

ELECTRONIC  
TEXTUAL  
EDITING

---

---

EDITED BY  
*Lou Burnard,*  
*Katherine O'Brien O'Keeffe,*  
AND  
*John Unsworth*

---

---

THE MODERN LANGUAGE  
ASSOCIATION OF AMERICA  
NEW YORK  
2006

# CRITICAL EDITING IN A DIGITAL HORIZON

---

---

DINO BUZZETTI AND JEROME MCGANN

Just as the machinery of the codex opened radically new ways to store, organize, study, transform, and disseminate knowledge and information, digital technology represents an epochal watershed for anyone involved with semiotic materials. For scholars of books and texts, and in particular for editorial scholars, digital tools have already begun their disciplinary transformations, and we can see as well the promise of further, perhaps even more remarkable, changes on the near horizon of our work.

In this essay, we describe and reflect on these changes, but first we briefly review the forms and procedures of scholarly editing that are licensed by codex technology. This survey is important because present work and future developments in digital scholarship evolve from critical models that we have inherited. The basic procedures and goals of scholarly editing will not change because of digital technology. True, the scale, range, and diversity of materials that can be subjected to scholarly formalization and analysis are all vastly augmented by these new tools. Besides, the emergence of born-digital artifacts creates entirely new critical opportunities, as well as problems, for librarians, archivists, and anyone interested in the study and interpretation of works of culture. Nonetheless, the goals of the scholar remain unaltered—preservation, access, dissemination, and analysis-interpretation—as does the basic critical method, formalization.

If our traditional goals remain, however, these new technologies are forcing us to revisit and rethink some of the most basic problems of textuality and theory of text. We address these matters in the two central sections of this essay, and in the final section we reflect on certain practical methodological implications of these reflections. First, however, we must step back

and make a brief review of the current state of text-editing theory and method.

## CODEX-BASED SCHOLARSHIP AND CRITICISM

Scholarly editing is the source and end and test of every type of investigative and interpretational activity that critical minds may choose to undertake.<sup>1</sup> Well understood by scholars until fairly recently, the foundational status of editorial work is now much less surely perceived. Hermeneuts of every kind regularly regard such work, in René Wellek's notoriously misguided description, as "preliminary operations" in literary studies (Wellek and Warren 57). Odd though it may seem, that view is widely shared even by bibliographers and editors, who often embrace a positivist conception of their work.

Scholarly editing is grounded in two procedural models: facsimile editing of individual documents and critical editing of a set of related documentary witnesses. In the first case, the scholar's object is to provide as accurate a simulation of some particular document as the means of reproduction allow. Various kinds of documentary simulation are possible, from digital images and photoduplications on one end to printed diplomatic transcriptions on the other. Facsimile editing is sometimes imagined as a relatively straightforward and even simple scholarly project, but in fact the facsimile editor's task is every bit as complex and demanding as the critical editor's. In certain respects it can be more difficult, precisely because of the illusions that come with the presence of a single documentary witness, which can appear as a simple, self-transparent, and self-identical object.<sup>2</sup>

Securing a clear and thorough facsimile brings with it more problems than the manifest and immediate technical ones, though they are real enough. In addition, the facsimile editor can never forget that the edition being made comes at a certain place and time. At best, therefore, the edition is an effort to simulate the document at that chosen moment. The document bears within itself the evidence of its life and provenance, but that evidence, because of the document's historical passage, will always be more or less obscure, ambiguous in meaning, or even unrecoverable.

Every document exhibits this kind of dynamic quality, and a good scholarly edition will seek to expose that volatility as fully as possible. Being clear about the dynamic character of a document is the beginning of scholarly wisdom, whatever type of work one may undertake (hermeneutical or editorial) and—in editorial work—whatever type of edition one has chosen to do.

The other foundational pole (or pillar) of scholarly editing is critical editing. This work centers in the comparative analysis of a set of documentary witnesses each of which instantiates some form or state of the work in question. We name, for example, Dante Gabriel Rossetti's "The Blessed Damozel" with that one name, as if it were a single, self-identical thing (which, in that special perspective, it in fact is—that is to say, is taken to be). But the work so named descends to us in multiple documentary forms. Critical editing involves the careful study of that documentary corpus. Its main purpose is to make various kinds of clarifying distinctions among the enormous number of textual witnesses that instantiate a certain named work.

The critical editor's working premise is that textual transmission involves a series of translations. Works get passed on by being reproduced in fresh documentary forms. This process of reproduction necessarily involves textual changes of various kinds, including changes that obscure and corrupt earlier textual forms. Some of these changes are made deliberately, many others not. A classical model of critical editing, therefore, has involved the effort to distinguish the corruptions that have entered the body of the work as a result of its transmission history. That model often postulates a single, authoritative, original state of the work. The scholar's analytic procedures are bent on the effort to recover the text of that presumably pristine original.

A key device for pursuing such a goal is stemmatic analysis. This is a procedure by which the evolutionary descent of the many textual witnesses is arranged in specific lines. A stemma of documents exposes, simply, which texts were copied from which texts. Understanding the lines of textual transmission supplies a scholar with information that guides and controls the editorial work when decisions have to be made between variant forms of the text.

The problematic character of every documentary witness remains a key issue for the critical editor. That difficulty emerges at the initial stage of editorial work—that is to say, at the point when a decision is made about which documents will come into the textual analysis. In no case can all the witnesses be included. On one hand, the number of actually available documents will be far too numerous; on the other, many documents that formed part of the transmission history will be inaccessible.

In some cases—they are a distinct minority—a relatively small and manageable set of documents offers itself for scholarly analysis. Print technology brought about a massive proliferation of textual works. These are passed on to us edition by edition, and of course each edition differs from every other, nor are the differences between editions always easy to see or understand.

But the play of these kinds of textual differences is still more extreme. An unschooled view, for example, will assume that every copy of a print edition of some work is identical to every other copy. Editorial scholars themselves often make this assumption, and sometimes deliberately (in order to simplify, for analytic purposes, the ordering of the editorial materials). But the textual scholar usually knows better, and in producing a critical edition from an analysis of printed documents, editors regularly understand that multiple copies of a single edition must be examined. (As we see below, even multiple copies that appear to be textually identical, always incorporate material differences that can be, from the scholar's point of view, crucial for anyone trying to understand the work in question.)

In recent years a special type of critical editing gained wide currency: the so-called eclectic editing procedure, promoted especially by Fredson Bowers. This method chooses a copy-text as the editor's point of departure. The editor then corrects (or, more strictly, changes) this text on the basis of a comparative study of the available readings in the witnesses that are judged to be authoritative by the editor. When variant readings appear equally authoritative, the editor uses judgment to choose between them.

In considering these matters, the scholar must never lose sight of the fundamentally volatile character of the textual condition. The pursuit of a correct or authoritative text is what a poet might call "a hopeless flight" (Byron 3.70.5). Editors can only work to the best of their judgment, for the texts remain, in the last analysis, ambiguous. One sees how and why by reflecting on, for example, the editorial commitment to achieve an authoritative text—which is to say, a text that represents the author's intention. That pervasive editorial concept is fraught with difficulties. Authors regularly change their works, so that one often must wrestle with multiple intentions. Which intentions are the most authoritative? first intentions? intermediate? final? Are we certain that we know each line of intentionality, or that we can clearly distinguish one from another? For that matter, how do we deal with those textual features and formations that come about through nonauthorial agencies like publishers? In respect to the idea of textual authority, more authorities sit at the textual table than the author.

Scholars have responded to that textual condition with a number of interesting, specialized procedures. Three important variations on the two basic approaches to scholarly editing are especially common: best-text editions, genetic editions, and editions with multiple versions. The best-text edition aims to generate a reproduction of the particular text of a certain work—let's say, the Hengwrt manuscript of Chaucer's *Canterbury Tales*—that

will be cleared of its errors. Collation is used to locate and correct what are judged to be corrupt passages. Unlike the eclectic edition, a best-text edition does not seek to generate a heteroglot text but one that accurately represents the readings of the target document. If a facsimile or diplomatic approach to the editorial task is taken, editors will even preserve readings that they judge to be corrupted. If the approach is critical, editors will try to correct such errors and restore the text to what they judge to have been its (most) authoritative state.

Genetic editing procedures were developed in order to deal with the dynamic character of an author's manuscript texts. These editions examine and collate all the documents that form part of the process that brought a certain work into a certain state. Usually these editions aim to expose and trace the authorial process of composition to some point of its completion (for example, to the point where the text has been made ready for publication).<sup>3</sup>

Multiple version editions may take a best-text, an eclectic, or a genetic approach to their work. They aim, in any of these cases, to present multiple-reading versions of some particular work. (Paul Baum edited Rossetti's "The Blessed Damozel" in this way, and Wordsworth's *The Prelude* as well as Coleridge's "The Rime of the Ancient Mariner" have regularly been treated in a versioning editorial approach.)<sup>4</sup>

Finally, we should mention the proposal for social text editing that was especially promoted in recent years by D. F. McKenzie. In McKenzie's view, the scholar's attention should be directed not only at the text—the linguistic features of a document—but at the entirety of the material character of the relevant witnesses. McKenzie regarded documents as complex semiotic fields that bear within themselves the evidence of their social emergence. The critical editor, in his view, should focus on that field of relations and not simply on the linguistic text. Unfortunately, McKenzie died before he could complete the project he had in mind to illustrate his editorial approach—his edition of William Congreve.

## TEXTUAL AND EDITORIAL SCHOLARSHIP WITH DIGITAL TOOLS

The advent of information technology in the last half of the twentieth century has transformed in major ways the terms in which editorial and textual studies are able to be conceived and conducted. This transformation has come about because the critical instrument for studying graphic and bibliographic

works, including textual works, is no longer the codex (see McGann, "Rationale"). Because the digital computer can simulate any material object or condition in a uniform electronic coding procedure, vast amounts of information that are contained in objects like books can be digitally transformed and stored for many different uses. In addition, information stored in different kinds of media—musical and pictorial information as well as textual and bibliographic information—can be gathered and translated into a uniform (digital) medium and of course can be broadcast electronically. We go online and access the card catalogs, and often the very holdings, of major research archives, museums, and libraries all over the world.

The implications of this situation for scholarly editing are especially remarkable. For example, one can now design and build scholarly editions that integrate the functions of the two great editorial models, the facsimile and the critical edition. In a codex framework these functions are integrated only at the level of the library or archive, so that comparative analysis—which is the basis of all scholarship—involves laborious transactions among many individuals separated in different ways and at various scales. A complete critical edition of the multimedia materials produced by figures like Rossetti, William Blake, or Robert Burns can be designed and built, and Shakespeare's work need no longer be critically treated in purely textual and linguistic terms but can be approached for what it was and still is: a set of theater documents. Digitization also overcomes the codex-enforced spatial limitations on the amount of material that can be uniformly gathered and re-presented. In short, digital tools permit one to conceive of an editorial environment incorporating materials of many different kinds that might be physically located anywhere.

The accessibility of these resources and the relative ease with which one can learn to make and use them have produced a volatile Internet environment. The Web is a petri dish for humanities sites devoted to every conceivable topic or figure or movement or event. Noncopyrighted texts are available everywhere as well as masses of commentary and associated information. And of course therein lies the problem, for scholarship and education demand disciplined work. Scholars commit themselves to developing and maintaining rigorous standards for critical procedures and critical outcomes. The truth about humanities on the Internet, however, is that tares are rampant among the wheat. Nor do we have in place as yet the institutions we need to organize and evaluate these materials. Those resources are slowly being developed, but in the meantime we have metastasis.

Here is an example of the kind of problem that now must be dealt with.

We have in mind not one of the thousands of slapdash, if also sometimes lively, Web sites that can be found with a simple *Google* search. We rather choose the widely used (and in fact very useful if also very expensive) *English Poetry Full-Text Database* (600–1900) developed and sold by Chadwyck-Healey. From a scholar's point of view, this work is primarily an electronic concordance for the authors and works in question. While its texts have been for the most part carefully proofed, they are nearly all noncopyrighted. The status of the source text therefore must be a primary concern for any but those whose use is of the most casual kind. Thomas Lovell Beddoes, for example, comes in the 1851 Pickering edition—an edition no scholar now would use except in the context of inquiries about Beddoes's reception history. In addition, although the database calls itself full-text, it is not. Prose materials in the books that served the database as copy-text are not part of the database. The prefaces, introductions, notes, appendixes, and so forth that accompany the poetry in the original books, and that are so often clearly an integral part of the poetry, have been removed.

Economic criteria largely determined the database's choice of texts (and, presumably, the removal of the prose materials). The decision brings certain advantages, however. The 1851 edition of Beddoes, for example, while not a rare book, is not common (the University of Virginia, which has strong nineteenth-century holdings, does not own a copy). The database is of course far from a complete collection of all books of poetry written, printed, or published between 600 and 1900, but it does contain the (poetical) texts of many books that are rare or difficult to find.

Two further important scholarly facts about the database. First, it is a proprietary work. This means that it does not lie open to Internet access, which would allow its materials to be integrated with other related materials. The database is thus an isolated work in a medium where interoperability—the capacity to create and manipulate relations among scattered and diverse types of materials—is the key function. Second (and along the same fault line), its texts can only be string-searched: they have not been editorially organized or marked up for structured search and analysis operations or for analytic integration with other materials in something like what has been imagined as a semantic web (see below).

Scholars whose work functions within the great protocols of the codex—one of the most amazing inventions of human ingenuity—appear to think that the construction of a Web site fairly defines digital scholarship in the humanities. This view responds to the power of Internet technology to make materials available to people who might not otherwise, for any number



of reasons, be able to access them. It registers as well the power of digitization to supply the user with multimedia materials. These increased accessibilities are indeed a great boon to everyone, not least of all to students of the humanities. But in a scholarly perspective, these digital functions continue to obscure the scholarly and educational opportunities that have been opened to us by the new technology.

Access to those opportunities requires one to become familiar with digital text representation procedures and in particular with how such materials can be marked and organized for formal analysis. That subject cannot be usefully engaged, however, without a comprehensive and adequate theory of textuality in general.

## MARKING AND STRUCTURING DIGITAL TEXT REPRESENTATIONS

Traditional text—printed, scripted, oral—is regularly taken, in its material instantiations, as self-identical and transparent. It is taken for what it appears to be: nonvolatile. In this view, volatility is seen as the outcome of an interpretive action on an otherwise fixed text. The inadequacy of this view, or theory, of text must be clearly grasped by scholars, and especially by scholars who wish to undertake that foundational act of criticism and interpretation, the making of a scholarly edition.

We may usefully begin to deconstruct this pervasive illusion about text by reflecting on the view of text held by a computer scientist, for whom text is “information coded as characters or sequences of characters” (Day 1). Coded information is data, and data is a processable material object. By processing data, we process the information it represents. But digital text, unlike the information it conveys, is not volatile. It is a physical thing residing in the memory cells of a computer in a completely disambiguated condition. That precise physical structure matters for digital text, just as precise physical structure, very different, matters for paper-based text. The digital form of the text defines it as an object on which computers can operate algorithmically to convey sense and information. A digital text is coded information, and a code has a syntax that governs the ordering of the physical signs it is made of. In principle, therefore, digital text is marked by the syntax of its code, by the arrangement of the physical tokens that stand for binary digits.

Any explicit feature of a text can be conceived as a mark. We may thus say that digital text is marked by the linear ordering of the string of coded characters that constitutes it as a data type, for the string shows explicitly its

own linear structure. The primary semiotic code of digital text is cast by the structural properties of a string of characters. The linearity of digital text as a data type puts an immediate constraint on its semiotics. It is a stream of coded characters, and each character has a position in an ordered linear succession.

But in common technical parlance, a string of coded characters is regarded as unmarked text. Markup is a special kind of coding, one laid on a textual object that has already been coded in another textual order entirely—that is to say, in the textual order marked by bibliographic codes. When we mark up a text with TEI or XML code, we are actually marking the pre-existent bibliographic markup and not the content, which has already been marked in the bibliographic object. This situation is the source of great confusion and must be clearly grasped if one is to understand what markup can and cannot do for bibliographically coded texts.

Markup is described, correctly, as “the denotation of specific positions in a text with some assigned tokens” (Raymond, Tompa, and Wood, *Markup*). In this sense, it adds to the linear string of digital characters its “embedded codes, known as tags.” A marked-up text is then commonly and properly understood as a tagged string of characters. But what function do tags perform with respect to bibliographic text, which is most definitely not a linear character string (though it can appear to be that to a superficial view)?

Let us continue to approach the problem from a computational point of view. In first-generation procedural markup systems, tags were used to add formatting instructions to a string of characters. With the introduction of declarative markup languages, such as SGML and its humanities derivative TEI, tags came to be used as “structure markers” (Joloboff 87). By adding structure to the string, semiotic properties of the digital text emerge as dependent functions of the markup with respect to the linear string of characters. It has been observed that adding structure to text in this way—that is, to text seen as flat or unstructured character data—enables “a new approach to document management, one that treats documents as databases” (Raymond, Tompa, and Wood, “From Data Representation” 3). But what does that understanding mean in semiotic terms?

The answer depends on the status of markup in relation to the bibliographically coded text. Markup makes explicit certain features of an originally paper-based text; it exhibits them by bringing them forth visibly into the expression of the text. It is therefore essentially notational. It affects the text’s expression, both digital and bibliographic, adding a certain type of structure to both.

But then we want to ask, How is that structure related to the content of the bibliographic object it is meant to (re)mark?

To show the crucial status of that question, let us make a thought experiment. Suppose we choose SGML as the markup language. Its syntax, a context-free grammar expressed by a document's DTD (document type definition), assigns a given hierarchical structure—chapters, sections, paragraphs, and so on—to the linear string of characters, the computer scientist's text. Text can thus be conceived as an "ordered hierarchy of content objects" (DeRose, Durand, Mylonas, and Renear 6; this is the OHCO textual thesis). But can textual content be altogether modeled as a mere set of hierarchically ordered objects? Are all textual relations between content elements hierarchical and linear? The answer is, clearly, no. Traditional texts are riven with overlapping and recursive structures of various kinds, just as they always engage, simultaneously, hierarchical and nonhierarchical formations. Hierarchical ordering is simply one type of formal arrangement that a text may be asked to operate with, and often it is not by any means the chief formal operative. Poetical texts in particular regularly deploy various complex kinds of nonlinear and recursive formalities.

Whatever the complexity of a bibliographic text's structure, however, that structure may be defined as "the set of latent relations" among the defined parts of the text (Segre and Kememy 34). Only through markup does that formal structure show explicitly at the level of textual expression. In principle, markup must therefore be able to make evident all implicit and virtual structural features of the text. Much depends on the properties of the markup system and on the relation between the markup tags and the string of character data. The position of the tags in the data may or may not be information-bearing. Forms of in-line markup, like those based in an SGML model, can exhibit only internal structure, that is, a structure dependent on "a subset of character positions" in textual data (Raymond, Tompa, and Wood, *Markup* 4). But textual structures, and in particular the content features of the text's structure, are "not always reducible to a functional description of subcomponents" of a string of characters (7). Textual structure is not bound, in general, to structural features of the expression of the text.

From a purely computational point of view, in-line "markup belongs not to the world of formalisms, but to the world of representations" (4). A formalism is a calculus operating on abstract objects of a certain kind, whereas a representation is a format or a coding convention to record and to store information. Unlike a calculus, a format does not compute anything, it simply provides a coding mechanism to organize physical tokens into data sets representing information. We can say, then, that markup is essentially a format,

or again, in a semiotic idiom, that markup is primarily notational. Inasmuch as it assigns structure to character strings or to the expression of textual information, it can refer only indirectly to textual content. In computational terms, it describes data structures but does not provide a data model or a semantics for data structures and an algebra that can operate on their values.

Attempts to use the DTDs in SGML systems as a constraint language, or formalism, to operate on textual data face a major difficulty in dealing with the multiple and overlapping hierarchical structures that are essential features of all textualities. Some circuitous ways out have been proposed, but in the end the solutions afforded provide “no method of specifying constraints on the interrelationship of separate DTDs” (Sperberg-McQueen and Huitfeldt, “Concurrent Document Hierarchies” 41). The use of embedded descriptive markup for managing documents as databases that can operate on their content is thus severely hampered by the dependence of SGML systems on internal structure. Content relations are best dealt with by forms of out-of-line markup, which “is more properly considered a specific type of external structure” (Raymond, Tompa, and Wood, *Markup* 4).

In general, we may therefore say that adding structure to textual data does not necessarily imply providing a model for processing the content of a text. A model applicable to document or internal structures can be appropriate only for directly corresponding content relations. To adequately process textual content, an external data model that can implement a database or operate some suitable knowledge representation scheme is required. The crucial problem for digital text representation and processing lies therefore in the ability to find consistent ways of relating a markup scheme to a knowledge-representation scheme and to its data model.

The Semantic Web project proceeds in that direction with its attempt to “bring structure to the meaningful content of Web pages” (Berners-Lee, Hendler, and Lassila). It is an effort to assign a formal model to the textual data available on the Web. The introduction of XML, a markup language profile that defines a generalized format for documents and data accessible on the Web, provides a common language for the schematic reduction of both the structure of documents—that is, their expression or expressive form—and the structure of their content. In this approach, the problem to solve consists precisely in relating the scheme that describes the format of the documents to the scheme that describes their content. The first would be an XML schema, “a document that describes the valid format of an XML dataset” (Stuart), and the second would be a metadata schema such as the resource description framework (RDF) being developed for the Semantic Web.<sup>5</sup>

61

AN XML SCHEMA AND A GENERAL MODEL

An RDF schema can be described as an “assertion model” that “allows an entity-relationship-like model to be made for the data.” This assertion model gives the data the semantics of standard predicate calculus (Berners-Lee). Both an XML schema and an RDF schema can assign a data model to a document, but in the first case the model depends on internal relations among different portions of the document, whereas in the second case it consists in an external structure independent of the structure of the document. In this context, XML documents act “as a transfer mechanism for structured data” (*Cambridge Communiqué*). XML works as a transfer syntax to map document-dependent or internal data structures into semantic or external data structures and vice versa. It is through markup that textual structures show up explicitly and become processable.

## MARKUP AND THE GENERAL THEORY OF TEXTUALITY

In this context, an important question rises to clear view. Since text is dynamic and mobile and textual structures are essentially indeterminate, how can markup properly deal with the phenomena of structural instability? Neither the expression nor the content of a text are given once and for all. Text is not self-identical.<sup>6</sup> The structure of its content very much depends on some act of interpretation by an interpreter, nor is its expression absolutely stable. Textual variants are not simply the result of faulty textual transmission. Text is unsteady, and both its content and expression keep constantly quivering. As Valentin Voloshinov has it, “what is important about a linguistic form is not that it is a stable and always self-equivalent signal, but that it is an always changeable and adaptable sign” (68).

Textual mobility originates in what has been described as “the dynamic[s] of structures and metastructures [that lie] in the heart of any semiotic activity” (Y. Neuman 67), and it shows up specifically in the semantic properties of those kinds of expression that set forth what linguists call reflexive metalinguistic features of natural language.<sup>7</sup> Diacritical signs are self-describing expressions of this kind, and markup can be viewed as a sort of diacritical mark. A common feature of self-reflexive expressions is that they are semantically ambiguous. They are part of the text and they also describe it; they are at once textual representations and representations of a textual representation. Markup, therefore, can be seen either as a metalinguistic description of a textual feature or as a new kind of construction that extends the expressive power of the object language and provides a visible sign of some implicit textual content.

A diacritical device such as punctuation, for instance, can be regarded as a kind of markup (Coombs, Renear, and DeRose 935), and by adding punctuation to a medieval text, a modern editor actually marks it up. Editorial punctuation, therefore, can be considered either as part of the text or as an external description related to it. In the first case, it produces a textual variant; in the second, a variant interpretation. Accordingly, any punctuation mark is ambivalent: it can be seen as the mark of an operation or as the mark of an operational result. If it is regarded as part of the text, it brings in a variant reading and has to be seen as a value for an operation of rephrasing; at the same time, by introducing an alternative reading, it casts a new interpretation on the text and must be seen as a rule to an action of construing. Yet the very same punctuation mark can be regarded as an external description of the text. In that case, it assigns a meaning to the text and must be seen as a value for an operation of construal. By providing a new interpretation, however, it adds structure to the wording of the text and must be seen as a rule to an action of “deformance” (for a definition of this term, see McGann, *Radiant Textuality*). Marks of this kind, viewable either way, behave just as Ludwig Wittgenstein’s famous duck-rabbit picture (*Philosophical Investigations* 2.11).

This sort of semantic ambivalence enables any diacritical mark, or for that matter any kind of markup, to act as a conversion device between textual and interpretational variants. Far from stabilizing the text, the markup actually mobilizes it. Through markup, an interpretational variant assumes a specific textual form; conversely, that explicit form immediately opens itself to interpretive indeterminacy. Markup has to do with structure or logical form. It describes the form or exposes it in the text. But the logical form of a textual expression is only apt to show or to express itself in language, and, as Wittgenstein puts it, “that which mirrors itself in language, language cannot represent.”<sup>8</sup> The only way to represent a logical form is to describe it by means of a metalanguage. The markup, on its part, may either exhibit or describe a logical form, but it can perform both functions only by changing its logical status: it has to commute between object language and metalanguage, so as to frame either an external metalinguistic description or an object-language, self-reflexive expression. Markup, therefore, is essentially ambivalent and sets forth self-reflexive ambiguous aspects of the text, which can produce structural shifts and make it unstable and mobile.

Text is thus open to indeterminacy, but textual indetermination is not totally unconstrained. Because of textual mobility, we may say that text is not self-identical. But putting things the other way around, we may also say that

text is, virtually, identical with itself, because the whole of all its possible variant readings and interpretations makes up a virtual unity identical with itself. Text in this view is not an arbitrary unity, for if it were seen as such, no text would differ from any other. The entirety of all latent capacities of the text is virtually one and the same, and this self-identity imposes limiting conditions on mobility and indetermination. The latent unity of the text brings about phenomena of mutual compensation between the stability of the expression and the variety of its possible interpretations or, conversely, between the instability of the expression and the steadiness of its conceptual import. With any given expression comes an indefinite number of possible interpretations, just as for any given conceptual content we may imagine an indefinite number of possible concrete formulations. But for a given text, the variation of either component is dependent on the invariance of its related counterpart, and such variation can come about only under this condition.

Semantic ambiguity may be thought of as an obstacle to an automatic processing of textual information, but actually it can serve that very purpose. Markup can provide a formal representation of textual dynamics precisely on account of its diacritical ambivalence and its capacity to induce structural indeterminacy and compensation. The OHCO thesis about the nature of the text is radically insufficient, because it does not recognize structural mobility as an essential property of the textual condition. The OHCO view builds on the assumption of a syntactically well-determined expression, not acknowledging that a fixed syntactic structure leaves the corresponding semantic structure open to indetermination. A nonsemantically identifiable string of characters is thus regarded as the vehicle of a specific content. A digital text representation need not assume that meaning can be fully represented in a syntactic logical form.<sup>9</sup> The automatic processing of the text does not depend on a condition of this kind and need not fall victim to the snares of classical artificial intelligence. A formal representation of textual information does not require an absolute coincidence between syntactic and semantic logical form. In this respect, the role of markup can be of paramount importance in bringing their interconnections to the fore. Markup, turning to account its operational dimension, can act as a transfer mechanism between one structure and the other. It can behave as a performative injunction and switch to a different logical function.

Viewing markup as an operator in this sense, we may say, as has been proposed, that “to describe the meaning of the markup in a document, it suffices to generate the set of inferences about the document which are licensed by the markup,” or even more assertively, that “in some ways, we can

regard the meaning of the markup as being constituted, not only described, by that set of inferences” (Sperberg-McQueen, Huitfeldt, and Renear 231). Actually, to describe markup in this way amounts to seeing it as a kind of “inference-ticket,” to use Gilbert Ryle’s locution—as an assertion, belonging to a “different level of discourse” from that to which belong the expressions it applies to (121). So described, markup functions as a higher-order object-language statement—as a rule that licenses the reader, or for that matter the machine, to interpret the text in a certain way and to assign dynamically a structure to its content. Markup can therefore be conceived as a transfer mechanism from a document’s structure to its semantic structure or, the other way around, from a semantic structure to a document’s structure (as in the implementations being imagined in the Semantic Web project).

Diacritical ambiguity, then, enables markup to provide a suitable type of formal representation for the phenomena of textual instability. By seeing markup in this way, we can regard it as a means of interpretation and deformance (see Samuels and McGann), as a functional device both to interpret and to modify the text. But in the OHCO view, the structure assigned to the expression of a text (by marking it up) and the structure assigned to its content coincide, with the result that the capacity of the markup to account for textual dynamics is prevented. Markup should not be thought of as introducing—as being able to introduce—a fixed and stable layer to the text. To approach textuality in this way is to approach it in illusion. Markup should be conceived, instead, as the expression of a highly reflexive act, a mapping of text back onto itself: as soon as a (marked) text is (re)marked, the meta-markings open themselves to indeterminacy. This reflexive operation leads one to the following formulation of the logical structure of the textual condition:

$$(A = A \iff A \neq A) \iff A \overset{m}{\text{-----}} A$$

(Buzzetti, “Digital Representation” 84)

In this view markup (*m*) is conceived as the expression of an operation, not of its value, for the nature of text is basically injunctive. Text can actually be seen as the physical mark of a sense-enacting operation (an act of *Besinnung*). But in its turn, the result of this operation, the expression of the text, must be seen not as a value but as an operation mark, otherwise its interpretation is prevented. Such an expression of the text is then regarded as a rule for an act of interpretation, an operation that is essentially underdetermined. Interpretation, as an act of deformance, flags explicitly its result as a self-reflexive textual mark, which imposes a new structuring on the expression of the text.



Again, the newly added structural mark, the value of the interpreting operation, converts back into an injunction for another, indeterminate act of interpretation.

Textual dynamics is thus the continual unfolding of the latent structural articulations of the text. Any structural determination of one of its two primary subunits, expression and content, leaves the other underdetermined and calls for a definition of its correlative subunit, in a constant process of impermanent codetermination. In more detail, and referring to the interweaving of textual content and expression, we may say that an act of composition is a sense-constituting operation that brings about the formulation of a text. The resulting expression can be considered as the self-identical value of a sense-enacting operation. By fixing it, we allow for the indetermination of its content. To define the content, we assume the expression as a rule for an interpreting operation. An act of interpretation brings about a content, and we can assume it as its self-identical value. A defined content provides a model for the expression of the text and can be viewed as a rule for its restructuring. A newly added structure mark can in turn be seen as a reformulation of the expression, and so on, in a permanent cycle of compensating actions between determination and indetermination of the expression and the content of the text.

This continual oscillation and interplay between indetermination and determination of the physical and the informational parts of the text renders its dynamic instability very similar to the functional behavior of self-organizing systems. Text can thus be thought of as a simulation machine for sense-organizing operations of an autopoietic kind. Text works as a self-organizing system inasmuch as its expression, taken as a value, enacts a sense-defining operation, just as its sense or content, taken as a value, enacts an expression-defining operation. Text provides an interpreter with a sort of prosthetic device to perform autopoietic operations of sense communication and exchange.

Textual indeterminacy and textual instability can thus be formally described, like most self-organization processes, through the calculus of indications introduced by George Spencer-Brown (see Buzzetti, "Diacritical Ambiguity"). His "nondualistic attempt" to set proper foundations for mathematics and descriptions in general "amounts to a subversion of the traditional understanding on the basis of descriptions," inasmuch as "it views descriptions as based on a primitive *act* (rather than a logical value or form)." In Spencer-Brown's calculus "subject and object are interlocked" (Varela 110), just as expression and content are interlocked in a self-organizing textual

system. Only an open and reversible deforming or interpreting act can keep them connected as in a continually oscillating dynamic process. Louis Kauffman's and Francisco Varela's extension of Spencer-Brown's calculus of indications (Kauffman and Varela; Varela, ch. 12) accounts more specifically for the "dynamic unfoldment" (Varela 113) of self-organizing systems and may therefore be consistently applied to an adequate description of textual mobility.

## FROM TEXT TO WORK: A NEW HORIZON FOR SCHOLARSHIP

Exposing the autopoietic logic of the textual condition is, in a full Peircean sense, a pragmatic necessity. As Varela, Humberto Maturana, and others have shown, this logic governs the operation of all self-organizing systems (Maturana and Varela; Varela et al.). Such systems develop and sustain themselves by marking their operations self-reflexively. The axiom that all text is marked text defines an autopoietic function. Writing systems, print technology, and now digital encoding license a set of markup conventions and procedures (algorithms) that facilitate the self-reflexive operations of human communicative action.

Scholarly editions are a special, highly sophisticated type of self-reflexive communication, and the fact is that we now must build such devices in digital space. This necessity is what Charles Sanders Peirce would call a "pragmatic" fact: it defines a kind of existential (as opposed to a categorical) imperative that scholars who wish to make these tools must recognize and implement. We may better explain the significance of this imperative by shifting the discussion to a concrete example. Around 1970, various kinds of social text theories began to gain prominence, pushing literary studies toward a more broadly cultural orientation. Interpreters began shifting their focus from the text toward any kind of social formation in a broadly conceived discourse field of semiotic works and activities. Because editors and bibliographers oriented their work to physical phenomena—the materials, means, and modes of production—rather than to the readerly text and hermeneutics, this tectonic shift in the larger community of scholars barely registered on bibliographers' instruments.

A notable exception among bibliographic scholars was McKenzie, whose 1985 Panizzi lectures climaxed almost twenty years of work on a social text approach to bibliography and editing. When they were published in 1986, the lectures brought into focus a central contradiction in literary and

cultural studies (*Bibliography*). Like their interpreter counterparts, textual and bibliographic scholars maintained an essential distinction between empirical-analytic disciplines on one hand and readerly-interpretive procedures on the other. In his Panizzi lectures McKenzie rejected this distinction and showed by discursive example why it could not be intellectually maintained.

His critics—most notably Thomas Tanselle and T. Howard-Hill—remarked that while McKenzie's ideas had a certain theoretical appeal, they could not be practically implemented (Howard-Hill; Tanselle, "Textual Criticism and Literary Sociology"). The ideas implicitly called for the critical editing of books and other socially constructed material objects. But critical editing, as opposed to facsimile and diplomatic editing, was designed to investigate texts—linguistic forms—not books or (what seemed even more preposterous) social events.

In fact one can transform the social and documentary aspects of a book into computable code. Working from the understanding that facsimile editing and critical editing need not be distinct and incommensurate critical functions, the *Rossetti Archive* proves the correctness of a social text approach to editing: it pushes traditional scholarly models of editing and textuality beyond the Masoretic wall of the linguistic object we call the text. The proof of concept would be the making of the archive. If our breach of the wall was minimal, as it was, its practical demonstration was significant. We were able to build a machine that organizes for complex study and analysis, for collation and critical comparison, the entire corpus of Rossetti's documentary materials, textual as well as pictorial. Critical, which is to say computational, attention was kept simultaneously on the physical features and conditions of actual objects (specific documents and pictorial works) as well as on their formal and conceptual characteristics (genre, metrics, iconography).<sup>10</sup> The archive's approach to Rossetti's so-called double works is in this respect exemplary. Large and diverse bodies of material that comprise works like "The Blessed Damozel" get synthetically organized: scores of printed texts, some with extensive manuscript additions; two manuscripts; dozens of pictorial works. These physical objects orbit around the conceptual thing we name for convenience "The Blessed Damozel." All the objects relate to that gravity field in different ways, and their differential relations metastasize when subsets of relations among them are revealed. At the same time, all the objects function in an indefinite number of other kinds of relations: to other textual and pictorial works, to institutions of various kinds, to different persons, to varying occasions. With the archive one can draw these materials into computable synthetic relations at macro- as well as microlevels. In the process the archive

discloses the hypothetical character of its materials and their component parts as well as the relations one discerns among these things. Though completely physical and measurable (in different ways and scales), neither the objects nor their parts are self-identical; all can be reshaped and transformed in the environment of the archive.

The autopoietic functions of the social text can also be computationally accessed through user logs. This set of materials—the use records, or hits, automatically stored by the computer—has received little attention by scholars who develop digital tools in the humanities. Formalizing its dynamic structure in digital terms will allow us to produce an even more complex simulation of social textualities. Our neglect of this body of information reflects, I believe, an ingrained commitment to the idea of the positive text or material document. The depth of this commitment can be measured by reading McKenzie, whose social text editing proposals yet remain faithful to the idea of the “primacy of the physical object” as a self-identical thing (“What’s Past” 274).

Reflecting on digital technology in his lecture “What’s Past Is Prologue,” McKenzie admitted that its simulation capacities were forcing him to rethink that primary article of bibliographic faith (272–73). He did not live to undertake an editorial project in digital form. Had he done so, we believe he would have seen his social text approach strengthened by the new technical devices. All editors engage with a work in process. Even if only one textual witness were to survive—say, that tomorrow a manuscript of a completely unrecorded play by Shakespeare were unearthed—that document would be a record of the process of its making and its transmission. Minimal as they might seem, its user logs would not have been completely erased, and those logs are essential evidence for anyone interested in reading (or editing) such a work. We are interested in documentary evidence precisely because it encodes, however cryptically at times, the evidence of the agents who were involved in making and transmitting the document. Scholars do not edit self-identical texts. They reconstruct a complex documentary record of textual makings and remakings, in which their own scholarly work directly participates.

No text, no book, no social event is one thing. Each is many things, fashioned and refashioned in repetitions that often occur (as it were) simultaneously. The works evolve and mutate in their use. And because all such uses are always invested in real circumstances, these multiplying forms are socially and physically coded in and by the works themselves. They bear the evidence of the meanings they have helped to make.

One advantage digitization has over paper-based instruments comes not from the computer's modeling powers but from its greater capacity for simulating phenomena—in this case, bibliographic and sociotextual. Books are simulation machines as well, of course. Indeed, the hardware and software of book technology have evolved into a state of sophistication that dwarfs computerization as it currently stands. In time this situation will change through the existential imperative—digitization—that now defines our semiotic horizon. That imperative is already leading us to design critical tools that organize our textual condition as an autopoietic set of social objects—that is to say, objects that are themselves the emergent functions of the measurements that their users and makers implement for certain purposes. Our aim is not to build a model of one made thing, it is to design a system that can simulate the system's realizable possibilities—those that are known and recorded as well as those that have yet to be (re)constructed.

McKenzie's central idea, that bibliographic objects are social objects, begs to be realized in digital terms and tools, begs to be realized by those tools and by the people who make them.

## NOTES

1. The best introduction in English to this broad subject is Greetham, *Textual Scholarship*; see also his important *Theories of the Text*. A brief introduction can be found in Williams and Abbott.
2. On the nonself-identity of material objects, see McGann, *Radiant Textuality*, chapters 5 and 6.
3. This genetic work was initiated with Friedrich Beissner's project (begun in 1943) to edit the work of Hölderlin. It was continued in the edition of D. E. Sattler, begun in 1975. The best known English-language genetic edition is Hans Walter Gabler's *Ulysses: A Critical and Synoptic Edition* (1984; Gabler, Steppe, and Melchior).
4. A good example of versioning is provided by Wordsworth's *The Prelude: 1799, 1805, 1850*.
5. For good bibliographic sources on RDF, see [www.w3.org/RDF/](http://www.w3.org/RDF/).
6. For a more thorough discussion of this assertion, see McGann, *Radiant Textuality*, especially chapter 5 and the appendix to chapter 6.
7. Cf. Hjelmslev 132: "Owing to the universalism of everyday language, an everyday language can be used as metalanguage to describe itself as object language."
8. *Tractatus* 4.121: see also 4.1212: "What can be shown cannot be said."
9. "If you take care of the syntax, the semantics will take care of itself" (Haugeland 23).

10. Since its initial conception, the *Rossetti Archive* has been subjected to further digital transformations—most notably a translation into XML format—that extend the archive's translinguistic critical functions. The digital logic of the archive's structure leaves it open to more comprehensive scales of interoperability, such as those being developed through the Semantic Web and the Open Knowledge Initiative (OKI). For an introduction to OKI see [www.okiproject.org](http://www.okiproject.org).