

Scholarly editions: does form matter?

Federico Meschini

1. Medium, Ecdotique and Textual Criticism

In 1975 Aurelio Roncaglia, one of the most important Italian philologists, in his *Principi ed applicazioni di Critica Testuale*¹, wrote that textual criticism is the central core of *ecdotique*², a larger discipline which presides over text editions. Therefore *ecdotique* also includes publishing and pragmatic issues, such as page and font layout or publication structure. If we make this assumption, it follows that, more than thirty years after Roncaglia's writing, the marriage, or at least the cohabitation, between electronic publishing and textual scholarship is not only a technical issue, but also and mostly a legitimate concern for humanities and should be made under *ecdotique*'s approval.

But what if textual criticism is agnostic about its medium, both not interested and overall not influenced by the way its results will be disseminated and preserved? Is textual criticism only a subset of *ecdotique*, even if a fundamental one? This seems to be the influential position of G. Thomas Tanselle in his foreword to the TEI/MLA Volume *Electronic Textual Editing*³ clearly in contrast with the well-know, and very often misused McLuhan's slogan about the medium and the message or Nietzsche's aphorism of the gaze into the abyss⁴.

Perhaps using holistic and systemic thinking can help us to put the question in another light. If textual criticism is an element of *ecdotique*, it is very improbable, that it would not interact, influence and be influenced by the other elements of this system, including the ones about technical aspects. The question is how much and in what way? Do the product characteristics influence the process methods, or is it the other way round?

Surely Tanselle is partly right in claiming some independence from the medium, both as an object and as a product of textual criticism, and the preponderance of substance over form. It goes without

¹ Aurelio Roncaglia, *Principi e applicazioni di critica testuale*, Bulzoni, Roma 1975, p. 26.

² Word introduced by Dom H. Quentin in his *Essais de critique textuelle (Ecdotique)*, Paris 1926.

³ Lou Burnard, Katherine O'Brien O'Keeffe, John Unsworth, *Electronic Textual Editing*, Modern Language Association of America 2006 <<http://www.tei-c.org/Activities/E/TE/Preview/index.xml>>.

⁴ In his Society for Textual Scholarship inaugural address he recognized that textual criticism is part of a larger system, since he said that it is "but one branch of textual scholarship as a whole" (G. T. Tanselle, 'Presidential Address', *TEXT* 1(1984)) and later he also said that it is part of the "single great enterprise" together with palaeography, codicology and bibliography (G. T. Tanselle, *Rationale of Textual Criticism*, University of Pennsylvania Press, 1989, p. 46). But when hinting at the existence of this system, he always avoids any allusion to the presence of more, but not only, technical-oriented disciplines.

saying that textual criticism is about ‘texts’, creations of human ingenuity made up by a particular kind of information which is composed by a set of graphical signs transcoding human language in a fixed state following some particular rules, and very different from other kinds of works, made up by other information types, like images or music. Tanselle’s use of the terms ‘substance’ and ‘form’ bring us into the world of Linguistics and Semiotics, where every sign system is composed of two different levels, substance (*signified*) and form (*signifier*). Even if substance could be predominant over form, the latter is certainly not without any meaning or influence. Perhaps we should remember the important role ‘accidentals’ had in W. W. Greg’s ‘Rationale of Copy-Text’⁵. In this case accidental do not means what happens by chance, but a formal feature (an accent) which is non-essential to the author’s meaning, the substance, and therefore is mostly about form. Furthermore, non-essential does not automatically mean non-relevant or non-influential.

Moreover, for the Danish linguist Hjelmslev both substance and form can be split in other two levels: content and expression. There is a core of textual criticism which is about the content substance, and it does not care if what it is dealing with is carved on the stone, written or typed with ink, or encoded using bit and bytes. It only cares whether what it is dealing with could be considered as ‘text’. This core could be surely considered as medium-proof. But what about the other levels, what about both the expression substance and the form, both content and expression, of textual works? For this reasons, my opinion is that the more we go up in layers which are above the content substance, the more the chosen medium has an active and influential role in the processes and products of textual criticism, and is not only a separated, static and isolated component in the discipline of *ecdotique*.

2. Textual Criticism and Computer Science

The relationship between textual criticism and computer science could seem an odd one at first, but they have many things in common, since they are disciplines which are both at the same time theoretical and practical, with these two aspects continually influencing each other and thus generating an epistemological virtuous circle.

Saint Isidore, bishop of Seville during the seventh century, could be considered the first point of connection between these two disciplines, since among his many skills he was also a biblical

⁵ W. W. Greg, “Rationale of Copy-Text” in *Studies in Bibliography*, 3(1), pp 19-37
<<http://etext.virginia.edu/etcbin/ot2www-sb?specfile=/texts/english/bibliog/SB/index/studbib.o2w&act=text&offset=1070642&textreg=1&query=>>.

exegetical scholar and his *Etymologiae*⁶ is recognized as the world's first database ever produced, conveying and organizing the knowledge of his time, being therefore similar to the current *WorldWideWeb*⁷.

In more recent times and also on a more concrete base, Dom H. Quentin can be considered the real precursor of computer science applications in textual criticism, since in 1926 he proposed a method for the evaluation of the variants based on a quantitative approach together with statistical elaborations to determine the existing relationships among the different texts, thereby leaving behind the traditional qualitative approach. The effectiveness of this method can be contested, but what really matters is the cultural and epistemological shift in preferring a mechanical method to a more subjective one. It goes without saying that in the twenties Dom Quentin did not have a computer at his disposal and therefore his manual system was translated into a real computer algorithm⁸ only in the late sixties, by Dom Jacques Froger⁹ and Gian Piero Zarri¹⁰.

But the real and official marriage between humanities and computing took place more than twenty years before this date (and exactly twenty years after Quentin's proposal): in 1946, just after World War II, when Father Roberto Busa was allowed to use an IBM punchcard machine to create his *Index Thomisticus* a lemmatization of the whole works of St. Thomas Aquinas, a pantagruelic effort which could be made possible by a single man only by the use of an automatic machine¹¹. Even if Busa's work was more in the field of Linguistics, being also the origin of the branch known as Computational Linguistics, it is also considered as the official birth of "Humanities Computing".

Humanities computing, or as it has been called more recently Digital Humanities, is considered by its supporters as a scientific discipline. This means that the application of computing to humanities is not only a mere technical issue, that has as its main aim to facilitate and speed up the scholar's work, but is a crossing that generates a whole new field of study, with its own theoretical and methodological issues¹². Being a specialized sector of Digital Humanities, electronic textual editing presents the same issue.

⁶ A digital copy is available in the virtual library of the Codices Electronici Sangallenses Project <<http://www.cesg.unifr.ch/>>.

⁷ For this reason he has been nominated as the patron saint of Internet and computer programmers.

⁸ An algorithm is a finite list of operations to solve a particular problem. See D. Harel, *Algorithmics The Spirit of Computing*, Addison-Wesley, 1992.

⁹ J. Froger, *La critique des textes et son automatisation*, Dunod, Paris, 1968.

¹⁰ G. P. Zarri, *Linguistica algoritmica e meccanizzazione della "collatio codicum"*, in "Lingua e stile", III, 3, pp.21-40, 1968.

¹¹ Between 1974 and 1980, 56 volumes of the *Index Thomisticus Sancti Thomae Aquinatis Operum Omnium Indices ed concordantiae* were produced. In 1989 a version on CD-Rom was made available, and in 2005, was also published online <<http://www.corpusthomaticum.org/>>.

¹² S. Schreibman, R. Siemens, J. Unsworth (eds.), *A Companion to Digital Humanities*, Blackwell Publishing, 2005 <<http://www.digitalhumanities.org/companion/>>

About this topic I would like to quote both Thomas Pynchon, who once wrote: “You know what a miracle is [...] But another world’s intrusion into this one”¹³ and J. F. Lyotard, who in ‘The Postmodern Condition’ stated that in a game with full or sufficient information, like our society, a better performance is obtained when two different kinds of information are put in a new relation, thus generating new knowledge.

For almost ten years after the conversion of Quentin’s method in an algorithm, the computer has, and in many cases still is, been considered as an autistic worker, only capable of saving the editor from some boring and repetitive work, without questioning the epistemological asset of the textual scholar’s work. It was only in 1978, during the international conference in Paris *La pratique des ordinateurs dans la critique des textes*¹⁴, that the computer started being considered not only as a solver of already known problems but also as a creator of new questions, thanks also to Cesare Segre who introduced the notion of text as a “diasystem”: if Text is as an interaction between two or more systems, no single text can be considered as the Text, and it is not possible to speak about the Text without considering its complex transmission line¹⁵.

The traditional scholarly activities of *collatio*, *emendatio* or *divinatio* are influenced by this paradigmatic and epistemological shift and new procedures have to be established. Moreover the two extremes of the scholarly process acquire a new dimension and importance: the transcription of the textual artefacts, which becomes a structural and semantic encoding, and the dissemination modalities, or in other words how the edition is initially assembled and subsequently published.

During the eighties a lot of theoretical studies about this subject were developed, granting it an academic dignity. At least three cases should be quoted: the publication of Peter Shillingsburg’s *Scholarly Editing in the Computer Age*¹⁶, together with the development of the software suite CASE¹; Hans Walter Gabler’s edition of James Joyce’s *Ulysses*, which made use of another piece of textual scholarly software, TUSTEP¹⁷; and the establishment in 1987 of the Text Encoding Initiative¹⁸, an international organization whose main aim is the creation of a standard for the electronic representation of literary texts. Nevertheless, the print medium was still the only proper solution for the publication of a critical edition, due to the lack of suitable supports, since the CD-Rom was still in its early development and the floppy-disk was too limited.

¹³ Thomas Pynchon, *The Crying of Lot 49*, Bantam, 1967.

¹⁴ J. Glennison, R. Marichal, J. Irigoin (eds.) *La Pratique des ordinateurs dans la critique des textes*, CNRS, 1979

¹⁵ C. Segre, *Critica testuale, teoria degli insiemi e diasistea*, in *Semiotica filologica. Testo e modelli culturali*, Torino Einaudi, 1979, pp. 53-64.

¹⁶ P. Shillingsburg, *Scholarly Editing in the Computer Age: Theory and Practice*, University of Georgia Press, 176.

¹⁷ Tübingen System of TExt processing Programs <http://www.zdv.uni-tuebingen.de/tustep/tustep_eng.html>.

¹⁸ <<http://www.tei-c.org/>>.

In the nineties this field of study continued to grow, with some important publications such as Jerome McGann's 'The Rationale of HyperText'¹⁹, the creation of research groups, involved both in theoretical and practical activities, such as the *Institute for Advanced Technology in the Humanities*²⁰ at the University of Virginia, or the informal Italian school of Humanities Computing.

The breakthrough of that decade, which also influenced the developments listed above, was the birth and the diffusion of digital media which could be used as real alternatives to the print for conveying scholarly editions of literary work. First, the CD-Rom, but above all the creation of the *WorldWideWeb*, a new infrastructure based on existing technologies which allows the publication of electronic texts accessible from every computer connected to this network.

Thanks to these premises, the new millennium is still witnessing an unstoppable expansion of scholarly editions in electronic forms. They are not anymore the subject of niche conferences and journals, as in the preceding decades, but they are talked about and proposed as a solution to overcome the natural limits of print editions even in more traditional circumstances.

What seems to be the start of an age of gold for electronic editions should sound more like a warning, because it is not solution in itself, but a new medium which entails as many problems as possibilities.

3. Digital texts and electronic editions

E.T.A Hoffman's *Die Brautwahl* (The Choice of a Bride), which was published in the early nineteenth century, tells the story of three suitors of Albertine, a young and beautiful girl. Of course Albertine will marry only one of the suitors, but the other two will receive in return a gift which would be for them more precious than the girl's hand. Therefore Chancellor Tusmann, an old bibliophile who is clearly not suitable for the young girl, is more than happy when, even if rejected by Albertine, he receives a magic book. All the book pages are blank, but they can change at will, becoming whatever book the chancellor desires to read.

It is evident that the chancellor's book is the fantasy version of a computer screen, or eventually more akin to an ebook reading device. Therefore the magic features which are incredibly valuable for the old bibliophile are embedded in the very nature of electronic texts. But then why is everyone who loves or works with books at least suspicious of, if not adverse, towards the digital counterpart of books?

¹⁹ <<http://www.iath.virginia.edu/public/jjm2f/rationale.html>>.

²⁰ IATH <<http://www.iath.virginia.edu/>>. It is the home of many important projects, such as *The Complete Writings and Pictures of Dante Gabriel Rossetti* <<http://www.rossettiarchive.org/>>, *The William Blake Archive* <<http://www.blakearchive.org/>>, *Piers Plowman Electronic Archive* <<http://www.iath.virginia.edu/seenet/piers/>> and *Dickinson Electronic Archives* <<http://www.emilydickinson.org/>>.

Some cultural resistances are physiological, since with the introduction of a new medium there have always been a phase of opposition to and imitation of the old one.

About this issue there is a deep and at the same time paradoxical connection between printed and electronic texts. Electronic texts currently use the paradigm of printed ones; therefore we have Word, PDF and HTML documents (even if HTML pages look more like scrolls), while at the same time the current workflow of printed text production is mostly digital, since they are written, edited, and formatted using digital tools, such as word processors or Desk Top Publishing software. It is only in the last step that they leave the world of bits to embody a fixed state made of ink and paper.

In the physical world, libraries, archives and books are very different entities, even if with very strong relations between them. Printed scholarly editions, a particular kind of book, are created using archival and research materials and are preserved in libraries. In the digital world the differences and relationships between digital libraries, electronic archives, electronic books and scholarly editions are at the same time overlapping and blending. Moreover these terms are very often used in an improperly interchangeable way.

But what is so special about electronic editions? Why do they intrigue so much both young and experienced scholars, who, when not intrigued, are at least concerned about them?

This can be tautological but a literary work, a text, is not a book, a book is only the top of the iceberg, the most visible and tangible part of what instead is a complex and layered system. The composition and evolutionary process of a literary text recalls Frost's Road not taken or Borges' Garden of Forking Paths, being, at every step in this process, a choice of a single route in a maze with both endless entries and exits. What a single book represents and embodies is only one of these routes, a fixed and synchronic state of a textual system which also has a diachronic axis. The role of a critical edition is therefore to overcome this limit, to represent as many routes as possible. This is done through the work of the editor, who gathers as much as possible of the scattered bits of information about the text, not only from its varied physical manifestation but from related documentary historical evidence. Then he or she tries to organize this information through the application of some specific scholarly methods and the imposition of a structure. The result is the creation of a critical edition, from which new knowledge about the text can be inferred.

But when the critical edition has a printed form there is clearly an ontological contradiction involved, since we are using the book form to overcome the limits of the book itself. Given this situation the largely adopted "copy-text plus apparatus" structure is very likely the best possible approach, but the problem is not solved, but only put in the background.

Can an electronic edition solve this Gordian knot, and if yes, why and in which way? Before answering this question we should think a little about the nature of an electronic edition and its

differences compared to a printed one. As it should be clear now a critical digital edition is not a mere transformation of a print edition into computer-readable file, but it is something which in some way extends the paradigm of critical editing itself and can produce a deeper understanding of the textual work²¹ Electronic editions have the innate potentiality of conveying more knowledge (both on a quantitative and qualitative base) than a printed one. It's no wonder that Cazalé and Mordenti spoke about "Editio sapiens" (wise edition), Paolo D'Iorio of "Hypertexte savant"²² and Peter Shillingsburg about "Electronic Knowledge Sites"²³. But why? What really is an electronic edition? A good metaphor is the old story about the six blind people and the elephant, where each of them thinks that it is something different: a snake, a rope, a pillar, a spear, because none can see the elephant as a whole: the judgment of each depends on the part of the elephant's body he has touched.

The printed medium is a technology for structuring and conveying information. These characteristics apply also to the electronic medium which has also a distinguishing feature: it is able to process and to work on information. In my opinion this is the crux of the difference between these two media, which should always be considered when thinking, talking and writing about them. An electronic edition can be a mere representation of a text, and this can be considered its degree zero, but it can, or better, it *should* also be a software application, able to process data, responding to users' actions pro-actively and therefore producing new information.

This is therefore the first main ingredient of electronic editions: the operating logic, what is possible or not to do in that edition, starting from the visual display to basic hyper-textual browsing, full-text search or multi-layer annotation. The basic building blocks of such an edition are composed of raw data, which can be made by one of the four possible kinds of digital information: text, image, audio and video. So we can have the transcription of a text together with the digitized page images or an audio version of the book. But an intermediate layer is needed between the operating logic and the raw data, and this role is carried out by structured data. This could be of two kinds: raw data transformed in structured files (such as an encoded text file) or metadata which impose a structure over the basic raw files (such as the relationship between the digitized images and the order of the pages or among the different versions of the same work).

Finally there is the user interface, which allows the interaction with the edition's features and its content.

²¹ P. Robinson, *What Is a Critical Digital Edition? Variants 1* (2002), pp. 43-62.

²² P. D'Iorio, *HyperNietzsche. Modèle d'un hypertexte savant sur Internet pour la recherche en sciences humaines. Questions philosophiques, problèmes juridiques, outils informatiques*. Presses Universitaires de France, 2000
<<http://www.hypernietzsche.org/doc/puf/>>.

²³ P. Shillingsburg, *From Gutenberg to Google*, Cambridge University Press, pp. 80-102.

What I have proposed here is only an overview, which at the same time imposes a pattern but also simplifies the architecture of an electronic edition.

The last thing I would like to discuss the “incompatibility principle” of electronic editions. While in the digital library world the term interoperability is considered as a mantra, in most cases electronic scholarly editions are architecturally dysmorphic from each other. This happens for several reasons. First, a digital library usually is based on a collection level, grouping digital items with a heterogeneous nature, both in term of content and form, while an edition has a more granular level, allowing deeper penetration and manipulation of the single item. Second, every digital edition adopts a particular inner logic (ontology or *Weltanschauung*) with characteristics influenced by many factors: the nature of the primary materials being edited, the scholarly vision of the editor together with the vision of the technical manager and the interaction between these two. It is almost natural that every edition is a microcosm of its own.

If, as I have argued, the electronic medium is very different from print the question remains whether this means “good different or bad different?”, as David Tennant asked Billie Piper, shortly after his regeneration as the new Doctor Who. In the present case also the answer is “Just... different”. Even if electronic publishing has a lot of potentialities they all come with a price, and the creation of an electronic edition is a very complex process.

But as scholars we are bound to three main tasks: collecting information, pondering this information and making judgements. Therefore what should be our judgement about electronic scholarly editions? The Bible, one of the most, if not the most important text in the history of western textual criticism, could be helpful here: “Judge not according to the appearance, but judge righteous judgment”²⁴ and “Ye shall know them by their fruits.”²⁵

²⁴ John 7:24.

²⁵ Matthew 7:15.