

# Fragmentary Texts and Digital Libraries

by

Monica Berti  
([monica@monicaberti.it](mailto:monica@monicaberti.it))

Digital technologies and on-line repositories are providing classicists with new tools for collecting and studying ancient sources. Digital philology is devising models and instruments for representing critical texts with variants of manuscripts and conjectures of philologists. It is now possible to develop also tools for realizing digital corpora of fragments of lost authors and works, in order to preserve an inestimable cultural heritage built across the centuries.

I write here some remarks on the main characteristics of fragmentary texts that should be represented in digital format. These remarks are part of the work which is being conducted at the *Perseus Digital Library* for a project on fragmentary authors.

The first question is to remind that *textual fragments* refer to a particular phenomenon, because they are not portion of an original larger whole like material fragments, but the result of a work of interpretation conducted by scholars, who extract and collect information pertaining to lost works embedded in other surviving texts. This use of the term fragment may be misleading, because the original text of the excerpt is usually covered by the context of transmission and distorted by the style and purpose of the author who has extracted and quoted it (usually called the “witness” of the fragment).

*Print collections of fragmentary texts* consist of textual excerpts drawn from many different sources and arranged according to various criteria, such as chronological order or thematic disposition. The length of these excerpts can be significantly different from one edition to another and depends on the editor’s choice. Moreover, when an extracted portion of text is printed, it immediately acquires a sort of materiality due to its typographical representation: it has very definite margins like a real fragment, but it is actually the result of a modern extraction and interpretation; it can give false illusions because the fragment in itself doesn’t exist, and it is only like a shadow, whose shape is blurred and can lead to a distorted perception of reality.

Secondly, one of the main concern when raising evidence of lost works is reconstructing the *complex relationship between the fragment and its source of transmission*, which means weighing the level of interference played by the author who has reused and transformed the original context of the fragment, measuring therefore the distance between the source text and the derived text, and trying to perceive the degree of text reuse and its effect on the resulting target text.

This interpretative process is usually explained in the commentary of a fragment edition or in papers and monographs pertaining to various aspects of fragmentary authors and works, but it is completely lost in the print representations of the fragments, which are simply typographical reproductions of extracts of derived texts.

Digital libraries and hypertextual models allow to rethink the fundamental question of the relation between the fragment and its context, representing and expressing every element of print conventions in a more dynamic and interconnected way.

## Representing Fragments

Print editions of fragments contain extracts from many different sources and are thus paper representations of hypertexts. Now that the source editions from which fragments are extracted are becoming available in digital form, it is possible to construct editions that are truly hypertextual, including not only excerpts but links to the scholarly sources from which those excerpts are drawn.

Encoding fragments is first of all the result of interpreting them, developing a language appropriate for representing every element of their textual features, thus creating meta-information through an accurate and elaborate semantic markup. Editing fragments, therefore, signifies producing meta-editions that are different from printed ones because they consist not only of isolated quotations but also of pointers to the original contexts from which the fragments have been extracted.

On a broader level, the goal of a digital edition of fragments is to represent multiple transtextual relationships as they are defined in literary criticism (cf. Gérard Genette), which include intertextuality (the presence of a text inside another text, such as quotations, allusions, and plagiarism), paratextuality (i.e., all those elements which are not part of the text, like titles, subtitles, prefaces, notes, etc.), metatextuality (critical relations among texts, i.e. commentaries and critical texts), architextuality (which means the generic quality and status of a text), and hypertextuality (i.e., the derivation of a text from a preexisting hypotext through a process of transformation or imitation).

Designing a digital edition of fragments also means finding digital paradigms and solutions to express information about printed critical editions and their editorial and conventional features. Working on a digital edition means converting traditional tools and resources used by scholars such as canonical references, tables of concordances, and indexes into machine actionable contents.

Standards, protocols, and tools now available permit us to express the hypertextual and hermeneutical nature of fragmentary texts, providing scholars with an interconnected corpus of primary and secondary sources of fragments that also includes critical apparatuses, commentaries, translations, and modern bibliography on ancient texts. The first requirement for building a digital collection of fragmentary texts then is to make the semantic contents of critical print editions machine readable, defining a general architecture for representing at least the following main elements pertaining to the domain of fragmentary texts in a digital library:

- ***Fragment as machine actionable link***

The fragment should be linked to the whole text of the source in which it is transmitted. This is the first function for a proper representation of fragmentary texts: in this way it is possible to see the excerpt directly inside its context of transmission, avoiding the misleading idea of an independent material existence of fragmentary texts, which derives from typographical representation of excerpts that are actually the result of modern reconstructions of lost works.

- ***Start and end of a fragment***

Linking the fragment to its source means collocating it again in its original context. The next step is providing a mechanism for marking the beginning and the end of a fragment in this context according to the choices of different editors. The result is that the reader, while visualizing the excerpt inside its source of transmission, is able to see

simultaneously the representation of different lengths of the same fragment based on editions that have adopted different textual criteria.

- ***Numbering and ordering fragments***

Numbering and ordering fragments may vary in a significant way from one edition to another. These differences depend on the choices of the editor, who can decide to order the fragments – and consequently number them – according to different internal or external characteristics of the fragments themselves or of their sources. A digital representation of fragments should provide the possibility of encoding this kind of information, which is usually registered in the table of concordances of a printed edition: aligning multiple references to the same textual object can help the reader visualize different numberings and orderings of fragments in different editions, including new data if new editions are added.

- ***Representing information on fragmentary authors and works***

Within the source transmitting the fragment, it is necessary to specify that a given segment of the text is the name of the author to which the fragment is attributed, and in some cases also the title of the work and the book number to which the fragment originally belonged. Attributing a fragment to an author and a work can be a difficult task, because we can have homonymous authors and also because managing titles of ancient works can be quite challenging: in most cases, witnesses do not cite the title of the work from which they have drawn the fragment; moreover, in ancient sources, the title of a work may be attested with more or less significant variants, and the result is that different editors may attribute the same fragment to different authors and works. The goal is to develop a comprehensive catalog of unique identifiers for every fragmentary author and work that will include multiple expressions of the same author and work and where each entry will have associated meta-data, providing the scholar with a sort of canon that simultaneously includes all available information on fragmentary authors and works, with pointers to primary and secondary sources. This function, beside providing the scholar with an innovative tool, can be very helpful in enhancing one of the “theoretical questions” suggested by Glenn Most when collecting fragments, i.e. the relationship between fragmentary authors and the “shifting boundaries of canon formation over time.”

- ***Classifying fragments***

Fragments are classifiable according to multiple criteria ranging from internal to external factors. The first classification is based on literary genre, which ancient fragments of classical literature cover almost entirely, from epic and poetry to oratory and historiography. Inside the same collection, fragments are usually distinguished as *testimonia* (i.e., fragments providing biographical and bibliographical information about fragmentary authors) and *fragmenta* (i.e., fragments of lost works). Other criteria for classifying fragments belonging to the same literary genre can also be applied, as is shown by the monumental work of Felix Jacoby in editing Greek historical fragments, which is one of the most important results achieved in the field of ancient historiography. Nevertheless, the print representation of these categories has many limitations, because it is impossible to draw a demarcation line among many different genres of fragmentary authors and works that can be inserted in different overlapping categories, and the result is that the same fragment is often repeated in many different sections corresponding to different categories. A digital collection in which every

fragment is preserved in its original context and represented with multiple pieces of meta-data can express the complexity of modern classifications, while not scattering and repeating the same excerpt many different times. In this way it is possible to avoid the strictness of printed categories, allowing scholars to compare a fragment with many other excerpts and visualizing its belonging to different categories in a more dynamic and simultaneous way.

### **Representing Textual Variants and Conjectures**

Print collections of fragments often include a critical apparatus, which is normally not based on a new examination of the original manuscripts that bear witness to the text, but on a selection of variants and conjectures drawn from the best critical editions of fragment sources. This choice is principally due to the fact that it would take too much time to examine every manuscript, and also because a work of this kind would go beyond the competencies and purposes of the editors of fragments, who are primarily interested in reconstructing content and characteristics of lost works.

Both the emerging cyberinfrastructure for the humanities and the research conducted in the field of ePhilology have devised a new concept of Greek and Latin textual corpora, where the aim is to provide scholarly services and methods for tracking and comparing multiple versions of the same text across time, affecting in a fundamental way as well future work on fragmentary texts:

- ***Multiple editions and alignment of citation schemes***

The first step is to collect every edition of the sources preserving fragments as well as collections of fragmentary works, so that a particular passage can be visualized in different versions of the same text reconstructed by different editors.

- ***Dynamic collation of multiple editions and digital criticism***

Collecting multiple critical editions of the same text means building a “multitext,” which is a “network of versions with a single, reconstructed root,” so that scholars can compare different textual choices and conjectures produced by philologists (the concept of multitext is the result of work conducted by the Homer Multitext Project of the Center for Hellenic Studies, which aims at producing a new digital representation of the textual tradition of the Homeric poems: see Blackwell-Crane and Dué-Ebbott in DHQ 2009.3.1). This process involves a new way of conceiving literary criticism because it produces a representation and visualization of textual transmission completely different from print conventions, where the text that is reconstructed by the editor is separated from the critical apparatus that is printed at the bottom of the page. In addition, the inclusion of images of manuscripts, papyri, and other source materials allows the reader to have a dynamic visualization of the textual tradition and to perceive the different channels of both the transmission and philological production of the text that is usually hidden in the static, concise, and necessarily selective critical apparatuses of standard printed editions. Producing a multitext, therefore, means producing multiple versions of the same text, which are the representation of the different steps of its transmission and reconstruction, from manuscript variants to philological conjectures. This process has fundamental consequences for the study of ancient sources in general and for fragmentary ones in particular, given that, while studying fragments and evaluating their distance from the original version, it is imperative to examine the manuscript

variants of the source text, in order to see what can be attributed to the witness or to the transmission of the text across centuries.

### **Secondary and Tertiary Sources**

Collecting fragments also means looking for many other kinds of information directly or indirectly connected to fragmentary authors. These data are usually labelled as “secondary” and “tertiary sources,” and may be summarized into the following fundamental categories: 1) *Loci Paralleli*, i.e. secondary ancient sources parallel to the witness of a fragment. Even if the relationship of a *locus parallelus* to the main quoter of a fragmentary text may involve many aspects, *loci paralleli* form two principal groups: a) sources quoting or paraphrasing the same fragment (in most cases these sources are chronologically later than the witness); b) sources treating the same subject of the fragment. 2) *Tertiary Sources*, i.e. modern bibliography consisting of monographs, papers, encyclopedia, grammars, translations, and other bibliographical tools giving information and commentaries on a wide range of materials pertaining to the fragment, its author, and its source of transmission.

A digital representation of fragmentary texts should provide links to secondary and tertiary sources, identifying passages in articles and monographs related to the fragment and the context from which the fragment has been drawn. Mass digitization projects like Google Books, Google Scholar, and Internet Archive are providing many collections of secondary and tertiary sources useful to classicists. Moreover, repositories like JSTOR and Project MUSE offer access to the titles of leading academic journals within a variety of disciplines, as well as monographs and other materials fundamental for scholarly activities: these archives are full-text searchable and offer many possibilities of interdisciplinary research, including high-quality images and interlinked citations and references.

In addition to these resources, there are other projects and electronic publications for digital classicists developed by organizations such as The Stoa Consortium and founded on the principle of open access. One of the most significant Stoa projects is the Suda On Line (SOL), which is particularly important for those interested in building a digital collection of Greek fragmentary authors, because the Suda preserves a lot of fragments of classical authors, which in most cases can be classified as *loci paralleli*. The aim of the project is to create an on-line version of this encyclopedic lexicon, providing, for the first time, a translation and interpretive apparatus for each entry thanks to the international cooperative efforts of many scholars. All of these resources represent the types of sources that should be included when devising a digital representation of fragmentary texts, in order to build a dynamic and interconnected corpus of primary, secondary, and tertiary sources.

### **Translation and Commentary**

Two other fundamental elements of modern collections of fragmentary works that may receive a great improvement in digital libraries are translations and commentaries. Translating texts means not only providing a service for those who do not have a good knowledge of ancient languages, but it is first of all an essential part of the scholarly interpretation produced by the editor. Through the collection of multiple editions of the same work, a digital library will allow scholars to also consult multiple translations into multiple languages, comparing different interpretations and linguistic restitutions of the

same passage. At a deeper level, aligning multiple editions enables us to create machine actionable dictionaries and dynamic lexica of Greek and Latin words and their corresponding terms in modern languages, providing an inestimable tool for scholars and for a wide range of linguistic, grammatical, and syntactic analyses.

As far as concerns fragments, Guido Schepens has pointed out that the commentary to the text is constituted by two fundamental tasks: the first is the effort to “deconstruct” the context that preserves the quotation in order to find the original characteristics of the fragment, and the second one is to try to “reconstruct” the fragment and the lost work to which it belonged. As for textual variants, conjectures and translations, a digital library should provide every passage with links to multiple commentaries drawn from the editions of fragments and source texts. A true digital commentary, however, can be conceived as something broader than that, because it can include every possible annotation identifying every phenomenon pertaining to the text, thus providing traditional commentaries with a wide series of services, ranging from morphological and syntactic analysis to named entity identification and different explanations or disputes on every aspect of the textual content.

### **Conclusion**

Devising a model and an architecture for representing fragmentary texts in a digital library is a fundamental contribution toward a systematic and structural analysis of the multiple layers of production and interpretation that constitute a textual fragment. In particular, the two most important goals of such a work are: 1) *Representing a textual fragment as a hypertext*, i.e. as a text derived from another text and interconnected to many other different typologies of texts: this means envisioning and building an expansible set of links that express multiple relations of the text of the fragment with itself (i.e., with the text embedding and transmitting the fragment) and with a wide range of secondary and tertiary sources (i.e., ancient evidence, commentaries, and many other kinds of bibliographical tools). 2) *Representing a textual fragment as a multitext*, i.e. as the result of a work of stratification of manuscript variants and scholarly conjectures that form the path through which the fragment has survived and without which it wouldn't exist as evidence.