What text really is not, and why editors have to learn to swim

Peter Robinson
Institute for Textual Scholarship and Electronic Editing, University of Birmingham, UK

Abstract

This article attempts to ask some fundamental questions about editing in the digital age, and give some answers to these questions. It is argued that a concentration on digital methods, for themselves, may neglect the base questions facing any editor: why is the editor making this edition; from whom is the editor making this edition? Indeed, in some respects thinking about text encoding for digital purposes has been built on assumptions which are, for editors, simply wrong. In particular, the concept of what 'text' is, upon which (for instance) the Text Encoding Initiative principles are based (what Renear calls 'realist'), is positivist, overconfident, simplistic and neglects the materiality of actual text instances. This view is opposed by what Renear calls 'anti-realism': texts do not have an independent existence, but are constructed by individual and collective acts of perception. In concrete terms, 'anti-realism' sees editions as made to serve the needs of the reader, as acts of interpretation and not as representations of some concrete reality: this is Pichler's view of the Wittgenstein transcripts, and the author's views of the Canterbury Tales project transcripts. However, it is argued that both realist and anti-realist extremes are dangerous: 'realism' can lead to editions which are arrogant and out-of-touch; anti-realism to editions which are reductionist and etiolated. In place of either extreme, we should substitute a different aim: to challenge readers to make new texts for themselves as they read, by finding new ways of presenting material so that both we editors and those who use our editions become better readers.

Correspondence:
Peter Robinson, Elmfield House, Selly Oak Campus, University of Birmingham, Edgbaston B29 6LG, UK.
E-mail:
p.m.robinson@bham.ac.uk

This essay was written in 1997. It has not been significantly updated but reflects that particular moment in the debate on electronic text.

Conference papers have a way of reshaping themselves, according to the exact circumstances in which they are conceived, delivered and published. I prepared one version of this conference paper well before I came to Toronto for the Conference on Editorial Problems, in November 1997. Then, I altered the paper just before the conference when I realized that there was to be a conference at Brown University, the weekend after the Toronto conference, to mark the tenth anniversary of the Text Encoding Initiative. At the Toronto conference itself, I altered the article again, after seeing the title and abstract of Michael Sperberg–McQueen's paper. And now, as the adrenalin of the conference recedes to be replaced (in theory) by considered reflection, I present yet another version of these thoughts.[1]

This evolution, of course, is rather like editions themselves. One uses the tools available, on the materials one has, for the audience who (one imagines) will use the edition. As the tools, the materials, the audience change, so the edition will change. Over the past two decades we have seen the most

Literary and Linguistic Computing, Vol. 24, No. 1, 2009.
doi:10.1093/llc/fqn030

dramatic change in the tools available to scholarly editors: a change arguably greater than the advent of print in the late fifteenth century. This change is the coming of computer technology, which is altering every aspect of how editions are researched, published and read.[2] It is to be expected that the impact of a dramatic tool would cause thinking editors to concentrate their thoughts on the tool itself. The composition of the panel of invited speakers reflects this. Where other conferences in the Toronto series of editorial problems series have been constructed around questions of editing a particular genre or period,[3] the common thread of the speakers at the November 1997 conference was their use of computing technology towards the making of editions: the 'how?' of this tool rather than the 'what' or 'why' of editions.

We are struggling just to learn to use this tool. In the course of this struggle, I believe we have neglected other aspects of the making of editions. In particular, we have not considered adequately the 'what' and 'why': what sort of editions are we making? Why and for whom are we making them? We are beginning to master the tool, and the first electronic editions are appearing.[4] It is time then that we turned from the question of 'how' to the questions of 'what' and 'why'. This is particularly important, because the understandable and necessary concentration on 'how' over the last years has implied answers to 'what' and 'why' which are not appropriate to the making of editions. In this article, I wish to review the kinds of edition prompted by the concentration of the last years on 'how', explain how these are inappropriate for the kinds of material and audiences which are my concern, and suggest other models for editors.

Ten years ago, about the time I first began to start the work which has now become the Canterbury Tales Project, many of us already had computers and there had already been a great deal of activity in the area of texts and computing. But an electronic edition like my *Wife of Bath's Prologue on CD-ROM* or Jerome McGann's Rosetti, was quite impossible, and indeed inconceivable except for the widest eyed dreamers.[5] Ten years on, we have gone past dreaming, and the first electronic editions have appeared. We now have the tools, or some of them; we have tested them; we have begun to use them.

What are these tools, which now make possible electronic editions, which were not possible ten years ago? The advance in personal computing, which has now brought a fast computer with internet access to the desk of most academics, is the most obvious. Less obvious, but actually more important and indeed crucial, has been the work of the Text Encoding Initiative. There is a neat apposition in the occurrence of the Conference on Editorial Problems and the Text Encoding Initiative tenth anniversary conference within a week of each other. There is no doubt that without the TEI there would be no Canterbury Tales project and no Wife of Bath's Project CD-ROM, and any conference on editorial problems focusing on the use of computers would have been very different—or would not have occurred at all. More than powerful hardware able to display huge images in a fraction of a second, more than clever software able to find cunning patterns in vast text corpora, editors need a way of saying what they have to say: of saying this is a variant reading, that is a scribal deletion. The encoding devised and described by the Text Encoding Initiative permits editors to say just this, and much more.[6] One cannot overstate the importance of this work to editors trying to work in the new medium. Before TEI, an editor who wished to create an electronic edition had to devise his or her own scheme of encoding, and find or (more likely) develop software that could make use of the encoding.[7] The editor would then be trapped on a treadmill, of having to maintain and extend a specialized encoding and software system: in the special compartment of computer hell, as Murray McGillivray puts it, reserved for single-purpose *ad hoc* systems.[8] TEI freed editors from the private hells of their own encoding, and—because TEI is an implementation of Standard Generalized Markup Language (SGML; hence TEI/SGML)—offered a range of methods for making and publishing the texts they made.

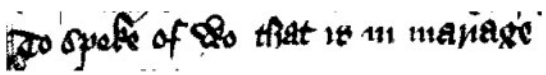
In the past few years, developing and using these systems based on the TEI/SGML encoding has been intoxicating. For me, at least, the struggle to master them, and then to apply them usefully, was so absorbing as to leave little space to ask the other

questions any maker of any edition must answer: why am I doing this; who am I doing it for? Now, we have reached the point where these questions must be asked. If editions are to be of interest for more than their technical mastery or of use to more than the people who made them, they must find a rationale and an audience. There is a danger that the marvelous tool of TEI/SGML encoding is accompanied by assumptions about rationale and audience which are actually inappropriate to textual scholars. In this article, I want to question some of these assumptions, and suggest other ways of thinking about editions which editors and textual scholars might find fruitful.

What I have to say in this article stems from what I think text is, and what text is not. Sometimes, we define ourselves most clearly not by what we believe, but by what we find viscerally, passionately, utterly impossible to accept. Looking back over the years, I can recall three incidents where I found myself in total disagreement with statements presented by other people. In each case, my disagreement was the more surprising to me, because those other people set forth their case as if what they were saying was something so perfectly obvious that surely no one could disagree. But I did disagree.

The first occasion was during a workshop in Oxford in 1991, to introduce the TEI guidelines to what was hoped would be the first users of the TEI. Jeremy Sinclair was sitting a few rows in front of me. In the course of one of those Jesuitical discussions on the minutiae of text encoding and character sets for which the TEI is justly famous, I could see that Jeremy was becoming increasingly agitated. Finally, he could endure it no longer. He stood up and declared 'What is all this fuss? When I see a stroke with a dot above it, I see an ''i''. So I encode it as an ''i'' and that is that'. He looked about himself as if daring anyone to challenge this obvious wisdom, and sat down to prolonged applause.

I could not join in that applause, then or now. There seemed something very, very wrong with that statement, but it took me a little time to see what it was. It came to me, though, the next time I sat down to transcribe Chaucer manuscripts. There I saw many marks which I encoded as 'i'. Some of these were strokes with dot above them: but many times, the dot was to one side, or it was a diagonal mark, or it was not there at all. And many times the stroke was not a single upright stroke at all: it was joined to the letter before or after, it had seraphs or horns attached at odd points. Hardly a single one of these conformed to Jeremy Sinclair's formula. Nevertheless, I transcribed all these marks as 'i': approximately some 70,000 'i's in the manuscripts of the Wife of Bath's Prologue. And here is an odd thing: nobody, in the two years since we published this, has come to me and said: this mark or that mark you have transcribed as an 'i' is not a stroke with a dot above it. You could argue that these were manuscripts, and so this is not a fair test of Sinclair's formula. Being nothing if not obsessive, I found myself looking at printed 'i's. And once more, I found that many of these printed marks which I and all the people in this audience agree are 'i', do not fit Sinclair's prescription of a single upright stroke with a dot over it. The stroke has seraphs; or is rounded at either end; or the dot is a square, or an ellipse—by now I was using a magnifying glass, so desperate was I to find, somewhere, a real 'i'. By now, too, I was puzzled. Sinclair's prescription declared that a stroke with a dot over it is an 'i', as if by some natural law. There seemed no such natural law. Yet how do we come to agree that all these marks are 'i's? Consider this:



The three strokes here which constitute the second to last word are near identical to the three strokes which commence the last word (and indeed, they are identical if one takes the somewhat larger seraph on the first stroke of the second last word as purely ornamental). Yet, we declare in our transcript that the first set of three strokes stands for the letters 'in', while the strokes which begin the last word stand for the letter 'm': thus, 'in mariage'. Why do we transcribe one set of three strokes as 'in' then, and another as 'm'? We do this because this is the only reading which makes sense. And none of the hundreds of Chaucer scholars who have looked at this have declared they disagree with our transcript.

I can now see—it has taken me so many years to come to this point—exactly why Sinclair was wrong.

An 'i' is an 'i' not because it is a stroke with a dot over it. An 'i' is an 'i' because we all agree that it is an 'i'. In forming this agreement, the exact visual appearance of the marks on the page is one of the determining factors, but only one of the determining factors. Other determining factors include our agreement that these marks actually have a linguistic meaning and are not some kind of random scattering of ink or bird dung; that they should be read from left to right; that together they constitute a passage of Middle English belonging to a poem we call the *Canterbury Tales* and so on.

I can vividly recall the circumstances of Sinclair's assertion. However, I can not offer so exact a context for the second statement, with which I am in profound disagreement. This is because the person who made it has said it so often that, in my mind, every time I see him (and I see him quite often) I am reminded of it. The person is Michael Sperberg-McQueen, and his assertion, roughly represented, goes like this. I have these two texts:

This is a text

This is a text

Michael has said, many times in my presence and I imagine many times elsewhere, that these two texts are actually the one text. They differ only in that the first is in a roman font, the second is in an italic font. But what, precisely, do we mean by saying 'these two texts are really the one text'? Consider this third version of the same text:

txet a si sihT

Most people would, I think, agree that this is not 'the same text' as our first example, here. Yet, graphically it is far closer to the first version than is the second, italic, example. They have exactly the same letters, in exactly the same shape, letter by letter; but in this third example these letters are reversed. Now, in a writing system in which one can read letters from right to left, or left to right, as the mood takes you, the first and third instances are identical. You have the same letters; you have them in precisely the same order. It is just that in the first instance you start reading at the left, in the third you start at the right. In such a writing system, we would say that the first and third texts are identical, while the second text might be quite different.

My difference with Sperberg-McQueen on this point is apparently rather slight, but on slight differences vast empires are built. His statement is 'these two texts are the same text, differing only in their graphic realization'. Instead, I would say 'we agree, for the purposes of this encoding, that these texts are the same text, differing only in their graphic realization'. The point of my qualification 'for the purposes of this encoding' is that one can well imagine different encodings in which indeed the texts are different. An encoding which foregrounded the presentation of text, for example, would see the roman and italic forms of the text as different. A graphic designer would certainly see these texts as different, as would a child who is learning to read, as would an editor who was encoding for 'bibliographic' rather than 'linguistic' codes. There are, already, such alternative encodings. And we can be sure that over time there will be many, many, more. The consequence of this is that any encoding, any editing, any reading, is conditional. And because they are so conditional, any pretence that any encoding is for all time is a delusion. We are bound by what we, as readers, see, and what we see is determined by our linguistic community: by the systems of signs and conventions we share. In time, this linguistic community will change. And when it does, and it will, our encodings, the systems of signs which make up our editions, will be inappropriate and will have to be redone.

I have introduced here the concept of an edition as a system of signs. I am not a philosopher, and those who are can see behind this characterization a century's thinking about semiotics, from Charles Saunders Peirce on. During 1992 and 1993, Elizabeth Solopova and I were developing the system of transcription which now underpins the whole Canterbury Tales project. In the course of this, we formulated the relationship between the marks on the page in the original manuscript, the original primary text and the electronic transcription that we were making. We stressed that the transcription, the edition, is a distinct linguistic object, in these words (Robinson and Solopova, p. 21):

> Any primary textual source . . . has its own semiotic system within it. As an embodiment of an aspect of a living natural language, it has

> its own complexities and ambiguities. The computer system with which one seeks to represent this text constitutes a different semiotic system, of electronic signs and distinct logical structure. The two semiotic systems are materially distinct, in that text written by hand is not the same as the text on the computer screen. They are formally distinct, in that a manuscript may contain an unlimited variety of letter forms but a computer fount ordinarily will not. They are logically distinct, in that the computer transcription will attempt to resolve ambiguities present in the natural language of the primary source (e.g. the same graph being used for distinct letters; [as in the example of forms of 'i' and 'm' above]): if the transcription does not do this, it will betray its principal aim of decoding of the primary source. Transcription is both decoding and encoding; the text in the computer system will not be the same as the text of the primary source.[9]

From this, we concluded that transcriptions, and hence editions are translations:

> In the course of our work we have come to realize that no transcription of these manuscripts into computer-readable form can ever be considered 'final', or 'definitive'. Transcription for the computer is a fundamentally interpretative activity, composed of a series of acts of translation from one system of signs (that of the manuscript) to another (that of the computer) …. Like all acts of translation, it must be seen as fundamentally incomplete and fundamentally interpretative.

In essence, we abandoned the premise that our transcriptions could aspire to anything like 'objective truth'. Many translations, many transcriptions, are possible. And so we faced a new problem: in our terms, what is a good transcription? What distinguishes a good transcription from a bad transcription? The answer we proposed was this:

> … our transcripts are best judged on how useful they will be for others, rather than as an attempt to achieve a definitive transcription of these manuscripts. Will the distinctions we make in these transcripts and the information we record provide a base for work by other scholars? How might our transcripts be improved, to meet the needs of scholars now and to come?[10]

At about the same time as Solopova and I were coming to this conclusion, a researcher in the Wittgenstein Archive in Bergen, Alois Pichler, was also considering the nature of computer transcription, and was coming to identical conclusions. In a paper published in 1995, he puts it thus: 'Machine-readable texts make it … clear to us what texts are and what text editing means: Texts are not objectively existing entities which just need to be discovered and presented, but entities which have to be constructed.'[11]

Pichler thus found himself faced with exactly the same question that we faced: how does one determine, what is a 'good' transcription, a 'correct' edition? And, he arrived at exactly the same answer as we had: 'the essential question is not about a true representation, but: whom do we want to serve with our transcriptions? Philosophers? Grammarians? Or graphologists? What is ''correct'' will depend on the answer to this question. And what we are going to represent, and how, is determined by our research interests … and not by a text which exists independently and which we are going to depict.'[12]

In a much quoted article published in 1990, DeRose, Mylonas, Durand and Renear asked the question 'What is text, really?'[13] To this question, Pichler proposed the answer: text, really, is not. Text does not exist outside the meanings we create: and these meanings are all the text we will ever know. Each text we create is a system of signs: we share these signs with others, and thus the individual text of each of us becomes part of our communal experience. Text is not a tree, which will live whether we perceive it or not. For, text without human perception is just marks on paper, or sounds in the air. These marks and sounds only become text when we find meaning in them.

This brings me to the third of the three statements, emanating from the TEI, with which I find myself in fierce disagreement. In late 1995, Claus Huitfeldt of the Wittgenstein Archive initiated an

online discussion on philosophical issues in text encoding, under the auspices of the philosophical journal *The Monist*. To start this discussion, Allan Renear of Brown University wrote a paper 'Theory and Metatheory in the Development of Text Encoding', and published this to the discussion group.[14] In this paper, Renear took issue with the position presented by Pichler (he was apparently not aware that we held identical views). He labelled this view 'antirealism'. This seems to me a rather good term, in that we deny that text has any reality beyond that which we give it through our individual and collective acts of perception. Renear seems to find antirealism very disturbing. Here is the passage from this paper in which he tries to dismiss it:

> Suppose a transcription has nothing at all to do with the text but helps the researcher win a prize which allows her, by providing money and time, to figure out, finally, the crux she was working on. In that case a (false) transcription has served the researcher's interests quite well, but no one would claim that it is thereby a reasonable encoding or one which is to be in any sense commended as a transcription.

After several years of thinking about this passage, I am still astonished by this argument.[15] First, Renear appears to ask us to judge the transcription by the motives of the transcriber. Because the transcriber has produced a transcription in order to 'win a prize', and not in order to provide a 'true' transcription, therefore the transcription must be false. I thought we had done away with the intentionalist fallacy long ago. Is *The Brothers Karamazov* a bad book because Dostoyevsky wrote it to clear his gambling debts? If we are to judge editions by the motives of their editors, who shall scape whipping? One scholar edits to win tenure (a deluded soul); another out of vanity; another out of frustrated creative impulse and another to make a gift for a person he or she may never see. Does one of these motives make a better edition than another? Do I care, or know, why Lachmann edited Lucretius?

Second, more seriously, Renear misses the point of his own example. Why did this transcription win a prize? Unless the judges are totally corrupt, we must accept that it won a prize because at least one scholar found it useful: that is, at least one scholar found that the transcript helped illuminate something which was unclear, or opened up a new and fruitful avenue of enquiry. Renear may think the transcription false. But enough scholars thought otherwise to judge it worthy of the prize. In the terms set out by Alois Pichler and ourselves, this transcription is therefore successful. Someone found it useful; therefore it is useful. The researcher deserved the prize, regardless of what his or her motives might have been.

It seems to me significant that both Pichler and we work with manuscript materials and not with printed sources. In manuscripts, one is faced at nearly every word, even at every letter, with the need to choose: to choose to ignore this mark or that; to represent this mark by this letter or that; to encode this meaning or that. For us, 'antirealism' is not a contention. It is a simple description of what we do when we encode. In order to choose what meaning we encode, we have to determine, at every point, the purposes of our encoding. It might be possible for a transcriber of printed text to transcribe mechanically, simply substituting for the 'i' of the printer's fount the 'i' of the computer fount: Jeremy Sinclair certainly thought this. But for a transcriber of manuscript materials, or of spoken word, transcription is never simple substitution.

This returns us to the point of difference I found between me and Michael Sperberg-McQueen. I have no doubt that indeed Michael would agree with my statement: that decisions about what one should and should not encode are to be determined according to the purposes of our encoding. Where we differ is this: we see every aspect of our transcription and edition as determined by the need to determine the purposes of our encoding. We do not ask: what is the right encoding of this word. We ask: who is to use the text we make? What use do they want to make of it? What do we think this text is saying? How can we, as editors, help the text speak to its readers? A transcription, an edition, is 'right' only in that it might serve these purposes. An edition is an act of communication. If it does not communicate it is useless.

Yet, while I am clearly more comfortable with Pichler's formulation than with those of

Sperberg-McQueen and Renear, there are elements in Pichler's analysis with which I am uncomfortable, too, and where I find myself closer to Renear than to Pichler. In the last paragraph, I argue that our function is to help the text speak to its readers; that we perceive the text as saying something, and that we seek to help readers discover what this something is. This takes us away from what one might term Pichler's hard-line 'antirealism' to something nearer Renear's argument: the text does have an independent existence; it is saying something; we are not just 'constructing' an artefact for the use of our readers, but we are trying to interpret an utterance. A danger of the pure antirealist argument is that it could lead to transcriptions (and hence editions), which are narrowly sectional. If one is constrained in transcription only to record those phenomena for which one feels there is a clear use, then one might end with impoverished transcripts and with editions serving only rather rigidly determined needs. Who is to determine what these needs are? We do not know, and can only guess at, the various uses our transcripts might serve.

I stated above that at every point, as we transcribe, we ask ourselves: who is to use the text we make and how will our encoding help them to use the text. So far, this agrees exactly with Pichler. But there are times—many times in fact—when we find ourselves thinking: we do not know who is to use this text, or how they will use it, but there is something here in the text which seems important, and which we will therefore encode. Here, we part company with Pichler, towards something nearer Renear's realist model. We find ourselves thinking: what is the text saying at this point? Even if we cannot think of a use for this information, or even a transparent way of encoding it, we feel bound to try to encode what it is saying, somehow. A good example of this is our treatment, in the *Canterbury Tales* transcripts, of ornamental capitals and other kinds of emphasis in the manuscripts. One could formulate the aims of the Canterbury Tales Project in strict linguistic terms: we are interested in Chaucer's words, and our transcripts are to be used in trying to reconstruct the textual affiliations of the manuscripts as a crucial stage towards determining Chaucer's words. In these terms, the presentation of the manuscripts, as in their use of hierarchies of script, different types and gradations of ornamentation, could be seen as irrelevant. The presentation is determined by the producers of each manuscript, and is notoriously unstable from copy to copy. It is therefore difficult (and perhaps impossible) to use information about presentation in analysis of manuscript relations, and we would be quite justified, on a narrowly antirealist view, in ignoring all this.

At times, we have indeed discussed doing just this. But we have not been able to bring ourselves to do this, and our transcripts do record considerable detail about the mise-en-page of the manuscripts. Clearly, the appearance of the manuscript page was of great importance to the medieval scribe: why else should they have spent so long inscribing the page with these patterns? Clearly, too, the appearance of the manuscript page has considerable impact on modern readers. Because appearance is 'non-textual', encoding it in any kind of character-based transcription presents real difficulties. But the appearance of these manuscripts is undeniably part of their utterance, and we therefore feel obliged to try to represent this, somehow, in our transcripts.

In summary, one could represent the extremes of the 'realist' and 'anti-realist' approaches as the Scylla and Charybdis of editing. Extreme realism might encourage editors to think that they are gifted with a unique insight into the text they are editing, and so encourage them to produce editions which are highly expressive of that insight, with the most intricate and idiosyncratic encoding, but which are of little use to anyone else: the electronic equivalent of Bentley's *Paradise Lost*. Extreme antirealism might encourage editors to produce editions which are sectional, partisan and conformist in the worst sense, like the great majority of schools and so-called reader's editions. Allen Renear reminded me, when considering this, that a fundamental tenet of the realist position is the fallibility of perception: while 'the text' might have its own defining reality, independent of how we perceive it, yet we are able only to comprehend the one reality through the various misleading instruments of our senses and knowledge. The contradiction is as old as

philosophy. Heraclitus, in the fragment used by T. S. Eliot as an epigraph to *Four Quartets*, explained it thus: though the word (Greek 'logos') is common to all, yet we behave according to our own private vision.[16] Realists who forget they are fallible will produce editions which are arrogant and out-of-touch; antirealists who are not willing to trust their own private vision will produce editions which are reductionist and etiolated.

It is with these thoughts in mind that we have been reconsidering the work of the Canterbury Tales Project. I confess that the manner of presentation of our first CD-ROM, of the Wife of Bath's Prologue, might make it appear that all we are doing is compiling facts, in the realist mode: some ten million facts, indeed. Some commentators have supposed from this that it is therefore an archive, not an edition; that it represents some kind of impersonal statement of the condition of the text in all those manuscripts. This is not our aim. Our intention was and is to recover the history of the making of the text, in order to find better ways to read it: if you like, to present our private vision of the *Tales*. In order to do this, we have to reconstruct the earliest states of the text from the massive jigsaw puzzle of all these readings, in all these witnesses. From this we might determine the state of the text as Chaucer left it, and then how this text was transformed into what we now have in the manuscripts. Through the use of various computer tools, and through long enquiry, we think we have come a considerable distance towards these aims.

Let us summarize rapidly what we have found.[17] First, we have arrived at a more sophisticated classification of the manuscripts into family groupings than Manly and Rickert were able to achieve. They found four groups of manuscripts, ABCD, but could only fit around half the total number of manuscripts into these groupings. We have confirmed the existence of these groupings and found three further groupings of manuscripts: what we call EF and O, and together these groupings account for all of the manuscripts of the *Tales*, and not just half of them. Manly and Rickert were unable to find that the same set of manuscript relations applied to any two tales, and so thought that every tale had a separate textual history. Recently, our work on the General Prologue tradition suggests that Manly and Rickert were wrong about this too.[18] Different parts of the *Tales* do have the same set of manuscript relations and therefore they have the same textual history. To put this in another way: our analysis so far gives no evidence of there having been separate publication of any part of the *Tales* in Chaucer's lifetime. Further, our analysis of the patterns of variation in the manuscripts suggests that there is no evidence that Chaucer ever made a word-by-word, line-by-line revision of the text. However, there is considerable evidence that he did delete and add passages; that he altered ascriptions of the tales, that he was experimenting with different orders of the tales and with different links between them and that he was still doing this at the time of his death. In the particular case of the so-called 'added passages' in the Wife of Bath's Prologue, I argue that Chaucer indeed wrote these as part of the WBP, at a time when he intended the Wife to tell what is now the Shipman's Tale. However, when he changed his mind and gave the Wife the tale she now tells, he marked these passages for deletion. These passages present the Wife as violently and aggressively promiscuous: they would fit a Wife who told the Shipman's Tale, but not the Wife who tells the tale she now has.[19]

We think that what lies behind all the manuscripts of the *Canterbury Tales* is a single body of papers, which we call O. We believe this body of papers was Chaucer's working draft of the Tales, left unfinished at the time of his death. These papers were in a state of some disorder. Thus, much of the variation in the earliest manuscripts, of tale order, links and additional passages, is due to different scribes interpreting these papers differently. Further, over time this rather loose sheaf of papers would be likely to become yet more disordered, as sections were let out for copying and replaced in the wrong order, or not replaced at all. We believe that the Hengwrt manuscript, alone of all the extant manuscripts, is a direct copy made from O, made from this pile of papers. However, it was not the only copy made from O. There is strong evidence of at least one copy, now lost, having been made from O before Hengwrt was copied. This copy we call the alpha exemplar. We believe that many of the manuscripts descend from

alpha; it appears that the Ellesmere manuscript is copied in part from alpha; and that, perhaps, most crucially, Caxton used alpha or a manuscript very close to it when making his second edition. Alpha is particularly important as it appears that it was copied before the pile of papers constituting O became disordered. Thus, alpha had the Tales in the order we know as the a order: the order found in the Riverside and most modern editions, and the order which it appears Chaucer intended.[20]

We have found a good deal else. From close examination of the punctuation in the early manuscripts, Elizabeth Solopova has concluded that the punctuation system found in the Hengwrt and Ellesmere manuscripts was devised by Chaucer himself. She has also found, from study of the metre in the early manuscripts, tell-tale signs of the habits of copying of individual scribes.[21] From all this, and from the close study of the spelling of the manuscripts we are now able to carry out (using the spelling databases on the CD-ROM), we are close to developing a rationale for a new edition of the *Tales*.

You will not find any of this on the first CD-ROM. This was a mistake, and we will rectify this in future publications. Accordingly, the General Prologue electronic book will include a stemmatic commentary. In this, I will try to explain the significance of (for example) some manuscripts reading Auerill in the first line, others reading Aprill—or is it Aurille or Aprille? Underpinning the stemmatic commentary will be an 'Analysis Workshop' of the whole tradition: a reconstruction of the history of the copying of the text as I have just summarized it, and based on the cladistic and database tools we have used. We will include those same cladistic and database tools as part of the electronic publication, so that readers may use them to question our hypotheses, or to make their own.

We have beyond this yet further aims: we want our readers to experience the multiplicity of these many *Canterbury Tales*, as we have experienced them. We want readers to discover the text for themselves, from these marks on these manuscripts, as we have discovered it for ourselves: to develop a private vision, and then to communicate this to others, as we are attempting. And more than this: we want to bring this experience to a far wider range of readers, to readers who have never thought to look at a manuscript, or to puzzle over different versions of the one text. Indeed, this is happening. Within the first year of publication, Cambridge sold some 180 copies of the Wife of Bath CD-ROM. This seems rather good, considering that only some ten scholars in the world had published on the text of the Tales in the last decade. The most enthusiastic response I have had to our work has been from school children: they see the point of it immediately.

Indeed, the greatest promise of electronic editions is, I think, their potential to bring the scholarly tools of the highest quality to a far wider range of readers. But this will not happen if we simply pile all the materials together into a massive electronic archive and say to the reader: here you are. Great quantities of original source material are already readily available for the renaissance period; yet scholars and readers make very little use of it. In order to have people use our work on Chaucer as we want to have it used, we will need to do far more than just collect it together, encode it in beautiful politically correct SGML, and park it in a text archive somewhere. We will have to find a publisher; we will have to present it in the most attractive form possible and we will have to persuade people to use it.

It is here I part company most sharply with the model of editions presented by Michael, and indeed by the whole TEI endeavour. In his talk, Michael declares that scholars should not 'teach their editions to swim'. But if we want our editions to reach the widest possible audience, if we want to transform the way people read by the electronic editions we make, then we will have to make them far more attractive. We will have to include far better tools for viewing multiple texts, for viewing transcriptions alongside images, for navigating scholarly commentary. To do this, editions will indeed have to learn to swim.

And, editions can swim. In the Seattle Art Gallery in late 1997, you could see at any moment scores of people using the Corbis CD-ROM of the Leicester Codex of Leonardo da Vinci. This CD-ROM has a brilliantly designed tool that you pass over the manuscript image. This tool rotates Leonardo's mirror writing, presents it in Italian transcription, and in English translation, all simultaneously.

On one Saturday afternoon in October 1997, I counted eighty people sitting at terminals in the Gallery enraptured by this, with others queuing to take their place. This one tool, I suggest, does more to bring this Leonardo Codex alive to the thousands of people who have used it than any number of printed papers or lectures.

We too could have such tools; we too could present the many texts of Chaucer, the many manuscripts, so that they are a living speech. I am quite happy to use TEI encoding if it helps me to this end. But if it obstructs this aim, or if I find a better way, then I will use that better way. In his abstract, Michael argues that one should concentrate, as an editor, on ensuring that an edition will 'survive' by concentrating on the 'content' and not on the 'behaviour' of an edition. I believe this formulation is false. First, the content of an edition—any edition—will become outdated, as the interests and concerns of readers change. Text, really, is not: content must alter as new readers come. In 100 years time, I doubt that anyone will be using the transcripts of the *Canterbury Tales*, into which we have put so much effort. In our transcription, we have not distinguished the different forms of each letter, but only the different spellings. In 100 years time, scholars will be interested in these different letter forms, and will want transcriptions which record them. Our transcripts will be outdated and of no interest to anyone except the occasional digger into archives.

There is a second reason why this formulation, that editions will survive because of their content alone, is false. One of the factors that has most shaped the TEI is its roots in the electronic text archive world. TEI rhetoric, and the rhetoric of Michael's formulation, implies that all the editor has to do is capture the 'content' of the primary source, encode it in bullet-proof TEI-approved SGML and place this in an electronic archive where devoted acolytes will preserve it through new generations of computer equipment in return for large sums of public money. This cosy formula has left out one rather important factor: the reader. Let me say it as clearly as possible. Editions do not survive because they are preserved in elegant encoding and in government-maintained electronic archives. They survive because they are read. They survive because people find them useful, they survive because scholars, students and schoolchildren find they help to them read. Our aim over the next decades is to transform the way people read the *Canterbury Tales*. We want readers to understand just what it is they are reading, so that they can make texts for themselves with new intelligence. If they do this, then our text will be outdated: and frankly, it will not matter then if people can no longer read our text, and I will not care if they cannot.

The great promise of electronic editions, to me, is not that we will find new ways of storing vast amounts of information. It is that we will find new ways of presenting this to readers, so that they may be better readers. To do this, we will have to teach our editions to swim to the readers. Over the last few years, some of our foremost textual scholars have spent their time puzzling over whether we should use this tag or that. This is dry-land swimming and does not get us very far. The water is fine. Come on in; start swimming, let us make some real editions for real readers.

Notes

1 The arguments put forward in this article have been shaped in discussions with the participants in the 1997 Conference on Editorial Problems conference, with my colleagues in the Canterbury Tales Project over many years, and through repeated encounters with partners in the enterprise of making scholarly publications for this new age. Several of these debts will be clear to the reader of the article; but one particular debt I must pay is to Allen Renear: a long discussion with him in 1998 provided the germ for the view of editors and 'truth' which I here put forward. For the protection of all those who have helped me struggle with these ideas I emphasize that the responsibility for what is here argued is my own. This article was substantially written in 1999: all relative temporal references ('Ten years ago', etc.) should be referred to that time.

2 There is already a flourishing literature surrounding electronic texts. Examples include **Finneran, R. F.** (ed.) (1996). *The Literary Text in the Digital Age*. Michigan: University of Michigan Press; **Landow, G. P.** (1992). *Hypertext: The Convergence of Contemporary Critical Theory and Technology*. Baltimore: Johns

Hopkins Press; **Landow, G. P. and Delaney, P**. (eds) (1991). *Hypermedia and Literary Studies*. Cambridge, MA: MIT Press; **Landow, G. P. and Delaney, P**. (eds) (1993). *The Digital Word: Text-Based Computing in the Humanities*. Cambridge, MA: MIT Press; and **Sutherland, K**. (ed.) (1997). *Electronic Text: Investigations in Method and Theory*. Oxford: Clarendon Press. I too am guilty of using a sentence beginning 'We are undergoing the greatest change in publishing since . . .,' but can claim the slight originality of finishing the sentence with 'Aldus Manutius' and not 'Johann Gutenberg' [in my essay New Directions in Critical Editing. In Sutherland, K. (ed.), *Electronic Text*].

3 For example: the twelfth annual conference focused on the editing of medieval texts: see **Rigg, A. G**. (ed.) (1977). *Editing Medieval Texts*. New York and London: Garland. Topics addressed by other conferences include: the editing of sixteenth-century texts; eighteenth-century texts; Renaissance dramatic texts; and the relation between editors, authors and publishers.

4 By 'electronic edition' I mean an edition conceived and executed exclusively for electronic publication, and impossible in any other form. This is not just a difference in quantity (an electronic edition is larger) but in kind (an electronic edition does things one could not do in print: for example, amalgamate full-text transcripts, images and tools for research into all this). By this test, **Robinson, P.** (1996). *Wife of Bath's Prologue on CD-ROM*. Cambridge: Cambridge University Press; **McDermott, A.** (1996). *Johnson's Dictionary on CD-ROM*. Cambridge: Cambridge University Press and **Huitfeldt, C.** (2001). *Wittgenstein's Nachlass: The Bergen Electronic Edition*. Oxford: Oxford University Press and the Wittgenstein Archive are all electronic editions. Ventures such as the Chadwyck-Healey publications, or the Chaucer, Woolf, Johnson and Boswell productions from Primary Source Media, or the Cambridge University Press Ruskin, are essentially electronic republications of printed material (what one might call 'digital microfilm') and so not electronic editions.

5 Something very like these electronic editions, permitting books and manuscripts from the whole world to be brought to the reader in an electronic virtual library, was described by Ted Nelson in his *Literary Machines* (for example, the 1990 printing, published at Swathmore by Mindfull Press). It may be symptomatic of something in Nelson's vision that the book itself and its publication details are as evanescent and fugitive as we sometimes fear electronic texts themselves may be [see, the discussion in **Jakob N.** (1995). *Multimedia and Hypertext: The Internet and Beyond*. Boston: AP Professional, p. 371]. It is notable that while Susan Hockey describes many uses of computers in the making of scholarly texts [**Hockey, S.** (1980) *Computer Applications in the Humanities*. London: Duckworth], there is no discussion in this book of the possible electronic publication of these texts themselves: these tools are described as aids for the making of printed texts, which is exactly how they were then used. The focus of the first edition of **Shillingsburg, P.** (1986). *Scholarly Editing in the Computer Age*. Athens and London: University of Georgia, was similarly on the use of computers in the making of printed scholarly editions; one can track the approach of electronic editions themselves through Shillingsburg's later writings, not least his contribution to the 1997 conference.

6 Of particular reference to scholarly editors are the chapters on transcription of primary sources and on editorial apparatus in the Text Encoding Initiative Guidelines; certain of the 'core' tags are also crucial to successful editorial work. See **Sperberg-McQueen, M. and Burnard, L.** (1994). *Guidelines for Electronic Text Encoding and Interchange*, 2 Vols, Chicago and Oxford: Text Encoding Initiative. One can identify the sections for which I was responsible by the incidence of quotations from Chaucer in the examples in the TEI Guidelines.

7 Some of these 'pre-TEI' systems of encoding are described in **Robinson, P.** (1994). *Transcription of Primary Sources using SGML*. Oxford: Office for Humanities Communication.

8 See **McGillivray, M.** (1993). Electronic Representation of Chaucer Manuscripts: Possibilities and Limitations. In Lancashire,I. (ed.), *Computer-based Chaucer Studies*. Toronto: University of Toronto Press, pp. 1–16.

9 **Robinson, P. M. W. and Solopova, E.** (1993/1997). Guidelines for Transcription of the Manuscripts of the Wife of Bath's Prologue. In Blake, N. F. and Robinson, P. M. W. (eds), *The Canterbury Tales Project: Occasional Papers*, vol. 1. Oxford: Office for Humanities Communication, pp. 19–52, at p. 21.

10 Ibid, pp. 19, 21.

11 **Pichler, A.** (1995). Advantages of a Machine-Readable Version of Wittgenstein's Nachlass. In Johannessen, K. and Nordenstam, T. (eds), *Culture and Value: Philosophy and the Cultural Sciences*. Vienna: Austrian Ludwig Wittgenstein Society, p. 774b.

12 Ibid, p. 690.

13 **DeRose, S. J., Durand, D. G., Mylonas, E., and Renear A. H.** (1990). What is text, really? *Journal of Computing in Higher Education*, 1.2 (1990): 3–26.

14 The paper was 'published' to the 'MII-PESP: Philosophy and Electronic Publishing' discussion group on 27 November 1995. The discussion group was established as part of a paper 'Philosophy and Electronic Publishing', organized by Claus Huitfeldt for publication in an interactive issue of the journal *The Monist*, published as volume 80, no. 3, 1997: see http://www.univie.ac.at/philosophie/bureau/intro.htm. Some of Renear's arguments are also presented in his article ''Three (Meta)Theories of Textuality'' In Sutherland, K. (ed.) *Electronic Text*, pp. 107–26.

15 For Renear's argument, compare Wittgenstein's criticism of Bertrand Russell's account of meaning as 'causal': thus, a word is used correctly 'when the average hearer will be affected by it in the way intended' [**Russell, B.** (1921) *Analysis of Mind*. London: Unwin, p. 198]. To this Wittgenstein retorts 'If I wanted to eat an apple, and someone punched me in the stomach, taking away my appetite, then it was this punch that I originally wanted' [**Wittgenstein, L.** (1975) In Rhees, R. (ed.) *Philosophical Remarks*London: Blackwell, p. 64; reported in **Monk, R.** (1990). *Ludwig Wittgenstein: The Duty of Genius*. London: Jonathan Cape, p. 291]. Thus, there must be a consonance between the original need and the final result (that is, both reflect the same 'reality'); or the text must have been originally written to allow the transcriber to win the prize. The second possibility is patently absurd; therefore there must be the one 'reality' underlying both text and transcription.

16 This is fragment III in Charles Kahn's edition of Heraclitus [**Kahn, C.** (1979) *The Art and Thought of Heraclitus: An Edition of the Fragments with Translation and Commentary*. Cambridge: Cambridge University Press, p. 28], corresponding to fragment two in Diels–Kranz. Kahn translates this as 'Although the account [logos] is shared, most men live as though their thinking were a private possession'. Kahn explains 'logos' as 'not simply language but rational discussion, calculation and choice: rationality as expressed in speech; in thought; and in action.'

17 A full discussion of our findings for The Wife of Bath's Prologue is printed in Robinson, P. M. W. A Stemmatic Analysis of the Fifteenth-century Witnesses to the Wife of Bath's Prologue. In Blake and Robinson (eds), *The Canterbury Tales Project: Occasional Papers*, vol. 2., pp. 69–132; and Robinson, P. M. W. The Problem of Authorial Variants in the Wife of Bath's Prologue. Blake and Robinson (eds), *The Canterbury Tales Project: Occasional Papers*, vol. 2, pp. 133–42. The discussion of The General Prologue tradition is in **Robinson, P.** (2000). 'Analysis workshop' section. In Solopova, E. (ed.), *The General Prologue of The Canterbury Tales on CD-ROM*. Cambridge: Cambridge University Press. This section also provides the tools themselves we have used for the analysis, and many examples of their use.

18 See Robinson, Analysis Workshop In *The General Prologue on CD-ROM*.

19 **Robinson, P.** (1996). A Stemmatic Analysis of the Fifteenth-Century Witnesses, p. 126; and Computer-Assisted Stemmatic Analysis and ''Best-Text'' Historical Editing. In Van Reenen, P. and Van Mulken, M. (eds), *Studies in Stemmatology*. Amsterdam: John Benjamins, pp. 71–103.

20 See **Robinson, P. M. W.** (1999). Can we trust Hengwrt? In Lester, G. A. (ed.), *Chaucer in Perspective: Middle English Essays in Honour of Norman Blake*. Sheffield: Sheffield Academic Press.

21 See **Solopova, E.** (2001). The Survival of Chaucer's Punctuation in the Early Manuscripts of the *Canterbury Tales*. In Minnis, A. J. (ed.), *Middle English Poetry: Texts and Traditions.* York: Boydell and Brewer; and Chaucer's metre and scribal editing in the early manuscripts of the *Canterbury Tales*. Blake and Robinson (eds), *The Canterbury Tales Project: Occasional Papers*, vol. 2, pp. 143–64.