

Neglected, not rejected: Is there a future for literary computing?

Willard McCarty
Centre for Computing in the Humanities
King's College London
staff.cch.kcl.ac.uk/~wmccarty/

Talk for the School of Humanities 20/20/20 series
12 March 2008
(note references [SLIDE] to the accompanying ppt)

Abstract. The application of computing to literary studies originated in machine translation and philological analysis in the late 1940s. By the early 1960s the revolutionary promise of a "telescope for the mind", as the visionary computational linguist Margaret Masterman declared, seemed imminent. By mid decade the American poet and professor of English Ephem Fogel was advising his colleagues not to be discouraged by the "Vision-Actuality Interval" but clearly thought that traversing it was only a matter of time, fortitude and resources. But by 1978, the whistle was being blown on revolutionary claims, e.g. by Susan Wittig, who noted that the incremental gains could not be denied but that the world of aesthetic discourse remained unchanged. The problem, she said, was that no one had questioned "the concept of text" on which literary computing had been based. In 1989 Rosanne Potter explained that literary computing had not been rejected, simply neglected. It didn't matter. In fact, from Wittig's declaration in 1978 to Jerome McGann's considerably more profound and exciting reiteration of her point at the beginning of this century, literary computing has remained in "the ghetto of specialist periodicals", as Thomas Corn has remarked.

In this talk I will consider what may be gained by asking the Wittig-McGann question from the perspective of humanities computing. I will suggest that the solution to being neglected is to look the other way and to go wide rather than deep.

What remains indubitable is that the new is never as mellow as the old, and that therefore the worship of mellowness is incompatible with new excellence.

Bertrand Russell, *Portraits from Memory*, p. 93

1. Lessons of history, 1962-1989

For many years practitioners who apply computing techniques to the study of literature have lamented the minority status of what they do. In 1989, for example, Rosanne Potter introduced a collection of essays on the topic, *Literary Computing and Literary Criticism*, with the remark that "It has not been rejected, but rather neglected" – a complaint repeated many times since.¹ Like artificial intelligence, literary computing had early successes leading to visionary predictions, though on a somewhat more modest scale than AI's. For example, in her contribution to the series *Freeing the Mind*, published by the *Times Literary Supplement* in 1962, the British computational linguist Margaret Masterman declared that computing offered us a "telescope of the mind" that would "enlarge the whole range of what its possessors could see

¹ Potter 1989: xvi; see also 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.

and do" and so change "their whole picture of the world" (38). But also like AI, literary computing ran into unforeseen difficulties. Three years after the *TLS* series, poet and English professor Ephim Fogel was cautioning enthusiasts against underestimating what he called the "Vision-Actuality Interval", though he clearly thought that traversing it was only a practical matter of time, fortitude and resources (1965). The wherewithal was certainly in hand by the latter half of the 1970s (Rommel 2004: 93), but already it was clear that Fogel's metaphor was the wrong one. As Susan Wittig pointed out in 1978, computing had undoubtedly allowed for incremental progress, making performance of old tasks more efficient and accurate, but had not delivered on its promise (211). Attempting to traverse the interval had revealed a chasm yawning beneath.

Looking back on 20th-century work, Susan Hockey has summarized the prevalent view of the computer "as an invaluable assistant" to literary scholarship, most usefully deployed to probe for textual surface-features and so to prompt the enquirer to reflect "on the methodology used to interpret the results" (2000: 84). But the problematic separation that this view implies between theory and practice was already clear by the late 1970s, when blame for practitioners' failure to realise Masterman's vision was fixed squarely on inattention to theory. A computing without theory, Colin Martindale argued, was no better than a method in search of a paradigm to direct and explain it (1978: 275f). Wittig argued that as a result literary computing was vulnerable to covert influence by a positivistic "concept of text" derived ultimately from New Criticism (1978: 211). 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 (1989/1978: 7). Since then, with few exceptions, literary computing has had very little to say in response to the discourse of criticism.

It would be a mistake, however, to say that literary computing has been neglected simply because it has been neglectful of theory. For the past several decades the concerns of literary theorists have offered few points of contact, and so few possibilities of coaxing literary computing out of its theoretical silence. In the article I have already cited, Martindale pointed out that both "have strengths and weaknesses, but the 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 empirical and theoretical approaches have alternated rather than meshed, as if each supplied the whole answer. As a result, Leonard Forster noted in his Presidential Address to the Modern Humanities Association in 1978, we get dogmatic abstractions, the criticism formed around them becoming what he called "a flight from literature". 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 tool metaphor is a good one, not only because it privileges the task over the tool but also because in the context of skilled use a tool is not "just a tool" but an *active interface* between craftsman and material.

2. The matchmaker's tasks

Martindale's Leibnizian marriage of criticism and computing requires two things of the matchmaker. The first is to articulate emergent theoretical tendencies of existing tools so that the cognitive implications of the empirical toolkit may be known. The second is to engage with the answer provoked by the question that the first task implies: if literary text is more than what these tools say that it is, then what is it? From that engagement comes not merely better tools (hence another round of provocations to theorizing) but also the spelling out of theory that tool-building requires.

In *Humanities Computing* (2005) I articulated the theoretical implications of computing as an analytical approach to 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 enabled to probe them. For literary studies this epistemology takes computing significantly beyond the standard view of an efficient but essentially mute handmaiden to scholarship by challenging us in detail to account for the failure of any rigorously analytical try for a systematic order of things.

So far so good. But the analytical machine orientates us toward the words on the page as if looking at a specimen under a microscope from which essential bits are puzzlingly absent or out of focus. (More accurately, we see not words but character-strings, not on any codex page contextualized by the riches of its material culture, but on screen.) This denuded microscopic orientation provides useful input to criticism, but in order to *do* criticism we have to make up for its radical deficiencies by supplying, often less than consciously, whatever we happen to know as readers. This is not in itself a fault, nor is it anything new: there's always been more to criticism than could be enumerated. What's new is the accounting machine. But accounting begins with the choice of what to account for, and so tends to exclude the uncountable from consideration. Hence it has left the big question hanging fire – the one Wittig asked in 1978 and Jerome McGann again more recently (2004): *what is text?* What is that which we are attempting to compute? The analytic approach I have outlined gives us no help here because it begins with

a finite textual corpus viewed more or less objectively; it allows reference to a context more or less arbitrarily defined by what the critic happens to know or chooses to include. Hence the crippling difficulties of infinite regress that the very idea of context appears to cause whenever anyone asks what it is a promissory note for (Scharfstein 1989).

This is an inherited problem, of course. But by failing to ask the Wittig-McGann question, literary computing has nothing to offer toward its solution. Furthermore it makes the critic vulnerable to a largely overlooked side-effect of the deluge of data, which as it increases raises the probability of arbitrary choice. Hence the poet's question, in Eliot's *Choruses from the Rock*: "Where is the knowledge lost in information?" (I.16).

3. What the tools say

There is, of course, no single answer to the Wittig-McGann question: there are many kinds of text and many ideas of what to do with them. But the matchmaker's first task gives us at least a start. True, the tools we have used consistently from the earliest days of philological and literary computing leave the big question unanswered. But to do anything useful at all, they must have a view on it, which we can excavate.

Initially the answers we get back are impoverished. A concordancer, for example [SLIDE], 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. Both relational database design [SLIDE] and formal ontologies imply an instantiated set of concepts and their interrelations, and prior to these, well-defined perspectives of enquiry. An annotation tool [SLIDE] affords a view of text as the occasion for commentary. A statistical analyzer [SLIDE] yields a complex population of verbal clues to a literary style. These are all valid but skeletal at best – not much for the matchmaker to work with.

We can, however, greatly improve the view by considering the historical origins of these tools. The most obvious is concordancing [SLIDE], a direct descendant of the remarkably stable device invented in the late 12th or early 13th-century to serve figural interpretation of the Bible (McCarty 1993). The keyword-in-context format, devised in the 1950s to satisfy the needs and capabilities of automation, shifted focus from concordant passages of a text to shared collocates of a word, and so moved the principal domain of use from literary studies to corpus linguistics. Nevertheless the mechanized idea of semantic triangulation basic to the figural scheme remains implicit in the tool and so in the results the tool produces. So also the tools and techniques of annotation [SLIDE] 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 the like, together with their social networks (McCarty 2004). Relational database design [SLIDE] and textual ontologies are similarly emergent from older practices of categorization and tabular layout, and more recently from the strong cultural predisposition toward discontinuous plurality. Lev Manovich's argument for the database as a symbolic form provides a starting point here (2001). [SLIDE]

4. Turning the other way

Conceptual archaeology is a necessary step in the theoretical maturation of literary computing, but at best it yields a semi-coherent miscellany. In contrast the point of the Wittig-McGann question is to afford opportunity for modelling *and continually remodelling* the relationship between text and software. It is to allow software development a view of textual artefacts that transcends any finite accounting of them. It is to put this relationship always in question. It is, by the *modus operandi* of the humanities, to put questioning at the heart of literary computing.

Zeno's Paradox tells us that no matter how quick we are, respecting Fogel's interval is no way to think about the problem of literary computing. McGann is right: the fundamental move is to effect within literary computing the reversal of perspective already proclaimed by the Bakhtinian situating of text in "the immense, boundless world of others' words" (Bakhtin 1986: 143). To use an analogy central to my forthcoming book, its point is comparable to the methodological reversal implicit in Erwin Schrödinger's question, "what is life?", which he asked in response to the strange failure of physics as he knew it, and largely as we know it, to provide an answer (1992/1944). In his commentary on Schrödinger's project in *Essays on Life Itself*, Robert Rosen argues that "our universes [of scientific discourse] are limited, 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). The alternative for which he argues amounts to a turn toward the quasi-teleological but non-deterministic idea of self-organizing systems, hence the ideas of complexity, emergence and autopoiesis coming these days primarily from the biological sciences.

To an audience of humanist scholars a swerve from literary theory to theoretical biology may seem eccentric if not dangerous. The fact is, however, that our matchmaker is not well served by either humanities computing or literary criticism and so must borrow elements of a bridging discourse from wherever. Being orientated to instruments and measurement, the nearby sciences are the obvious places to go, though not the only ones. Theoretical

biology (from which McGann borrows) and biological anthropology offer essential clues, but I am also developing connections with semiotics, anthropological linguistics, improvisational musicology and the cognitive sciences, which likewise study how, to quote a London improvisational musician, one gets "from A to C when there is no B" (Bailey 1992: 136).

For computing the central idea is modelling, an iterative, manipulatory and progressive operation. There are three kinds. Clifford Geertz has distinguished two of them: analytical "modelling-of", which manipulates a representation of an object or phenomenon in order to find out more about it, and "modelling-for", which works from an idea or design to its realisation. In the forthcoming book I am concerned with a third kind, which resembles modelling-for but without a pre-existing design. Using the musicological term, I call it "improvisational modelling" to denote its serendipitous moment-by-moment development in performance of the unrealised potential of a text. Pek Van Andel points out that *pure* serendipity cannot be programmed (1994; cf. Merton and Barber 2004), but then modelling is not pure. It is a compromising interplay between algorithms and the world. So there is no contradiction, only questions to be asked.

5. Improvisational criticism

For literary computing, the largest and most urgent of the specific questions is how great a role improvisational modelling can play in the act of criticism. It is to ask, how close can such modelling come to the development of scholarly reading that is fully attentive to Bakhtin's "immense, boundless world of others' words" (which is all of literature). How can we build not the microscope literary computing has been taken to be but the "telescope of the mind" Masterman envisioned? How can literary computing strengthen literary criticism's pursuit of the humanities' fundamental purpose, which I take to be exploration, in Jerome Bruner's words, of "the alternativeness of human possibility" (1986: 53)?

In 1989 Northrop Frye pointed the way by suggesting the potential of modelling to converge on fundamental structures of literature through systematic investigation of its recurring conventional units (1991: 6). Since then, with rather different emphasis, McGann and colleagues have developed a computer-assisted game, *IVANHOE*, in which player-modellers develop unrealised possibilities of a given literary text by selectively rewriting it from an authorial "inner standing-point" (2006: 148-71). In writings about this and other "speculative provocations" of computing, they have promisingly sketched out possibilities toward a theoretical discourse.² Much work needs to

² See the work of the Speculative Computing Laboratory, University of Virginia, www.speculativecomputing.org/ (9 March 2008).

be done, as I have suggested, by extending the range of borrowings and working out how to assimilate them. Similarly, *IVANHOE* suggests improvisational modelling but does not carry it out. *IVANHOE* uses computing to record improvisational moves; it doesn't *model* these moves.

So we are left with two principal questions. First, can the co-creative process of reading itself be accommodated sufficiently well that we might say it is being modelled? Second, if we begin with reading conceived in this way, are there common threads, perhaps of thinning relevance, that we may follow out from it into the disciplines concerned with how new things come about? Are the pidgins one speaks with each of the relevant epistemic cultures dialects of a single language of method?

To explore these questions I centre on the trope of personification in Ovid's *Metamorphoses*, or more precisely the "personification figure" (Paxson 1994: 35), which is an occurrence too brief to have direct narrative consequence. Following Morton Bloomfield's lead (1963), I define it as an ontological disturbance, anthropocentric but not necessarily anthropomorphic, created in the act of reading by attribution of unusual characteristics to a non-human entity. This disturbance moves the affected entity ontologically toward but not necessarily to the human. In *Humanities Computing* I studied these personification figures analytically, and thus encountered the problem of arbitrary context. Now, by modelling the process improvisationally with the aid of a teachable interactive environment, I propose to explore how they are synthesized. The aim is simultaneously to design such an environment and to develop its explanatory theory. Success of the former will be measured by the degree to which the two thousand year-old reception-history of the *Metamorphoses* provides matching adaptations and borrowings. Success of the latter will be demonstrated by the fruitfulness of the expanded theoretical amalgam (perhaps, one day, even a compound) to stimulate other efforts.

In the abstract I said that the solution to literary computing's neglect is "to look the other way and to go wide rather than deep". I meant, of course, to turn from the analytical microscope to Masterman's "telescope of the mind" and from narrow obsession with the inadequate vocabularies to help wherever it may be found.

Works cited

- 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.
- 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.
- Bloomfield, Morton. 1963. "A Grammatical Approach to Personification Allegory". *Modern Philology* 60.3: 161-71.
- Bruner, Jerome. 1986. "Possible Castles". In *Actual Minds, Possible Worlds*. 44-54. Cambridge MA: Harvard University Press.
- 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.
- Fogel, Ephim G. 1965. "The Humanist and the Computer: Vision and Actuality". *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.
- 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.
- Geertz, Clifford. 1993/1973. "Religion as a Cultural System". In *The Interpretation of Cultures: Selected Essays*. New York: Basic Books. Reprint: London: Fontana 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).
- Manovich, Lev. 2001. *The Language of New Media*. Cambridge MA: MIT Press.
- 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.
- 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.

- McCann, Jerome. 2004. "Marking Texts of Many Dimensions". In Schreibman, Siemens and Unsworth 2004: 198-217.
- . 2006. "Visible Language, 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.
- 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.
- Paxson, James J. 1994. *The Poetics of Personification*. Literature, Culture, Theory 6. Cambridge: Cambridge University Press
- Potter, Rosanne, ed. 1989. *Literary Computing and Literary Criticism. Theoretical and Practical Essays on Theme and Rhetoric*. Philadelphia: University of Pennsylvania Press.
- 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.
- Russell, Bertrand. 1958. *Portraits from Memory and other essays*. London: Reader's Union, George Allen and Unwin.
- 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.
- Van Anandel, 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.
- Wittig, Susan. 1978. "The Computer and the Concept of Text". *Computers and the Humanities* 11: 211-15.