996/1954): 13

Abstract. Even a sketchy chronology of the interrelations between computer science and the humanities gives evidence of unrealised potential for mutually beneficial interrelation. But although collaborations date back to the early days of computing, one suspects that these have been dependent far more on circumstance, accident and individual good will rather than on a shared disciplinary understanding, which seems continually to elude us. In this talk I argue that the emphasis on problem-solving in computer science and the characteristic aversion to method in the humanities have rendered each side opaque to the other. Recent attempts to bridge the gulf have demonstrated yet again that the humanities remain ill equipped to formulate their styles of reasoning in terms of methodical procedure, hence to speak to computer science comprehensibly. In turn computer science is apt to miss the point of humanities research or to see very little of the challenge. Echoing Tertullian I ask bluntly, what does computing fundamentally have to do with poetry, music or the arts? Surveying work in the digital humanities, I outline three kinds and attempt an answer for each. In conclusion I consider how the projects of CATCH manifest the potential we have been so poor at articulating.

2

1. Chronology and history

My subject is the now much discussed question of how computer science and the humanities might be most productively engaged with each other. In this talk I will approach the question both historically and philosophically. As Bertrand Russell remarked with reference to history, both disciplines are "an essential part of the furniture of an educated mind" (1958/1956: 191). So, imitating him in spirit, I try to speak historically and philosophically, though not as an historian or a philosopher, neither of which I am.

The origins of the engagement between computer science and the humanities are elusive. Indeed, we can make the starting point as old as we please by allowing the word "compute" to have a generic meaning, and so take "computer" to signify anything that calculates, including people; by giving word "science" its predominately European sense of "acquaintance with or mastery of any department of learning"; and by playing similarly fast-and-loose with the term "humanities". Alternatively, we can make it considerably younger by dating computer science from the first usage of the name, in 1961 (*OED*), or wait until 1965, when the first department was founded at Carnegie-Mellon University in the US. We can tighten the chronological screws a bit further by restricting scope to the first explicit discussions, probably sometime in the early to mid 1960s, when people of mixed disciplinary origins seem first to have started addressing the possibilities of a relationship.[1] Finally, we can be *very* strict, and by laying down the condition that we intend a relationship of mutual understanding, say with confidence that we are speculating about possibilities far from being realised even today.

We do learn something from such historically naïve efforts: that there is strong argument for a great affinity of concerns which computer science and the humanities share; that this affinity goes back as far as one wishes to seek for it; that practitioners in the disciplines have in fact collaborated in one way or another for the last 40 years; that thoughtful scholars with experience on both sides have been thinking about the relationship for almost as long; and that we haven't really begun to open up the cornucopia of possibilities. The problem is not merely that the chronology of interactions is much more thickly populated with complicating evidence than I have admitted or know to admit. What's really lacking here is the historiography that would transform what seem explanatory facts into matters requiring explanation. Until we have such an historiography, we do not even know what people, events, inventions and publications to include. And until we have a genuine *history* of the interactions between CS and the humanities, we cannot really claim to have a proper subject.

3

One way into a history is to start with those who were around at the time, for example your own beloved Edsger Dijkstra, who was disallowed at his wedding from declaring himself a "programmer" because, he was told, the profession did not exist (1972: 860); or Alan Perlis, who before his

[1] In his forthcoming doctoral dissertation, Edward Vanhoutte argues for the beginnings of the relationship in the Machine Translation project, for which see Wilks 1972; Locke and Booth 1955. Elsewhere he refers to the informal mingling of computer scientists and humanists at the conferences held by IBM in 1964-5, proceedings of the first of which are published in Pearson 1965. See also the remarks in Oakman 1987: 227.

appointment as the first ever chair of a computer science department in the United States was simultaneously a mathematics professor and director of a computing centre (Denning 1990: 604f); or, from my own side of the street, Fr Roberto Busa, Jesuit medieval philologist, who having heard of computers, travelled in 1949 from war-torn northern Italy to the United States and there persuaded Thomas Watson, Sr., head of IBM, to give him free computing (1980: 93f). These are wonderful stories. What they teach us *inter alia* is that the lines of intellectual engagement cannot be cleanly drawn, even if the professional boundaries seem at times very definite indeed. They teach us that, as in the story which runs back from Alan Turing to David Hilbert and further, we must look to the metamorphic continuity of intellectual problems for our history – or as the great Romance philologist Erich Auerbach said, we must follow "the changing aspect… of the permanent" (1984/1944: 12).

The situation with computer science and the humanities, as with computer science itself, is greatly complicated by the near contemporaneity of the raw material. Indeed, the term "recent history" is not so much a kind of history as a hotly debated question: can there be such a thing – a history involving people, events and objects not entirely of the past? (McCarty 2004: 162-4). For computing a good start has nevertheless been made by the Princeton historian Michael Mahoney, who shows how the amalgam which is computer science was formed by the intersecting and interrelating of several originally quite separate agendas (2002). For humanities computing the British practitioner Susan Hockey (1980, 2000, 2004) and the American Robert Oakman (1980) have written useful narrative chronologies. But to my knowledge no one other than myself and the Belgian scholar Edward Vanhoutte have even begun work toward a history of interrelations between computer science and the humanities. Once it is published, Vanhoutte's historical account of humanities computing should put our investigations on a much firmer footing.

The start I wish to make here begins with very recent events, from which I will extract an important clue to the work which needs to be done. Following that clue, I will then examine the vexed relationship of the humanities to method as a way of getting at their relationship to computing, and so to computer science. Finally I will take a brief look at the projects of your own CATCH programme as exemplary of the current state of the art (which is far more interesting than what I hear people saying about that state). There is, however, much work to be done on both sides of the house.

4

2. Very recent attempts

I began to pay attention to the question in 2004, when in the course of writing a book on humanities computing I discovered, as George Steiner said of Martin Heidegger, that there was a massive presence blocking my path with which I simply had to deal (1978: 21). So I read everything I could find on the subject of computer science as a discipline, and wrote a chapter on it for the book (2005: 158-98). What I have to say here is an abridged development from that chapter, so I will not mention that work further.

5

A more reliable guide to majority thinking, or the lack of it, are two conferences that took place last year, one in Canada, the other in the United States. I gave a paper at the former, held in New Brunswick by the Canadian Symposium for Text Analysis.[2] Apart from my own contribution (which, again, I will pass by in silence), thoughts on the relationship took the form of suggestions for projects already with the "capacity to inspire different 'problem statements' that sound normal for each discipline" involved, as one colleague put it. While happy stories are not to be dismissed, indeed celebrated, this one does not give us an answer to the question I am asking but leaves us with a different version of the same question: how to *translate* problem statements across the divide, as another colleague remarked.[3] The analogy of translation is a telling one. In Umberto Eco's sense, to translate a text means to interpret it in two different languages, *involving the culture of each* (2001: 14, 17). How can you do that unless you are or can become a participant-observer of both?

6

The other conference in 2006 was the first *Chicago Colloquium on Digital Humanities and Computer Science* (which met again for the second time about a month ago). We have the benefit of a commentary on this event by Dan Cohen, a young scholar who keeps a blog.[4] In it he quotes a mordant joke made at this conference by Martin Mueller, a distinguished professor of English and Classics (Northwestern, US) who has been deeply involved in the digital humanities for many years. "I will go away from this conference", Mueller remarked, "with the knowledge that intelligence analysts and literary scholars are exactly the same." There was laughter, and then "the core truth of the joke settled in". Commonalities of method are beneficial, in that they allow practitioners to exchange techniques and tools, but from scholars you

---

[2] CaSTA 2006, www.lib.unb.ca/casta2006/.

[3] Joseph Gilbert and Bethany Nowviskie, respectively, by e-mail, 12/10/07, here quoted by permission.

[4] For Cohen's commentary, see www.dancohen.org/2006/11/13/intelligence-analysts-and-humanities-scholars/; for the Chicago Colloquium, dhcs2006.uchicago.edu/program.

7 might expect to hear much more about what makes their practices distinct from the spooks'. You might expect methodological self-awareness – that is, were it not for the prevalent tendency throughout the humanities to keep method tacit, however methodical the research may be, even if (as in some cases) the method can be spelled out. In other words, the Chicago Colloquium does not present a surprising, perverse or even unusual case, but one more serious, widespread and consequential. Why are humanists on the whole so methodologically unconscious?

8 **3. The humanities and method**

Let me give you an atypically intelligent but still characteristic statement of the problem. In his Forward to the essay by Erich Auerbach from which I quoted earlier, Paolo Valesio asks rhetorically, "Do we regard literary criticism as an essentially aseptic set of metalinguistic operations, akin to those practiced in the exact sciences?" His answer is very revealing: "Then we will always be able to find, in the work of any remarkable critic, elements that may legitimately support and comfort this kind of analysis: for *no critic can escape the embrace of Methodology*. But to say that such a gesture is legitimate does not mean that it is adequate" (1984: viii, my emphasis). What makes this statement characteristic of the humanities is the image of an unwelcome embrace. What makes it intelligent is the recognition of the difference between the legitimacy of method and its inadequacy for criticism. I will return to this difference later.

9

10 At the beginning of *Warheit und Methode* (2000/1960) Hans-Georg Gadamer explains the humanities' silence about method as integral to their particular way of apprehending the world, and he contrasts it with that of the sciences. "[T]he specific problem that the [humanities] present to thought", Gadamer remarks, "is that one has not rightly grasped their nature if one measures them by the yardstick of a progressive knowledge of regularity", as in the sciences. Research in the humanities "does not endeavour to grasp the concrete phenomenon as an instance of a universal rule", he goes on to observe. "The individual case does not serve only to confirm a law from which practical predictions can be made. Its ideal is rather to understand the phenomenon itself in its unique and historical concreteness" (2000/1960: 4f). Since finding regularities is not the goal of humanities research but at best a starting point, the instantiation of regularity in investigative procedure, which we call method, is quite secondary. By emphasizing it, many would say, one goes in the wrong direction.

11 We reach much the same conclusion via cognitive psychologist Jerome Bruner's rather different argument (1986). He begins by pointing out that both

the sciences and the humanities generate imaginative hypotheses, but they differ crucially in what they do with them. Scientific research attempts to falsify hypotheses in order to come to a singular abstract law by which phenomena of the physical world may be explained. Research in the humanities demands not that hypothetical constructs are falsifiable but that they are verisimilar to human experience – that they are *humanly possible*. The goal of that research, Bruner argues, is not to zero in on the particular and unique so much as to open up and explore "the alternativeness of human possibility" – to follow the imagination wherever it may lead, to foster what William Blake called "expanding eyes". In common philosophical language, the humanities are concerned with possible worlds, the sciences with "the possible world we happen to live in" (Sparshott 1990: 7).

If method, then, is a investigative form of the law-like behaviour for which the researcher is looking, we might conclude that it suits the sciences down to the ground but the humanities not at all, or only as something to be kept suspiciously at more than arm's length. Looking closely, however, neither turns out to be quite that simply true. On the scientific side, Ever since
12 Thomas Kuhn started the historicization of the sciences with *The Structure of Scientific Revolutions* (1962), Paul Feyerabend began the recognition of scientific method's plurality with *Against Method* (1975) and the biologists a vision of *Science without Laws* (2007), we have begun to see something much closer to the humanities than previously. Nevertheless, the trajectory of scientific research still moves toward singular formulation, and explicit method, though plural, remains an important aspect of scientific practice. On the humanities' side, convergence has been much slower, with much more modest results – because the problems are so much more complex. My purpose here is to question the humanities' new relationship to method as mediated by computing. That it has brought the humanities and the sciences closer together is a very important result, to which I will return briefly later, although it is not my topic on this occasion.

13 **4. Athens and Jerusalem**

Since a very strong analogy connects physical law (expressed mathematically) and computational procedure (stated algorithmically), the foreignness of explicit method to the humanities may be extended to the present case. To
14 echo Tertullian's famous question – *Quid ergo Athenis et Hierosolymis?*, "What therefore does Athens have to do with Jerusalem?" (*De prae. haer.* 7.9) – we can ask, what therefore can computing have to do with the humanities? European culture has, I think, answered Tertullian abundantly. Computing remains in question.

You may well be eager to point out that this question has a one-word answer, namely "Behold!". Where indeed in the arts and humanities is computing *not* to be found, except perhaps in the studies of a very few retired Oxford and Cambridge professors? Recall nevertheless Martin Mueller's joke at the Chicago Symposium, provocation to which can be found throughout the episodic history of interactions between humanists and computer scientists. We may be able to demonstrate with authority that humanists' lack of methodological self-awareness has solid justification, but computing changes the situation profoundly by requiring that the method to be implemented be spelled out. When humanists pick up computational tools, they're rarely if ever picking up instruments embodying their own ways of working. These instruments are far more likely to embody rather different working methods and goals. Furthermore, the historical development of computing itself is increasingly toward empowerment of an end-*maker* of tools and away from mere facilitation of a passive end-*user*. Hence I think myself justified in asking these blunt questions: why are we humanists being so dull-witted to our own ways of knowing and to the demands of computing as to be methodologically indistinguishable from the spooks? why, to echo Jerome McGann, are most of us at the beginning of the 21st Century so much like Henry Adams at the end of the 19th, whose humanistic training "left him unprepared for the dynamo of the twentieth-century"? (2005: 105-6). "For the interesting puzzle in our times", political scientist Langdon Winner has written, "is that we so willingly sleepwalk through the process of reconstituting the conditions of human existence" (1997/1986: 61).

15

Even if we were able to drug curiosity, even if it were possible to relegate computing strictly to the role of a knowledge jukebox for vending information to be applied in traditional ways, there would still be revolutionary trouble for the old and by now cozy two-cultured arrangement. As in the European Age of Exploration, the volume and variety of primary data now available is challenging old taxonomies, and semantically insensitive searching across large collections of secondary literature is bringing scholars into contact with ways of thinking and speaking that their *Doktorväter* would hardly recognize. To use the language of economics, it is impossible (or at least *very* difficult) to alter a component of a tightly interrelated system, such as the academy, without causing system-wide effects. One does not have to go very far to find these.

16

Two illustrative anecdotes.

(1) A classicist friend of mine, who once jokingly exclaimed, "Thank God the Library at Alexandria burned!", now must cope not only with the *Thesaurus Linguae Graecae* (extended to the fall of Constantinople in 1453) but also with

the masses of epigraphic material coming online from sites throughout SE Europe and the Middle East. No longer can he confidently stick to a rather limited canon of edited texts or can he breezily separate literature and inscriptions into conveniently distinct genres.[5]

(2) When I look for secondary literature online, I now make little use of the common techniques for refining a search – which have never worked very well anyhow. Technologically naïve or otherwise deliberately wide searching allows me to see in what disciplines my topic surfaces, to locate prominent writings from each and so to benefit from the often considerable ethnographic variety of ways in which the topic is treated. As linguist Catherine Ball remarked years ago, there is much to be said for poor precision in retrieval: with perfect precision, "we find exactly what we said we are looking for, *and no more*" (1994: 296, my emphasis).

Whether quite so deliberately, whether judiciously handled, the same widening effects of imprecise searching must now be commonplace. It does not take much imagination to see what must be happening to disciplinary boundaries, indeed to our whole conception of disciplinarity, as a result of relativizing encounters with variant discourses. This was, of course, going on before the Web. Back in 1980 American anthropologist Clifford Geertz took time out to notice "the enormous amount of genre mixing in intellectual life", interpreting it as a sign of "the refiguration of social thought" (1980: 19). The lesson to learn here is that the Web didn't *cause* genre-mixing, rather it gave very long leash to an already strong desire to expand the scholarly mind into the wide-open fields of the academic *heteroglossia*.

In any case, we cannot protect our disciplines against curiosity, at least not all of the time, nor our research against the intimations of a much greater promise. If we are to do any better than the Chicago Colloquium, as I think we must, we have to confront the question of what computing's Athens has to do with poetry's Jerusalem on the most fundamental level of the research that humanists do. It's my principle aim here to suggest what kinds of responses are underway or at least possible.

There are, I think, three: the collective, the analytic and the synthetic.

17

5. The collective

18

By the "collective" kind I refer to the massive effort now going on to stock the virtual shelves. This effort is analogous to the generations of hard work that

[5] For the breakdown of these genres, see Roueché 2008.

went into building up the network of printed resources which have allowed philological, literary and historical criticism to flourish. But so great is the felt need to make resources accessible that we have paid scant attention to the kind of questions that scholars in many fields have traditionally asked. The problem is that asking some of these questions is crucial to resource-building of the kind we need.

Among computer scientists, massive digitization throws up the well known technical challenges of retrieval. But in the rush back to the lab, the questions that I referred to tend not to get asked. As an example, I cite the fascinating set of problems in "music information retrieval", and in particular an interesting paper by Donald Byrd and my London colleague Tim Crawford, "Problems of music information retrieval in the real world" (2002). This paper rightly notes the early interest in the field that was submerged in the great flood of enthusiasm for the Web, which as elsewhere obscured the fundamental questions to which Byrd and Crawford turn (250). When the authors ask, "Why is music IR hard?", they reach for the analogy of words. This is not a bad place to begin, but they begin by assuming as fact that "in text, there are many ways to say the same thing", hence that "a basic requirement of text IR is conflating units of meaning". Then, since "it is not clear that music *has* units of meaning" comparable to words, they conclude that it is a "vastly more difficult problem" than text even in the most challenging of languages (260).

Perhaps it is. But in treating synonymy as a convenient fact of language rather than a hugely problematic notion which works only by interpretative fiat, they chart a path that takes all travellers who stay on it past the point of no return. Language simply isn't like that in the *real* world. Concepts are *not* data, or in the data, or anywhere else but in our heads; they are abstractions we create in order to fit a text to a particular and often limited way of thinking. When we assert that we are *saying the same thing in different words*, we are overlooking a highly complex set of mental operations that, if done knowingly, are done conjecturally, *as if* two different things were the same thing. We may express the result in metadata for convenience of retrieval, or achieve it by means of a clever algorithm, but in either case only the most intellectually trivial problems of linguistic processing are solved – those of little or no interest to scholars.

Retrieval is very important, to be sure, but it is problematic in proportion to expressive qualities of the text to be retrieved. If all one wants to do is to handle what Umberto Eco calls "closed texts" (such as technical manuals) as they are intended to be used – for *information* rather than for *meaning* – then retrieval in the classical sense will probably do. But if one wants real-world problems of the sort that the humanities live on, involving "open texts",

whose meaning is co-created by the reader, then a different path is required (Eco 1984/1979: 3-43).

Another activity allied with digitization is design of digital genres. For years now it has been clear that old genres (such as editions, dictionaries, commentaries and encyclopedias) are formally dependent on the medium for which they were designed, i.e. print. In attempting to translate the *Oxford English Dictionary* into digital form, computer scientists at the New OED Project at Waterloo, working together with lexicographers and dictionary users, discovered that their perspective on the OED raised a host of fascinating, difficult and consequential problems well outside the domains of computer science (Raymond and Tompa 1988). When designing the digital version of the *Greek-English Lexicon* of Liddell, Scott and Jones – the so-called "Big Liddle" – classicists and computer scientists in the Perseus Project discovered, but only partially implemented, a new sort of dictionary that is more an explicit component of a larger process than a work in itself.[6] Textual editors and commentary makers have been struggling for years to find a digital genre-design that works well enough to be widely accepted. In other words, the question of digital genres likewise takes us to problems far closer to scholarship than to retrieval.

19

6. The analytic

20

By "analytic computing" I mean the kind that matches the sort of analytic operations which scholars perform when, for example, they study texts for how these texts say what they say.

21

The analytic response begins by noticing that to render anything at all computable, a model of it must be built, and that the components of this model must satisfy two rigorous criteria: complete explicitness and absolute consistency. A representation is always different from the reality that it represents, but a computational model is radical in that respect. To get some idea of how radical the difference, attempt to imagine the rendition of a poem that makes its meaning completely explicit and expression of that meaning absolutely consistent from one instance to the next. That's the down-side. The up-side in this characteristic engineering trade-off is the manipulative power that computing brings. This power derives from the fact that to use a musical analogy, "there is no melody, there is [only] melodying", as David Sudnow wrote about jazz improvisation (1978: 146; cf. 2001: 126). In other words, what matters in this process is not the model but its perfective iteration at the hands of the modeller. What matters is not the model but the *modelling*.

[6] See e.g. Crane 1998; the online publications listed at www.perseus.tufts.edu/Articles/.

A brief note about terminology. "Model" is a highly polysemous term, to which here I am giving the sense of a temporary heuristic device that the modeller manipulates in order to find out more about the object of study. In this sense modelling is "perfective", i.e. it moves toward an increasingly better representation until the so-called law of diminishing returns kicks in. When that happens, the modeller is left with something that works fairly well but which (computational representation being as it is) leaves a certain residue of instances that won't fit the overall scheme. There is a gap between the representation and the reality. We know from the example of digital audio production, e.g. on music CDs, that this gap can in at least some instances be reduced to such a fine degree of granularity as to become unnoticeable – in the case of audio signals, below the level at which the human ear functions. We know from participatory virtual reality environments, such as *Second Life*, that the gap can simply be ignored. (One suspects that its near-total disappearance is only a matter of time.) The whole point of analytical digital scholarship, however, is not to make the gap between representation and reality unnoticeable or to overlook it but to track it to the point at which perfecting and iterating the model fails to make any significant improvements, then to inspect the residue. The job is well done when that residue turns out to be, in McGann's words, "the hem of a quantum garment" (2004: 201)– something that transforms the critical world in which one is operating.

22a

22b

22c

That's the analytic response. As Johanna Drucker has pointed out (2007), it is fundamentally reductive. It works by *mathesis*, the Foucauldian "science or practice of establishing a systematic order of things" (*OED*), then by comparison with reality as we know it, in order to raise the epistemological question of *how we know what we somehow know*. It foregrounds method.

23

7. The synthetic

24

The synthetic response to the question of computing's Athens versus poetry's Jerusalem is more difficult to talk about – possibly in part because the reductive language essentially derived from physics has dominated our intellectual culture for a long time. By my highly tentative reckoning, there are three emergent paths along which to develop the synthetic response.

The first, championed for example by the Canadian scholars Andrew Mactavish and Geoffrey Rockwell (2006), is toward the arts. The visual arts are engaged because of rapid development of tools for visualisation and construction of virtual reality spaces, with far greater importance these put on aesthetics and design than been true for computing previously. The

performing arts are engaged through developments in local hardware and communications technologies, which allow computing to involve multiple actors performing in real-time and virtual space – again, in *Second Life* or similar environments. The plastic arts and the crafts are engaged by the essential role of interface objects, both physical and virtual, which involve kinaesthesis or, cognitive scientists now teach us, kinaesthetic memories. The media arts, where the others combine, make kin of a long lineage of media artists and explorers, as the German scholar Siegfried Zielinski has shown (2006/2002). Engineering, with its characteristic, creative *agon* of "design under constraint" (Wulf 2000), becomes a close relation. Quite suddenly we need to hear from artists and craftsmen, performers, engineers, instrument makers and, from before our time, natural historians, as well as from palaeographers, book historians and other scholars of material culture. They all have something to teach us.

The second path turns to studies in language (sociolinguistics and pragmatics, poetics and literary theory) and to improvisational musicology. The chief strains of linguistics relevant to this path are the conversational and anthropological, with their focus on how discourses are dynamically generated and shaped (Sacks 1995, Duranti 1997, Schlegoff 2006). The poetics and literary theory of most interest comes out of Mikhail Bakhtin's notion of texts in "a world of other's words" (1986/1970-1: 143), reader-response and performative theories of language, semiotics (Eco 1984/1979), poetics (Hejinian 2000) and literary-criticism (McGann 2003). The bridge to musicology is provided both by jazz musicians, who themselves speak in terms of conversation (Berliner 1994), and by those whose interests span both (Duranti and Burrell 2004; Sudnow 1978, 2001). Broadly speaking, the diverse concerns of this path converge on the question of how in language or in music new arises dynamically, interactionally from old – for example, how we continue a conversation that remains recognizably the same conversation, or how text and reader exist in a "textual field" defined by a co-dependent relation that somehow the text has anticipated (McGann 2003). The question for computing is how such textual phenomena may be modelled.

The third path draws from biology and related fields, including biological anthropology and evolutionary theory (which have attracted considerable interest from computer science). Biology and its relations stray furthest from the sights and sounds familiar to humanists, but like the musicology of improvisation and the literary theory of co-dependent relations, it also is concerned with how the new emerges from the old dynamically and interactionally in the absence of teleological determinism. Instead of asking, how does the jazz musician know what next to play, or how do new readings of a text come about, the biologists ask, how is it that life emerges out of the

non-living? How is it that living systems continue to bring order out of
25 disorder, contrary to the otherwise universal rule of entropy? In a series of lectures given in Dublin in 1943, and later published as *What is Life?*, the theoretical physicist Erwin Schrödinger responded to the failure of physics to deal with living systems. He gestured prophetically toward what he called a "new physics".

Biologist Robert Rosen argues that we have only recently begun to explore the path Schrödinger glimpsed (2000). Three overlapping ideas are now variously used to characterize it, especially with reference to living systems. The first is *self-organization*, or the ability of a system to increase in complexity without external assistance. The second is *emergence*, which focuses on such a system's ability to generate new properties (Deacon 2006). The third and both most
26 comprehensive and challenging is *autopoiesis* (lit. "self-making"), in which the components of a system continuously both re-generate the processes that produced them and constitute the very system that realises them as a network. Systems which exhibit such behaviour are said to be "complex" – meaning not intricate or complicated but in a dynamic relation of co-dependency with their components. In logic they are said to be "impredicative", or indefinable except in terms of a totality of which they are a part (Rosen 2000: 82-95). Bertrand Russell, writing as a mathematical logician, invoked the "vicious circle principle" to condemn all such self-referential formulations (1908: 237). We may simply find them paradoxical. But as the Chilean biologists who developed the theory of autopoiesis say, we need new words in order to talk about such matters lest we fall "into the always gaping trap of not saying anything new because the language does not permit it" (Maturana and Varella 1980/1972: xvii).

"What remains indubitable", Russell remarked in his memoirs, "is that the new is never as mellow as the old, and that therefore the worship of mellowness is incompatible with new excellence" (1958/1956: 93). In other words, we're in for a very rough ride – and more. Our difficulty in struggling to generate, understand and adapt such new ideas gives the lie to the tiresomely constant, bland and naively optimistic talk of innovation. The ancients feared *res novae* for a reason. We welcome the new but should be under no illusion that it comes easily or undemandingly.

It may be too early to say very much about the probable yield of "new excellence" from my three emergent paths toward a theory for synthetic work in the digital humanities. But two things seem clear to me from my own experience with literary computing.

First is that current tools for textual studies are woefully inadequate. The chief problem I have encountered is their assumption, inherited from the school of New Criticism but profoundly resonant with reductionist thinking generally (Wittig 1978), that context can defined, and so digitized, or ignored altogether. Hence context is either arbitrarily limited or enters through the cognitive back-door, in the form of whatever scholar-users happen to know – but often do not know that they know, and so do not know critically.

Second is that highly promising work is proceeding nevertheless, and attempts, including my own, are being made to put together a discourse for talking about the problem. Already McGann has drawn from the second and third of these paths to explain his implementation in the *Ivanhoe Game* of an authorial "inner standing point" for literary criticism (2003). My theory of modelling for the digital humanities, though focused on the analytic side of the street, draws heavily from the first path and so is directly applicable to synthetic work. But we need much more. Robert Rosen's call to turn from the "mind-set of reductionism, of looking only downward toward subsystems, and never upward and outward" (2000: 2), though specifically addressed to biologists, seems to me an imperative for all the digital humanities.

In obeying that imperative humanists simply must reach out for help wherever it may be found. To whom they reach should be of considerable interest to computer scientists, since at minimum these reachings point either to areas of work already of interest or to new ways of thinking. Such reachings are easy to ridicule, as bad science, or trendy window-dressing, or whatever, but the need to grow, the impulse to assimilate and the desire to connect with the rest of the world are genuine. I don't deny that early efforts will in all likelihood seem rather foolish. Making them will require extraordinary bravery on the part of humanists. But I think as well that humanists can give strong assurance, from the richness and power of their subject matter and maturity of their questioning, that they will give back at least as good as they get, if not better.

27

8. CATCH as catch can

On the whole I was not fair to Byrd and Crawford, who demonstrate beyond doubt the worthiness of their perspective to throw new light on musicology. But I think that they and all others whose work comes under the unfortunate rubric of "information technology" labour under an unnecessary cognitive burden. Were it in my power, I would rid the world of the confusing word

"information", vexatious to those whose quest is for knowledge.[7] In a sense "music information retrieval" is accurate, since only information can be retrieved, and when it is, it does good service by raising the scholar's question, which is also the poet's: "Where is the knowledge lost in information?" (Eliot, *Choruses from the Rock* I.16). The danger is, however, that mere retrieval of information will be taken as the goal, rather than as the beginning of musicological questioning.

Whatever may be its technical accomplishments, the WITCHCRAFT project demonstrates the orientation to such questioning as arises from the involvement of computer science with musicology. I note these strenuous demands: a focus on practices and performances rather than on objects, and so an emphasis on processes to be modelled; a probing for we might call the "essential features" of a song through its many versions, which we know as versions but don't know how we know; an imperative to define similarity in perceptual and cognitive terms – which I would hope means in terms of perceptual and cognitive processes rather than states; a highly critical stance to existing systems; and an insistence on real-world complexity rather than satisfaction within a toy world.

Similarly the CHORAL project, by taking on topic-detection in spoken-word collections, would seem to run straight into the challenging problems of anthropological linguistics and so may well be close kin to my own research into how literature co-creates its own literary tradition. MUNCH considers, I suppose, what one might by analogy call "topic-detection" in visual media, which is at least as subtle a problem as the musicologists and literary critics face. I wonder, would it have in its sight such problems as this one, posed by a Roman mosaic from Carthage? [IMAGE]. MuSEUM, I would suppose, seeks somehow to do what humans do when confronted with variety. One way of putting this would be to echo what Auerbach almost said, that we triangulate on "the permanent" by following it through its "changing aspects". But the lesson of *heteroglossia* in literary studies and elsewhere makes this an even harder problem by teaching us somewhat differently: that this "permanent" is not a pre-existing object that humans find but a snapshot of their imaginative world-building. If so, then what MuSEUM and several if not all of the other CATCH projects aim at is the computational modelling of human creative processes rather than the rather old-fashioned retrieval task of mining data

28 - 32

[7] "Information" is a verbal amalgam which in current usage confusingly mixes the technical sense defined by Claude Shannon and Warren Weaver for their mathematical theory of communication with the older sense of communicated factual knowledge, with the result that the word now often implies "the notion of meaning as a measurable abstract quantity without respect to the circumstances and format of its production or reception" (McCarty 2005: 109-12).

from what humans have created in the past. What these projects have to offer *inter alia* is assistance in that imaginative world-building.

I've been talking about the interrelation between how we think and what we do, giving priority to the former because words are my medium of work. But computing is all about making things that act on the world rather more directly than words do – acting also on the world of ideas, including those of the humanities. In setting out to make things, computer scientists have to ask *dumb questions*, that is, questions which have been kept voiceless because the keepers of the domain of those questions have not thought to ask them. Giving these questions a voice – which is to say, *problematizing* that which provokes them – is a core function of the humanities. So in asking these *dumb questions* computer scientists enter into the game of the humanities on the common ground of humanities computing.

The history of computing which I hope for is one that will show the intimate intermingling of human an technical matters that close to our own time becomes the intermingling of computer science with the humanities. Because institutional forms are notoriously local, I won't hazard an opinion on how others might best manage this intermingling. It is clear that more computer science has to be involved than has been, for reasons Manfred Thaller, who names the activity "*Historisch- Kulturwissenschaftliche Informationsverarbeitung*". It is clear to me at least that the common ground of intermingling is a scholarly place of its own that stands outside the traditional disciplines, in relation to them somewhat like a merchant ship of exploration to the island cultures of an archipelago. It is clear, finally, that the folks involved have to change, the computer scientists to lose their triumphalism (though not their drive toward various unreachable goals), the humanists to regain their taste for the technical arts (though not their relentless questioning of the fruits of these arts).

Works cited

Auerbach, Erich. 1984. *Scenes from the Drama of European Literature*. History and Theory of Literature, Vol. 9. Manchester: Manchester University Press.

Bahktin, Mikhail. 1986/1970-1. "From Notes Made in 1970-1". In *Speech Genres and Other Late Essays*. Ed. Caryl Emerson and Michael Holquist. 132-58. Austin TX: University of Texas Press.

Ball, C. N. 1994. "Automated Text Analysis: Cautionary Tales." *Literary and Linguistic Computing* 9.4: 295-302.

Berliner, Paul F. 1994. *Thinking in Jazz: The Infinite Art of Improvisation*. Chicago Studies in Ethnomusicology. Chicago: University of Chicago Press.

Bruner, Jerome. 1986. "Possible Castles". In *Actual Minds, Possible Worlds.* 44-54. Cambridge MA: Harvard University Press.

Busa, R. 1980. "The Annals of Humanities Computing: The Index Thomisticus". *Computers and the Humanities* 14: 83-90.

Byrd, Donald and Tim Crawford. 2002. "Problems of music information retrieval in the real world". *Information Processing and Management* 38: 249-72.

Crane, Gregory. 1998. "New Technologies for Reading: The Lexicon and the Digital Library". *Classical World* 91.6: 471-501.

Deacon, Terrence W. 2006. "Emergence: the Hole at the Wheel's Hub". In *The Re-emergence of Emergence: The Emergentist Hypothesis from Science to Religion*. Ed. Philip Clayton and Paul Davies. Oxford: Oxford University Press.

Denning, Peter. 1990. "Alan J. Perlis 1922-1990:A Founding Father of Computer Science as a Separate Discipline". *Communications of the ACM* 33.5 (May): 604-5.

Dijkstra, Edsger. 1972. "The Humble Programmer". 1972 ACM Turing Award Lecture. *Communications of the ACM* 15.10 (October): 859-66.

Drucker, Johanna. 2007. "Philosophy and Digital Humanities: A review of Willard McCarty, *Humanities Computing* (London and NY: Palgrave, 2005)". *Digital Humanities Quarterly* 1.1 (Spring). www.digitalhumanities.org/export/sites/default/vol/001/1/000001.html.

Duranti, Alessandro. 1997. *Linguistic Anthropology*. Cambridge: Cambridge University Press.

--- and Kenny Burrell. "Jazz Improvisation: A Search for Hidden Harmonies and a Unique Self". *Ricerche di Psicologia* 3.27: 71-101.

Eco, Umberto. 1984/1979. *The Role of the Reader: Explorations in the Semiotics of Texts*. Bloomington IN: University of Indiana Press.

---. 2001. *Experiences in Translation.* Trans. Alastair McEwen. Toronto Italian Studies. Toronto: University of Toronto Press.

Gadamer, Hans-Georg. 2000/1960. *Truth and Method.* 2nd edn. Trans. Joel Weinsheimer and Donald G. Marshall. New York: Continuum.

Geertz, Clifford. 1980. "Blurred Genres: The Refiguration of Social Thought". *The American Scholar* 29.2 (Spring). Rpt in *Local Knowledge*. 19-35. New York: Basic Books, 2000/1983.

Hejinian, Lyn. 2000. "The Rejection of Closure". In *The Language of Inquiry*. 40-58. Berkeley: University of California Press.

Hockey, Susan. 1980. *A Guide to Computer Applications in the Humanities*, Baltimore: The Johns Hopkins University Press.

---. 2000. *Electronic Texts in the Humanities*. Oxford: Oxford University Press.

---. 2004. "The History of Humanities Computing". In Schreibman, Siemens and Unsworth 2004: 31-9.

Locke, A. D. and W. N. Booth, eds. 1955. *Machine translation of languages*. Cambridge MA: MIT Press, 1955.

Mactavish, Andrew and Geoffrey Rockwell. 2006. "Multimedia Education in the Arts and Humanities". In *Mind Technologies: Humanities Computing and the Canadian Academic*

Community. Ed. Raymond Siemens and David Moorman. 225-43. Calgary: University of Calgary Press.

Mahoney, Michael S. 2002. "Software As Science -- Science As Software." In *History of Computing: Software Issues. International Conference on the History of Computing ICHC 2000, April 5-7, 2000, Heinz Nixdorf MuseumsForum, Paderborn, Germany*. Ed. Ulf Hashagen, Reinhard Keil-Slawik and Arthur L. Norberg. 25-48. Berlin: Springer Verlag. www.princeton.edu/~mike/.

Maturana, Humbert R. and Francisco J. Varella. 1980/1972. *Autopoieses and Cognition: The Realization of the Living*. Dordrecht: D. Reidel

McCarty, Willard. 2004. "As It Almost Was: Historiography of Recent Things". *Literary and Linguistic Computing* 19.2: 161-80.

---. 2005. *Humanities Computing*. Basingstoke: Palgrave.

---. 2008. Rev. of Zielinski 2006. *Literary and Linguistic Computing*.

McGann, Jerome. 2003. "Texts in N-Dimensions and Interpretation in a New Key". Special issue on The Ivanhoe Game. *Text Technology* 12.2 (2003): 1-18. texttechnology.mcmaster.ca/pdf/vol12_2_02.pdf (12/10/07).

---. 2004. "Marking Texts of Many Dimensions." In Schreibman, Siemens and Unsworth 2004: 198-217.

---. 2005. "Information Technology and the Troubled Humanities". *Text Technology* 2: 105-21. texttechnology.mcmaster.ca/pdf/vol14_2/mcgann14-2.pdf.

Oakman, Robert. 1980. *Computer Methods for Literary Research*. Columbia: Univ. of South Carolina Press.

---. 1987. "Perspectives on Teaching Computing in the Humanities". *Computers and the Humanities* 21: 227-33.

Pierson, G.W., ed. 1965. *Computers for the Humanities?* A Record of the Conference Sponsored by Yale University on a Grant from IBM January 22-23, 1965. New Haven, CN: Yale University Press.

Raymond, Darrell and Frank Wm. Tompa. 1988. "Hypertext and the Oxford English Dictionary". *Communications of the ACM* 31.7 (July): 871-9.

Rosen, Robert. 2000. *Essays on Life Itself*. Complexity in Ecological Systems Series. New York: Columbia University Press.

Roueché, Charlotte. 2008 (forthcoming). "Digitising inscribed texts". In *Text editing, print and the digital world*. Ed. Marilyn Deegan and Kathryn Sutherland. Digital Research in the Arts and Humanities series, Methods Network. London: Ashgate.

Russell, Bertrand. "Mathematical Logic as Based on a Theory of Types". *American Journal of Mathematics* 30.3 (July): 222-62.

---. 1958/1956. *Portraits from Memory and other Essays*. London: Readers Union, George Allen & Unwin.

Sacks, Harvey. 1995. *Lectures on Conversation*. Ed. Gail Jefferson. 2 vols. Oxford: Blackwell.

Schegloff, Emanuel A. 2006. *Sequence organization in interaction: A primer in conversation analysis.* Cambridge: Cambridge University Press.

Schrödinger, Erwin. 1943. *What is Life? The Physical Aspect of the Living Cell*. Cambridge: Cambridge University Press.

Schreibman, Susan, Ray Siemens and John Unsworth, eds. 2004. *A Companion to Digital Humanities*. Oxford: Blackwell. www.digitalhumanities.org/companion/, 16. (12/10/07).

Schrödinger, Erwin. 1944. *What is Life? The Physical Aspect of the Living Cell*. In *What is Life?* with *Mind and Matter* and *Autobiographical Sketches*. Cambridge: Cambridge University Press, 1967.

---. 1996. *Nature and the Greeks* and *Science and Humanism*. Cambridge: Cambridge University Press.

Sparshott, Francis. "Imagination: The Very Idea". *The Journal of Aesthetics and Art Criticism* 48.1 (Winter): 1-8.

Steiner, George. 1978. *Heidegger*. Fontana Modern Masters. Glasgow: Fontana / Collins.

Sudnow, David. 1978. *Ways of the Hand: The Organization of Improvised Conduct*. Cambridge MA: Harvard University Press.

---. 2001. *Ways of the Hand: A Rewritten Account*. Cambridge MA: MIT Press.

Valesio, Paolo. 1984. "Foreword". In Auerbach 1984: vii-xxviii.

Wilks, Yorick. 1972. *Grammar, Meaning and the Machine Analysis of Language*. London: Routledge and Kegan Paul.

Winner, Langdon. 1997/1986. "Technologies as Forms of Life". In *Technology and Values*. Ed. Laura Westra and Kristin Shrader-Frechette. 55-69. Oxford: Rowman and Littlefield.

Wittig, Susan. 1978. "The Computer and the Concept of Text". *Computers and the Humanities* 11: 21115. [*]

Wulf, William. 2000. "The Nature of Engineering, the Science of the Humanities, and Gödel's Theorem (Interview)". *Ubiquity* 1.28. www.acm.org/ubiquity/interviews/w_wulf_1.html.

Zielinski, Siegfried. 2006/2002. *Deep Time of the Media: Toward an Archaeology of Hearing and Seeing by Technological Means*. Trans. Gloria Custance. Cambridge MA: MIT Press. Rev. McCarty 2008 (forthcoming).