Literary enquiry and experimental method: What has happened? What might?

Willard McCarty
Professor of Humanities Computing
King's College London
staff.cch.kcl.ac.uk/~wmccarty/

Crossing boundaries: History of science and computational linguistics
Università di Bari
28 April 2008

> No amount of refinement or subtlety within the world of mechanism can avail; once we are in that world, what we need is already gone. Thus, we must retreat to an earlier epistemological stage, before the assumptions that characterize mechanisms have been made.
>
> Robert Rosen, *Life Itself: A Comprehensive Inquiry into the Nature, Origin, and Fabrication of Life* (1991): xvi

> Rather than think of the computer as a kind of brain, think of it as a kind of book…. a machine for processing a variety of symbolic forms organized in looping autopoietic structures….
>
> Jerome McGann, "Visible language, Interface, IVANHOE", in *The Scholar's Art: Literary Studies in a Managed World* (2006): 156

1. The problem

[SLIDE] Consider the two passages of my epigraph.

The first, imported from a book on theoretical biology, puts the reductionism of physics as we know it beyond the pale of life. Rosen inverts the usual hierarchy of sciences, by which physics is the most basic and universal, biology somehow (we don't know exactly how) a special case, in order to make room for the question Erwin Schrödinger famously asked in 1943, "What is life?" Rosen argues that mechanism renders life incomprehensible, that in order to investigate life we must think in terms of a modelling relation between the causal processes of nature and the inferential processes of reasoning, not the identity of these two that classical physics assumes. I quote Rosen primarily because his fundamental reorientation of perspective suggests a magnanimous way of considering the relationship between algorithmic mechanism and imaginative language. The culturally ascendent

biological sciences serve us well with a range of powerful analogies for rescuing literary computing from the doldrums where most of it has impotently dozed for the last forty years or so.

My second epigraph, from a book on textual scholarship, represents the most imaginative attempt to date to awaken literary computing to its primary task of participating in the interpretative operations of criticism. It points the path I will be following. [SLIDE] Note in McGann's description of computer-as-book the adjectival form *autopoietic* (lit. "self-producing"),[1] not yet in the *Oxford English Dictionary* but prominent in a semantic field where evolutionary biology overlaps with other fields which study the way less sophisticated systems make themselves into more sophisticated ones.

Let me put the situation of literary computing in a nutshell: by placing the question of method in the path of criticism, computing gives us the chance to be as brave as Schrödinger. Feeling "the keen longing for unified, all-embracing knowledge", he could see no other escape from the dilemma of specialisation "than that some of us should venture to embark on a synthesis of facts and theories, albeit with second-hand and incomplete knowledge of some of them – and at the risk of making fools of ourselves" (1967/1944: 1). Like physics as he saw it, literary computing has been at the end of its conceptual tether for a long time. To loose the knot so that it may run free, knowledge from distant parts must be gathered and put to use. Disciplinary boundaries must be transgressed, gatekeepers outwitted. There's little risk in explaining computing's failure to do much for interpretation. There's great risk in doing something about it.

2. Lessons of history, 1949-1989

[SLIDE] Trouble began early in the history of literary computing. In 1966 the first professional journal in the field, *Computers and the Humanities*, made groundbreaking activity suddenly visible. Early, dramatic successes in computing had stirred popular and scholarly interest in what computers were doing and might be able to do. In 1949, for example, an American lawyer, inspired by cybernetics, thought that only the vaguery of legal terms stood in the way of deciding lawsuits by machine (Loevinger 1949: 471f). [SLIDE] In 1957 Herbert Simon went much further: "Put it bluntly (hard now to shock)", he wrote in his lecture notes, "Machines think! learn! create!"[2] [SLIDE] In a similar vein, British computational linguist Margaret Masterman asserted in her contribution to *Freeing the Mind*, a series published in the *Times Literary*

[1] See below, note 25.

[2] Simon 1957, later incorporated into Simon and Newell 1958a; the predictions, having encountered objections, were repeated with emphasis in Simon and Newell 1958b.

Supplement in 1962, that the computer was far more than the "menial tool" other contributors had described. It was, she declared, a "telescope of the mind" which like its astronomical namesake would soon "enlarge the whole range of what its possessors could see and do" and so change "their whole picture of the world" (38f) – words quoted approvingly four years later in a book on the automated analysis of text (Ellis and Favat 1970/1966: 126). [SLIDE] At the end of that decade, the editor of *The Economist* predicted that in the 1970s the computer would come into its own, perhaps even changing "the outlook of man". [SLIDE] A somewhat less restrained Canadian journalist had proclaimed five years earlier a "new age of miracles" [SLIDE] in terms strikingly similar to a 21st-century funding concept known as "the semantic web".

Battle-lines were being drawn over the significance of computing. Note, however, not the division formed by enthusiastic promoters on one side and grumpy Luddites on the other, but the one between those, like Masterman, who focused on augmenting human capacities,[3] and those like Simon, whose interest in automation is a lineal descendant [SLIDE] of Frederick Winslow Taylor's clerkish elevation of efficiency,[4] the factory production-line, a leisure-class of passive consumers and the rest of modernism's progeny.

[SLIDE] Two technical assessments from the mid 1960s told a sobering story, however. In *Alchemy and Artificial Intelligence*, with the clock still ticking on Simon's confident predictions, Hubert Dreyfus diagnosed stagnation in AI research, pointing to unexpected difficulties which had followed the early successes. What had been assumed to be a difference of degree, between current and envisioned systems, had turned out to be a difference in kind with no known way across the discontinuities (1965: iii, 9). [SLIDE] The following year the devastating "black book" on machine translation, *Language and Machines: Computers in Translation and Linguistics*, came to more or less the same conclusion (ALPAC 1966: 32). As Fr. Busa has noted, the major obstacle turned out to be not "the inadequacy of computer knowledge but rather… our insufficient comprehension of natural language" (1980: 86; cf Wilks 1972: 4).

Just as now, the majority of computer-friendly scholars in those days held a clerkish, Taylorian view of computing, which put its benefits wholly and cleanly on the side of efficiency, excluding the interpretative operations of criticism. For example, in his 1965 review of the Literary Data Processing Conference held by IBM the previous year, Franklin J. Pegues declared that

[3] For the idea of augmentation see esp Engelbart 1962; Bardini 2000.

[4] Taylor 1913/1911. Simon and Newell cite Taylor as precursor in 1958a.

the emergent purpose of computing was "to free the humanist for the important work of literary criticism by providing him with large and accurate masses of data that may be used by him in the work which only he can accomplish".[5] (The assurance given here runs nervously through the literature of the period.) Pegues argued that the scholar's horizons would be enlarged by more data with fewer errors and less drudgery, but he passed over in silence how those horizons might be affected. At the IBM conference Ephim Fogel, cautiously optimistic, argued for incremental progress along what he called the "Vision-Actuality Interval", which however wearisome to traverse was, he thought, only a matter of time, fortitude and steadily accumulating resources.[6] As with AI and machine translation, however, such progress faltered – not because of practical shortcomings but because of theoretical blinkers. Fogel's interval became a chasm.

[SLIDE] Not everyone was so inattentive, however.[7] A number of writers from 1966 to 1969 identified the need for theory, urging a turn away from the plodding "mechanical clerk" to imaginings of a creative, readerly instrument. Louis Milic's article, the first in the first issue of *Computers and the Humanities*, is an example.[8] [SLIDE] The point was again made forcefully in a remarkable clustering of articles a decade later, in 1978. Susan Wittig, referring to Masterman, pointed out that computing had undoubtedly allowed for improvements, making performance of old tasks more efficient and accurate, but it had not delivered on the vision's promise (1978: 211). She argued that inattention to theory had made literary computing vulnerable to covert influence by a positivistic "concept of text" derived ultimately from New Criticism.[9] A computing effectively without theory, Colin Martindale argued, was no better than a method in search of a paradigm to direct and explain it (1978: 275f). Richard Bailey, quoting both Wittig and Martindale, declared that practitioners were blindly groping their way through criticism's past, with a time-lag of about 50 years (1978: 7). John B. Smith argued that Formalist, Structuralist and phenomenological ideas could form the basis for a "computer criticism" but got no further than suggesting the influence of

[5] Pegues 1965: 107. The Conference proceedings were published as Bassinger, Parrish and Arader 1964.

[6] Fogel 1964; on the state of resources at the time see Rommel 2004: 93.

[7] This part of the story is well told by Potter 1991: 403-7.

[8] See Rosanne Potter's useful survey of early work (1991: 402-4); Milic 1966: 6;

[9] Note Paul de Man's comment that "from a technical point of view, very little has happened in American criticism since the in-novative works of New Criticism. There certainly have been numerous excellent books of criticism since, but in none of them have the techniques of description and interpretation evolved beyond the techniques of close reading established in the thirties and the forties" (1973: 27).

computing on "the questions one asks… and the way one sees the text and its meaning" (1989/1978: 14).

As Busa said about the machine translation project, the problem had little to do with computing, much to do with the operative *idea of text*, whose improvement and expansion is the real aim of the exercise (1980: 89).

Thus from the beginning of humanities computing's first professional journal, the whistle was being blown on naïve literary computing. But given the enormity of the challenge and the ease with which low-hanging fruit could be gathered by means of simple tools, it is not surprising that in 1989, in the year that the text-analysis program TACT was released to the public, Rosanne Potter remarked in the Preface to *Literary Computing and Literary Criticism* that computing had "not been rejected, but rather neglected" by mainstream criticism – a complaint repeated many times since.[10] Numerous reasons for this neglect have been offered, but the fact remains that literary computing has had very little to say in response to critical discourse for the last half-century. Twelve years after Potter, in the Preface to *Radiant Textuality* (2001), Jerome McGann noted the instrumental role of computing in "the technical and pre-critical occupations" on which scholarship depends but its almost total absence from interpretative procedures (2001: xii).

It would be a mistake, however, to say that the fault lies solely with literary computing for its neglect of theory. For the past several decades most theorizing, shading into what Jonathan Culler has called "just plain 'theory'" (1997: 1), has offered few points of contact with the study of textual data, and so few possibilities for coaxing computing practitioners out of their theoretical silence. As Paul de Man remarked in 1973, "from the technical point of view" criticism has had very little to offer beyond "the techniques of close reading established in the thirties and forties" (1973: 27).

In the article I have already cited, Martindale pointed out that both literary theory and literary computing "have strengths and weaknesses, but the

[10] Potter 1989: xvi; see also Potter 1991; Corns 1991; *Computers and the Humanities* 27.5-6, 1993; Opas and Rommel 1995; *Literary and Linguistic Computing* 18.2, 2003; Rommel 2004; Hoover 2007; McCarty 2008 (forthcoming). Note that despite the intention, declared two years ago at the Summit on Digital Tools in the Humanities at Virginia, to "enable new and innovative approaches to humanistic scholarship" (Frischer et al. 2006), discussions of tool-building have likewise been preoccupied with features of software, continuing to allow old ideas their shadowy power, ignoring the discourse of criticism as it has moved on. The Chicago Digital Humanities/Computer Science Colloquium, now in its third year, has not done much better; see lucian.uchicago.edu/blogs/dhcs2008/ for the Colloquium; note esp. Dan Cohen's blog for Monday, 13 November 2006, at www.dancohen.org/category/computer-science/ (12 April 2008).

striking thing is that the strengths of one are the weaknesses of the other. If the two were meshed," he suggested, "the weaknesses would largely be cancelled out." But rather than mesh, empirical and theoretical approaches have been taken up each in turn, each taken as the answer rather than as the answer's other half. The result, Leonard Forster noted in 1978, is dogmatic abstraction and a criticism that is little better than "a flight from literature".[11] He recommended a "flexible pragmatism" analogous to the craftsman's, who selects now this tool, now that one to accomplish whatever task is at hand. Forster's metaphor is a good one, not because it privileges the scholarly task over the tool (the timid mantra of service-orientated computing) but because it places tools in the context of an *active interface* between craftsman and material.

The moral of the story is that neither task nor tool holds the secret. What's needed is attention to the craftsmanship, to the process and practice of that which criticism entails.

3. The matchmaker's tasks

[SLIDE] Literary computing can doubtless continue as an "invaluable assistant" to scholarship (Hockey 2000: 84), following criticism wherever it goes and trying its best to be of service. But scattered evidence from literary computing, the compelling nature of the questions its failure raises and the obvious benefits of hooking up with the discourse of criticism suggest that brokering Martindale's Leibnizian marriage of theorist and empiric is worth the candle. However ghettoized it may be in specialist periodicals (Corns 1991), literary computing remains an underdeveloped rather than moribund research programme. The question is, what now can be done?

[SLIDE] In *Humanities Computing* (2005) I took up part of the task by concentrating on the theoretical implications of computing as an analytical approach to the study of the humanities as a whole. I presented a negative epistemology, arguing that the primary function of computing is not to automate knowing but to identify what we somehow know but cannot adequately specify. Because computing gives us manipulatory power over the models we construct, we are able rapidly to close on that which cannot be formulated. Thus we are confronted with our own quite specific ignorance of cultural artefacts and so are better equipped to question them. For literary studies this epistemology takes computing significantly beyond the standard view of an efficient but essentially mute and obedient handmaiden by challenging us in detail to account for the failure of any rigorously analytical

[11] Forster 1978; cf. de Man 1973: 27f.

try for a systematic order of things. But it takes us no further than the negative gift with positive consequences that lie somehow beyond what any such try can in principle do.

The fundamental problem raised by all observational instruments is, as Ian Hacking has argued for microscopy,[12] that we "don't just peer" through them to newly visible objects that are as we see them to be, independently of the viewing. We must also "interfere" with the incoming data based on what we know of what we are trying to observe. We must *make sense with* these data, sometimes by intervening in the observational process, sometimes by altering the object of study. This we simply cannot do, or do well, without a good idea of what we are looking at. In literary studies such knowing interference is not, as in the sciences, so much a preliminary step toward consensus about the object in view. Rather it is an ongoing, never-ending process. The literary object in view is hardly an object at all but the contingent, interactive, emergent outcome we wisely use a gerund to name: *reading*.

[SLIDE] For centuries the codex book has functioned as just such an instrument, or "machine to think with", as I. A. Richards called it in 1926.[13] For centuries it has encouraged interpretative interference with the flow of language, even (e.g. in critical editions and commentaries) providing optional sequences of interfering moves. To us, apt to view it from the outside, the codex appears as a metatheoretical statement articulated by such moves (cf McGann 2001: 75-97). Computing foregrounds these, but because "computing" is also a gerund – not a name for an action or set of actions but a name for *acting* – its kind of book is less a metatheoretical statement or even map of possibilities than a device for trying things out. But computing's shift of emphasis from representation to intervening, while clearly experimental in character, also drives constant theorizing in response to what has just happened, and in preparation for the next move. For this reason the tools we use must be theoretically capable to the highest degree at the most basic level of their design.

If computing is to be pursued as a way of critical acting on literature then much more than the epistemological question is at stake. To be brought to ask how we know what we somehow know but cannot represent computationally is a major step forward, [SLIDE] but it is preliminary to asking the ontological question Wittig and others raised at the end of the 1970s and

[12] Hacking 1983: 186-209. For the computer as microscope, see e.g. ALPAC 1966: 121, Gilmour-Bryson 1984: 11; for microscope and telescope, Denning 2007; more generally, Mahoney 2000: 31. Frege, echoing Leibniz, used the microscope as a metaphor for his *Begriffschrift* (notation of concepts), for which see Göranzon 1993: 44 and Crane 2003: 24.
[13] Richards 1926: 1; cf. McGann 2001: 54-7.

McGann again two decades later: *what is text that it eludes all such representation* – that it can be, in McGann's words, "the hem of a quantum garment" (2004: 201)? Analytical literary computing tells us how to exploit the unavoidable difference between textual representation and reality, but it has nothing at all to say about what we choose to represent. Even if we limit the textual object of study to its verbal data, trouble starts with the context required for interpretation. The dominant consensus within a critical specialism may obscure the problem and often does. But we are warned of it by the crippling difficulties of infinite regress that the very idea of context appears to cause whenever anyone asks what exactly it is a promissory note for (Scharfstein 1989). Context, Jonathan Culler remarked, is merely more text and so appeal to it solves nothing (1988: 93f). But appealing to it, particularly if it is to be modelled computationally across the open domain of literature (or of real life) reveals how unsatisfactorily arbitrary and limiting the unspoken notion or any analytic formulation of it is. The problem of context is the problem of text. What is it?

By failing to ask the Wittig-McGann question, literary computing is confined to providing evidence for or against what we already know or suspect. It is strongly inhibited in its capacity to surprise. Providing evidence seems justification enough, but evidence becomes increasingly problematic as the volume of data exceeds the norm for critical practices formed prior to the exponential growth of online resources. As this volume increases, so does the probability of arbitrary choice, and so the ease with which any statement may be connected to any other. Good critics may do better scholarship by finding more of what they need; bad critics may be swiftly becoming worse ones more easily. The point, however, is that literary computing has thereby served only as mutely obedient handmaiden, and so done nothing much to rescue itself from its position of weakness, from which it can hardly deliver the benefits claimed for it by the faithful. It has done little to educate scholars methodologically.

There is, of course, no single answer to the Wittig-McGann question, because there are many kinds of text, many ideas of what to do with each kind and every reason to think that these kinds and ideas are limited only by human ingenuity. There is only the questioning. Given the renewed prominence that McGann's work has brought to the question, what can be done is to develop ways of asking it such that responses can be made *in* as well as *with* software.

An obvious starting point is with inherited tools of reference, e.g. lexicons, critical editions and commentaries, inferring from them the ideas of text they implement. To the degree this has been done, in aid of speculating about or designing a software equivalent, results suggest the prominent role of tacit

uses in the social contexts of argument and in the building or maintaining of a social imaginary.[14] I will return to the importance of this later. Results also illumine the primitive crudity of our software tools. Let us take a brief look at them to see what they can teach us.

4. What the tools say

[SLIDE] We know from experience that none of these tools do very well with the Wittig-McGann question, but to do anything useful at all, they must afford a view of text that can be recovered.

Initially the answers we get back from existing tools are impoverished. [SLIDE] Nowadays a concordancer, for example, implies that by "text" we mean a corpus informed by verbal correspondence of passages and by the words that collocate with whatever word is in focus. [SLIDE] Both relational database design and formal ontologies imply an instantiated set of concepts and their interrelations, and prior to these, well-defined perspectives of enquiry. [SLIDE] An annotation tool affords a view of text as the occasion for commentary. [SLIDE] A statistical analyzer yields a complex population of verbal clues to a literary style. These are all valid, even highly valuable aspects of text, but again, they are isolated and so isolating.

We can, however, greatly enrich what each has to contribute by considering their historical origins. [SLIDE] The most obvious to be explored is the concordancer, a direct descendant of the late 12th or early 13th-century device invented to serve interpretation of the Bible by gathering together verbally correspondent passages (McCarty 1993). Once the format of the concordance stabilized in the late 13th Century, it remained broadly the same well into the 1970s. Thus an essentially medieval tool bore with it into computing the biblical scheme for triangulating from scattered verbal signifiers to their signified, verbally transcendent meaning, implying an idea of text – Erich Auerbach named it "figural interpretation" (1984/1944) – fundamental to the European tradition from the time of its origins. In the first few decades of computing, implementing concordances fostered this idea of text and drew specific attention to the mechanical operations of scholarship, provoking theoretical reflection which has yet to be harvested.[15] For practical reasons the so-called keyword-in-context (KWIC) format, devised in the late 1950s in response to automation, began to replace the medieval design. With its adoption, the KWIC layout narrowed attention from concordant *passages of a text* to shared *collocates of a keyword*, and so moved the principal domain of use

[14] For the commentary see McCarty 2004; for the lexicon, Raymond and Tompa 1988.
[15] See e.g. Markman 1965.

from literary studies to corpus linguistics and lexicography. The resultant idea of text was most famously articulated by J. R. Firth, echoing the ancient proverb, *noscitur e sociis*: "You shall know a word by the company it keeps!" (1957: 11).

[SLIDE] So also the tools and techniques of annotation imply a partial answer to the Wittig-McGann question. These have historical roots in ancient commentary practice, including manuscript glosses, marginalia, free-standing notes and other forms of intertextuality, together with their social networks (McCarty 2004). Their ideas of text remain largely unexplored, though attention to glossing and marginalia is on the rise.[16] [SLIDE] Relational database design and textual ontologies are similarly emergent from older practices of categorization and tabular layout beginning with ancient libraries, and more recently from the strong cultural predisposition toward discontinuous plurality. Lev Manovich's argument for the database as symbolic form provides a starting point here (2001: 218-43).

5. Turning to confront the context

[SLIDE] Writing a conceptual history of literary computing from its tools would help to give it a theoretical voice, but at best the exercise yields a semi-coherent and uncertain miscellany. Apart from the problem of verbalizing what Davis Baird has called "thing knowledge" (2004), evidence is lacking from actual choice of which tools, the sequences in which they are applied and the unexpected uses to which they are put (an especially acute problem in the humanities, where few tools have been designed for the purpose). Humanists lack a tradition of experimental work: they are not trained to record what they do as they are doing it and have no genre in which to describe research as process. They are under-educated technologically and so undervalue or simply do not see the mediation tools perform. Unsurprisingly, evidence from the scholarly record, in the rare instances in which it exists at all, is scattered through footnotes and asides in publications across many disciplines.

In any case, there is no whole for these parts to sum to, no great idea of text that may be assembled from the scattered fragments of its implementation. The point of asking the Wittig-McGann question is quite otherwise: to enable literary computing to make a great inductive leap from its mute servant's doldrums to an understanding of itself as a full participant in the interpretative operations of criticism. To devise new tools without the benefit

[16] See McCarty 2004 for commentary; Wieland 1985, Teeuwen 2002 and O'Sullivan 2004 on medieval glossing (for which it is clear that a corpus of glosses is quite different from a "text" in the Lachmannian sense); Jackson 2001 on marginalia.

of at least trying to make that leap has not and will not significantly increase the mildly helpful but severely cobbled abilities of literary computing no matter how much data accumulates.

McGann's own response to the question has been to argue for the reversal of perspective within criticism already implied by the Bakhtinian situating of text in an "immense, boundless world of others' words" (Bakhtin 1986: 143). The details of this response, including his online game *IVANHOE*, are best presented by his own writings, which are here taken as required reading and as a point of departure, to which I will return.[17] But that reversal of perspective is already inescapable given the problem of context, which itself seems inevitable once we free literary computing from the strictures of a knowledge jukebox to become a project for modelling literature.[18]

The term "modelling" is so polysemous that its meaning cannot be taken for granted, so I had better say what I mean by it. [SLIDE] Here I show the modelling relation between a formal or abstract system, such as a computer program, and a natural system or artefact, such as a poem. In *Humanities Computing* I argued for the analytical, mimetic kind that Clifford Geertz has called "modelling-of" (1993/1973: 93), which aims at refinement of the epistemological question, as noted earlier. Geertz distinguished this kind from its opposite, "modelling-for", a more or less creative realisation of an idea or design achieved through perfective, exploratory manipulation. (Design for a new airplane wing is a straightforward example.) [SLIDE] The Bakhtinian reversal, however, entails a different sort of modelling from either of those two, something that resembles modelling-for but begins without a pre-existing design, or at least not a consciously accessible one. It is a mapless modelling "forward", toward something that is not yet anything. Using the musicological term, I call it "improvisational modelling" to denote its moment-by-moment development in performance of an emergent potential. This sort of modelling is widely attested in the experimental sciences.[19] What it might be for text reflects, again, the Wittig-McGann question.

---

[17] See McGann 2006, 2004 and 2001, also www.iath.virginia.edu/~jjm2f/online.html (6 April 2008).

[18] The fundamental role of modelling is itself an inevitable consequence of Mr Turing's universal machine; see McCarty 2005: 170-2.

[19] Gooding, for example, focuses on the products of such modelling rather than the process; he names them *construals*, "flexible, quasi-linguistic messengers between the perceptual and the conceptual" (1986: 208) or "tentative representations of possible outcomes… continually constructed and revised to describe and communicate actual outcomes" (1992: 103).

6. Bridging discourse

[SLIDE] It is a truism that questioning is central to the humanities, and that good research turns good questions into better ones. The Wittig-McGann question is certainly worthy. How can we ask it better, more effectively, in software?

The difficulty that confronts us is bridging the gulf which separates the language of criticism from the language of implementation. This is hard, not because we don't know how to resolve scholarly problems computationally. We have decades of experience with collaborative projects in humanities computing, and ethnographic studies of collaboration, chiefly in the sciences, from which to learn.[20] There common ways of talking about problems and objects that have different meanings for the various participants are well studied.[21] But although this social form of bridging disciplines offers the great benefit of other-mindedness, alone it is an inefficient and only partially effective means of furthering research that is fundamentally the result of two or more intersecting, interacting practices. Collaboration wherever possible needs to be internalized so that the interacting can occur at the speed of thought as well as at the pace of meetings. Hence the need for a mediating discourse.

The time-honoured approach for building a new discourse can be found in the early history of each new way of construing the world: to reach into an older, better established field for promising "tropes and imageries of explanation", as Clifford Geertz said, then to adapt them.[22] In each case a connection is established from the poorly understood phenomenon or system with which one is working to a better understood analogue elsewhere. [SLIDE] The analogy links relationships, not things: as A is to B (within one system) so C is to D (in another). Its strongest claim is that the two systems, however different, are *isotropic*, i.e. the same governing laws or principles apply in both. Hence a good analogy not only holds up to examination and yields many insights, it also pulls the connected fields ever closer together as the imported ways of thinking take hold. Each analogical connection must be probed for its actual benefits as well as cognitive trajectory, but because its yield may not be known for some time, the best anyone may be able to hope for is plausibility at the outset. But note: because an analogy is conjectural, effort is required, in proportion to its strength, not to blur it into an identity,

---

[20] For the digital humanities see McCarty 2005: 116-29; for the sciences see e.g. Galison 1996.
[21] Galison 1997 s.v. "trading zone"; for humanities computing, see McCarty 1995: 121-9.
[22] McCarty 2005: 114-57; Geertz 2000/1980: 22. An early example is Aristotle's adaptation of *hyle* ('forest', 'wood', 'timber') to denote what a Form required to realise itself, for which see Solmes 1961: 396.

and so to render it "cosmological".[23] The danger is especially great when the analogy appealingly simplifies an intractable problem, such as the one under consideration here. Analogizing is as perilous as it is powerful.

In the words of a London improvisational musician, what we are looking for is how to get "from A to C when there is no B" (Bailey 1992: 136). If, that is, we begin, as readers do, with a text, and so with the question of how reading may be modelled, we need to bridge the gulf between Bakhtinian or other outward-directed language on the one hand and on the other a design strategy for a computing system capable of implementing its improvisational trajectory.[24] One promising place to begin is, as I suggested earlier, with evolutionary biology, whose fundamental problem is precisely to answer the improvisational question for living systems. Thus we return to Erwin Schrödinger's lecture at Trinity College Dublin in 1943.

In his commentary on Schrödinger's project, Robert Rosen (with whom I began) has argued that in pursuing his question, Schrödinger diagnosed the fatally constricting path of reductionist methods that had had such great influence on 20th-century thought. "[O]ur universes [of scientific discourse] are limited," Rosen declared, "not by the demands of problems that need to be solved, but by extraneous standards of rigor. The result is a mind-set of reductionism, of looking only downward toward subsystems, and never upward and outward" (2000: 2). What he does not say, but needs here to be said, is that the influence of positivistic discourse on all others has been so great that this "mind-set of reductionism" has been ours as well and is still with us, no matter how discredited among the professional philosophers. Hence the alternative Rosen presents is particularly welcome: a turn toward the quasi-teleological but non-deterministic idea of self-organizing systems, hence the ideas of complexity, emergence and autopoiesis (of which McGann makes extensive use).[25]

Biology and its nearest neighbours (which, after all, still lie at a formidable conceptual distance from criticism) are not the only fields concerned with the emergent properties of autopoietic systems. Other likely candidates include anthropological linguistics and conversation analysis; improvisational

---

[23] Dening 1996: 39f, 53, applied to a discipline one forgets is likewise conjectural; cf Taylor 1985: 1-3.

[24] Much closer to an actual language of design is Eco 1984: 3-43, but Eco sketches states and transitions between them rather than processes.

[25] These terms tend to overlap. For autopoiesis see esp Luisi 2003, which reviews work following from Maturana and Varela 1980/1972; in the context of computing, Winograd and Flores 1986; in application to literary studies, Tabbi 2002 and to language, Livingston 2006. For emergence see Deacon 2006.

musicology, including but not limited to studies of jazz; and the cognitive sciences. These are all promising sources for analogies.

7. An improvisational companion to criticism?

[SLIDE] Perhaps now it is worthwhile returning to Margaret Masterman's "telescope of the mind" to ask what sort of computational instrument might live up to the promise of enlarging "the whole range" of what we might see and do as critics and so change our "whole picture" of literature.

[SLIDE] The most imaginatively powerful attempt to date is *IVANHOE*, an online play-space in which participants intervene, change, add to and comment on the discourse field of a given cultural artefact.[26] The critical objective of the players is to explore in blog-like exchanges the possible worlds or imaginative trajectories of this artefact from an authorial "inner standing-point". Computational tools aid the interpretative play by managing communications and by visualising the interactions of players so as to stimulate their imaginations. Scope of play is constrained to the focal artefact, which players are assumed to know. Googling for whatever is permitted, but the game's tools do not aid or direct the search. *IVANHOE* is thus more closely analogous to a microscope than a telescope, but it is of Masterman's kind nevertheless because it is built explicitly and self-consciously for looking outward from the artefact toward its manifold possibilities.

In the rationale for *IVANHOE*, McGann borrows extensively from theoretical biology and elsewhere, as I have suggested we need to do, but the analogies are rhetorical rather than computational. My research question is this: can we do more? Can we use these analogies to design modelling machines capable of finding connections from a given literary text to others, or can we adapt whatever software may exist, for example to simulate evolutionary or improvisational development? In 1989 Northrop Frye mentioned in passing the possibility that modelling such as I have described might be used to converge on fundamental structures of literature through systematic investigation of its recurring conventional units (1991: 6). Is this a realizable goal?

It's clear from what I have said that although such a modelling machine must be able to search all text in digital form, mere searching is not only insufficient but perilous without some kind of automated guidance. It's clear from the massiveness of the collection to be searched that only the most rudimentary

[26] McGann 2006. Although it is not scholarly in its immediate objectives, the online interactive game *Façade* holds some promise. See www.interactivestory.net/ and Rauch 2006. I am indebted to Matt Kirschenbaum for pointing me to this game.

scholarly metadata, if any at all, can be expected (though metadata generated by search-engines could perhaps be exploited to advantage).[27] It's clear that whatever the instrument does, it must be far more of a cognitively intimate companion than a bot, however semantic the web that gets searched. Searching will need to start from a reading, somehow specified, of a given text, produce results from the textual collection and learn from the reader's response, modifying both future and existing results according to what it learns. Hence, because the envisioned operations are massively combinatorial, they may well require more computing power than is easily available, at least now. They may be supercomputerish. Finally, it's evident that tools of some kind, perhaps like *IVANHOE* offering visual representations, will be needed so that the investigator can direct the machine more effectively and imagine more generously than otherwise.

The question of how to build such a thing is in essence the question of where the permeable, moving membrane is between reader and device, or to put the matter differently, how great a role computing can play in criticism. This is, in effect, the question of artificial intelligence, and so presumably a matter of keen interest for AI research. It is the most intimate, most promising encounter possible between literary theorist and literary empiric. But can it be done?

[27] I'm indebted to Dr John Keating (An Foras Feasa, Maynooth) for the suggestion concerning search-engines.

Works cited

ALPAC (Automatic Language Processing Advisory Committee). 1966. *Language and Machines: Computers in Translation and Linguistics*. Report 1416. Washington DC: National Academy of Sciences, National Research Council. www.nap.edu/catalog.php?record_id=9547 (7 April 2008).

Bailey, Derek. 1993/1992. *Improvisation: Its Nature and Practice in Music*. Cambridge MA: Da Capo Press.

Bailey, Richard W. 1989/1978. "The Future of Computational Stylistics". Paper for the Fifth International Meeting, Association for Literary and Linguistic Computing, King's College London, 15 December 1978. Rpt. in Potter 1989: 3-12.

Baird, Davis. 2004. *Thing Knowledge: A Philosophy of Scientific Instruments*. Berkeley: University of California Press.

Bakhtin, M. M. 1986. "From Notes Made in 1970-71". In *Speech Genres and other Late Essays*. Trans. Vern W. McGee. Ed. Caryl Emerson and Michael Holquist. 132-58. Austin TX: University of Texas Press.

Bardini, Thierry. 2000. *Bootstrapping: Douglas Engelbart, Coevolution, and the Origins of Personal Computing*. Writing Science. Stanford CA: Stanford University Press.

Bessinger, Jess B., Stephen M. Parrish and Harry F. Arader, eds. 1964. *Literary Data Processing Conference Proceedings, September 9. 10, 11 – 1964*. Armonk NY: IBM Corporation.

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

Corns, Thomas N. 1991. "Computers in the Humanities: Methods and Applications in the Study of English Literature". *Literary and Linguistic Computing* 6.2: 127-30.

Crane, Tim. 2003. *The Mechanical Mind: A philosophical introduction to minds, machines and mental representation*. 2nd edn. London: Routledge.

Culler, Jonathan. 1988. *Framing the Sign: Criticism and its Institutions*. Oxford: Oxford University Press.

---. 1997. *Literary Theory: A Very Short Introduction*. Oxford: Oxford University Press.

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. 111-50. Oxford: Oxford University Press.

de Man, Paul. 1973. "Semiology and Rhetoric". *Diacritics* 3.3: 27-33.

Dening, Greg. 1996. "A Poetic for Histories". *Performances*. 33-63. Chicago: University of Chicago Press.

Dreyfus, Herbert L. 1965. *Alchemy and Artificial Intelligence*. Paper P-3244. Santa Monica CA: Rand Corporation. www.rand.org/pubs/papers/P3244/ (23 March 2008).

Denning, Peter. 2007. "A New Interview with Peter Denning on the Great Principles of Computing". *Ubiquity* 8.22. www.acm.org/ubiquity/interviews/v8i22_denning.html (7 April 2008).

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

Ellis, Allan B. and F. André Favat. 1970/1966. "From Computer to Criticism: An Application of Automatic Content Analysis to the Study of Literature". In *Science and Literature: New Lenses for Criticism*. Ed. Edward M. Jennings. 125-37. Garden City NY: Anchor Books.

Engelbart, Douglas. 1962. *Augmenting Human IntellectL A Conceptual Framework*. Summary Report AFOSR-3233, for the Director of Information Sciences, Air Force Office of Scientific Research . Stanford CA: Stanford Research Institute. www.bootstrap.org/institute/bibliography.html (25 April 2008).

Fogel, Ephim G. 1965. "The Humanist and the Computer: Vision and Actuality". In Bessinger, Parrish and Arader 1964: 11-24. Rpt in *The Journal of Higher Education* 36.2: 61-8.

Forster, Leonard. 1978. "Literary Studies as a Flight from Literature?" *Modern Language Review* 73: xxi-xxxiv.

Frischer, Bernie *et al*. 2006. *Summit on Digital Tools for the Humanities: Report on Summit Accomplishments*. Charlottesville VA: Institute for Advanced Technology in the Humanities. www.iath.virginia.edu/dtsummit/ (7 April 2008).

Frye, Northrop. 1991. "Literary and Mechanical Models". In *Research in Humanities Computing 1. Selected papers from the 1989 ACH-ALLC Conference*. Ed. Ian Lancashire. 3-12. Oxford: Clarendon Press.

Galison, Peter. 1997. *Image and Logic: A Material Culture of Microphysics*. Chicago: University of Chicago Press.

Geertz, Clifford. 2000/1980. "Blurred Genres: The Reconfiguration of Social Thought". In *Local Knowledge: Further Essays in Interpretative Anthropology*. 3rd edn. 19-35. New York: Basic Books.

---. 1993/1973. "Religion as a Cultural System". In *The Interpretation of Cultures: Selected Essays*. New York: Basic Books. Reprint: London: Fontana Press.

Gilmour-Bryson, Anne, ed. 1984. *Computer Applications to Medieval Studies*. Studies in Medieval Culture XVII. Kalamazoo MI: Medieval Institute Publications.

Gooding, David. 1986. "How do scientists reach agreement about novel observations?" *Studies in the History and Philosophy of Science* 17.2: 205-30.

---. 1992. "Putting Agency Back into Experiment". In *Science as Practice and Culture*. Ed. Andrew Pickering. Chicago: University of Chicago Press.

Göranzon, Bo. 1993. *The Practical Intellect: Computers and Skills*. Artificial Intelligence and Society. London: Springer Verlag.

Hacking, Ian. 1983. *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge: Cambridge University Press.

Hockey, Susan. 2000. *Electronic Texts in the Humanities. Principles and Practice*. Oxford: Oxford University Press.

Hoover, David. 2007. "The End of the Irrelevant Text: Electronic Texts, Linguistics, and Literary Theory". *Digital Humanities Quarterly* 1.2. www.digitalhumanities.org/dhq/vol/001/2/ (9 March 2007).

Jackson, H. J. 2001. *Marginalia: Readers Writing in Books*. New Haven CN: Yale University Press.

Livingston, Ira. 2006. *Between Science and Literature: An Introduction to Autopoietics*. Urbana: University of Illinois Press.

Loevinger, Lee. 1949. "Jurimetrics: The Next Step Forward". *Minnesota Law Review* 33.5: 455-93.

Luisi, Pier Luigi. 2003. "Autopoiesis: a review and a reappraisal". *Naturwissenschaften* 90: 49-59.

Mahoney, Michael S. 2000. "The Structures of Computation". *The First Computers – History and Architectures*. Ed. Raúl Rojas and Ulf Hashagen. 17-32. Cambridge MA: MIT Press.

Manovich, Lev. 2001. *The Language of New Media*. Cambridge MA: MIT Press. For the database chapter, see www.manovich.net/DOCS/database.rtf (25 April 2008).

Martindale, Colin. 1978. "Sit with the Statisticians and Commit a Social Science: Interdisciplinarity Aspects of Poetics". *Poetics* 7: 273-82.

Masterman, Margaret. 1962. "The Intellect's New Eye". *Freeing the Mind. Articles and Letters from The Times Literary Supplement during March-June, 1962*. 38-44. London: Times Publishing Company.

Maturana, Humbert R. and Francisco J. Varela. 1980. *Autopoiesis and Cognition: The Realization of the Living*. Boston Studies in the Philosophy of Science 42. London: D. Reidel.

McCarty, Willard. 1993. "Handmade, Computer-assisted, and Electronic Concordances of Chaucer". *CCH Working Papers 3: 49-65.*

---. 2004. "Network with a Thousand Entrances: Commentary in an Electronic Age?" In *The Classical Commentary: Histories, Practices, Theory*. 359-402. Ed. Roy K. Gibson and Christina Shuttleworth Kraus. Leiden: Brill.
---. 2005. *Humanities Computing*. Basingstoke: Palgrave.
---. 2008 (forthcoming). "Response: Perspectives of Literary Theory". *Journal of Literary Theory* 2: 476-8.
McGann, Jerome. 2001. *Radiant Textuality*. Basingstoke: Palgrave.
---. 2004. "Marking Texts of Many Dimensions". In Schreibman, Siemens and Unsworth 2004: 198-217.
---. 2006. "Visible Langage, Interface, IVANHOE". In *The Scholar's Art: Literary Studies in a Managed World*. 148-71. Chicago: University of Chicago Press.
Merton, Robert K. and Elinor Barber. 2004. *The Travels and Adventures of Serendipity: A Study in Sociological Semantics and the Sociology of Science*. Princeton: Princeton University Press.
Milic, Louis T. 1966. "The Next Step". *Computers and the Humanities* 1: 3-6.
Opas, Lisa Lena and Thomas Rommel. 1995. "Introduction. Special Section: New Approaches to Computer Applications in Literary Studies". *Literary and Linguistic Computing* 10.4: 261-2.
O'Sullivan, Sinéad. 2004. *Early Medieval Glosses on Prudentius'* Psychomachia. Mittellateinische Studien und Texte 31. Leiden: Brill.
Pegues, Franklin J. 1965. "Editorial: Computer Research in the Humanities". *Journal of Higher Education* 36.2: 105-8.
Potter, Rosanne, ed. 1989. *Literary Computing and Literary Criticism. Theoretical and Practical Essays on Theme and Rhetoric*. Philadelphia: University of Pennsylvania Press.
---. 1991. "Statistical Analysis of Literature: A Retrospective on *Computers and the Humanities*, 1966-1990". *Computers and the Humanities* 25: 401-29.
Rauch, Jonathan. 2006. "Sex, Lies, and Videogames". *Atlantic Monthly* 298.4. www.theatlantic.com/doc/200611 (12 April 2008).
Raymond, Darrell and Frank Wm. Tompa. 1988. "Hypertext and the *Oxford English Dictionary*". *Communications of the ACM* 31.7: 871-9.
Richards, I. A. 1926. *Principles of Literary Criticism*. London: Routledge & Kegan Paul.
Rommel, Thomas. 2004. "Literary Studies". In Schreibman, Siemens and Unsworth 2004: 88-96.
Rosen, Robert. 2000. *Essays on Life Itself*. Complexity in Ecological Systems. New York: Columbia University Press.
Scharfstein, Ben-Ami. 1989. *The Dilemma of Context*. New York: New York University Press.
Schreibman, Susan, Ray Siemens and John Unsworth, eds. 2004. *A Companion to Digital Humanities*. Oxford: Blackwell. www.digitalhumanities.org/companion/ (9 March 2008).
Schrödinger, Erwin. 1992/1944. *What is Life? Physical Aspects of the Living Cell* with *Mind and Matter* and *Autobiographical Sketches*. Cambridge: Cambridge University Press.
Simon, Herbert A. 1957. Untitled lecture notes for a talk given to the Operations Research Society of America, 14 November 1957. *Herbert Simon Collection*, Lectures and Talks 1951-1991. diva.library.cmu.edu/webapp/simon/item.jsp?q=box00053/fld03599/bdl0001/doc0001/ (23 March 2008).
----- and Allen Newell. 1958a. "Heuristic Problem Solving: The Next Advance in Operations Research". *Operations Research* 6.1: 1-10.
-----. 1958b. "Reply: Heuristic Problem Solving". *Operations Research* 6.3: 449-50.
Smith, John B. 1989/1978. "Computer Criticism". In Potter 1989: 13-44.
Solmes, Friedrich. 1961. "Aristotle's word for 'Matter'". *Didascaliae: Studies in Honor of Alselm M. Albareda, Prefect of the Vatican Library*. Ed. Sesto Prete. 393-408.
Tabbi, Joseph. 2002. *Cognitive Fictions.* St Paul: University of Minnesota Press.

Taylor, Charles. 1985. *Philosophy and the Human Sciences*. Vol. 2 of *Philosophical Papers*. 2 vols. Cambridge: Cambridge University Press.

Taylor, Frederick Winslow. 1913/1911. *The Principles of Scientific Management*. New York: Harper. www.marxists.org/reference/subject/economics/taylor (25 April 2008)

Teeuwen, Mariken. 2002. *Harmony and the Music of the Spheres: The* Ars Musica *in Ninth-Century Commentaries on Martianus Capella*. Mittellateinische Studien und Texte 30. Leiden: Brill.

Van Andel, Pek. 1994. "Anatomy of the Unsought Finding. Serendipity: Origin, History, Domains, Traditions, Appearances, Patterns and Programmability". *British Journal of the Philosophy of Science* 45: 631-48.

Wieland, Gernot R. 1985. "The Glossed Manuscript: Classbook or Library Book?" *Anglo-Saxon England* 14: 153-73.

Wilks, Yorick Alexander. 1972. *Grammar, meaning and the machine analysis of language*. London: Routledge & Kegan Paul.

Wittig, Susan. 1978. "The Computer and the Concept of Text". *Computers and the Humanities* 11: 211-15.