New Publication Cultures in the Humanities

Exploring the Paradigm Shift

Péter Dávidházi (Ed.)

New Publication Cultures in the Humanities

New Publication Cultures
in the Humanities

Exploring the Paradigm Shift

Edited by Péter Dávidházi

In memory of Irma Vogel (1954-2011)
who worked for this project to the last

Computer-assisted Scholarly Editing of Manuscript Sources

Andrea Bozzi

Introduction

Over the last months, various reports have appeared in the media regarding the creation of large digital libraries implemented by public institutions and private companies working in collaboration. More recently, many of the problems connected with copyright seem to have been resolved or have found feasible solutions. In the meantime, new tools for the dissemination and reading of *e-books* have started to appear on the market.

A new form of publication appears to be emerging: on the one hand this important issue tends to be oversimplified in the media, especially when the advent of these new technologies is presented as the first step towards the disappearance of the traditional paper book; on the other hand, the general reading public is not informed on another aspect of the problem, which probably does not have the same importance for them as it has for the specialists. What changes in study methods should scholars dealing with linguistic, philological and literary texts expect? In this contribution, I will try to show how it is possible to keep alive (and possibly improve) fields of study that have a long tradition and which are apparently threatened by the advent of radical transformations in the creation and dissemination of information.

My intervention will concentrate on the specific aspect of textual criticism. I realize this is a discipline that could be defined as being 'very exclusive,' as the scholars are not numerically equivalent to the community of people working in other Humanities disciplines; for example historians, philosophers, or those dealing with the history and criticism of literature. However, if we consider that textual criticism covers a very large period (Ancient, Medieval and Modern times) and many languages, there is also an increase in the population of specialists. If we add the specialists working on the philology of ancient printed texts and those studying the manuscripts of modern and contemporary authors, we realize that we are dealing with a very respectable community, which deserves to be helped to access the new digital world. An important element that needs to be considered is the inexplicable reluctance on the part of this community of scholars to

accept the use of new technology. For this reason, it is our responsibility not only to enhance the study and development of suitable technology, but also to promote appropriate and effective actions able to convince sceptics that these technological tools are crucial to accompany the disciplines of philology and textual criticism into the new field of Digital Humanities.

The Model

I have called the model I am using for this community of textual critics 'system for digital philology,' which can be summarized as follows. Firstly, it is necessary to adopt shared standards at the international level for text labelling, so that the elements contained in the texts (format values, style, capitals, onomastics, etc.) can be recognized by internet browsers or can interoperate with other data available on the Web. Even the annotations added to the text (as, for example, the lemma for each wordform or their part of speech, the indication of variants, etc.), produced either manually or using NLP[1] tools, must respect specific standards. This element currently has great importance since the standards have a crucial function for the creation of research infrastructures enabling the members of the same community of users and of other communities operating in the human sciences to share the resources made available on the Web.

Furthermore, a digital philology application should be provided with a series of fundamental research tools within a modular architecture. The system is developed by adding to the core of basic tools increasingly specialized ones designed to meet particular needs. This developmental strategy can customize the system to the many philology types and, at the same time, can be specialized for each one. Some of the basic tools are represented, for example, by an indexer, a concordance program, a statistic analyzer, an image enhancement system, etc.

The digital document, eventually supplied with additional information and annotated standard markup systems, is produced on a server so that it can be queried via the Web. In this phase, the administration of accesses is important because data visibility could initially be limited only to some members of the community (collaborative editions) and later become public (final publication). XML encoding of the documents and annotations allows for a second form of dissemination, that of traditional paper, at any time by adopting appropriate XML data transformers, also in view of the print-on-demand or ebook-controlled distribution. I shall not focus here on the standards issue; instead, let us browse through the basic tools for

scholarly editing of digital documents, bearing in mind the prospect of a digital philological system.[2]

We are convinced that an appropriately designed scholarly editing module of the computational philological system meeting the specific needs of classical and medieval philology, especially in the field of Romance languages, will also respond to other types of philology; for example, Greek and late Aegyptian papyrology, epigraphy, palaeo-Slavonic philology, philology of ancient printed texts, and – with appropriate adaptation – the philology of texts of modern and contemporary authors. The basic criteria are the following:

– linear transcription of a single source;
– positive apparatus for recording the variants of the collated sources;
– specific area of the apparatus for storing the readings selected or proposed by the critical editor;
– automatic generation of the *textus constitutus*;
– automatic generation of the text of all the other reviewed and collated sources;
– computer-assisted assessment of the variants and man-machine user interface to hypothesize stemmata resulting from the apparatus data.

From the Model to the Textual Criticism Web Application

The application that I am presenting here is aimed at achieving these ambitious targets through an architecture based on interconnected modules. In other words, it works with a nucleus of components for the treatment of both text files and digital image files, which form the core of the system. According to the specific needs, from time to time a number of programs are added both for the management of images (enhancement, segmentation, pattern recognition, etc.) and of text (natural language processing, information extraction, data mining, electronic editing, etc.).

1. Textual levels.
The main principle is the structural subdivision of the textual data to be represented in a digital system: the data can include images (of the document containing the text), text (namely, the diplomatic or interpretative transcription of the text contained in the images), or extra-and paratextual information. The first group (extratextual information) comprises the references of the text in the page (page number or manuscript folio, number of paragraph, comma, section, running titles in the case of a dictionary

or volume of encyclopaedia, etc.). The last one (paratextual information) includes annotations, apparatuses and bibliography.

This structure, consistent with the encoding and markup systems adopted at an international level for textual data, and referred not only to texts (see, for example, the markup language of the *Text Encoding Initiative*-TEI,[3] now available also in xml version), places the different types on different levels: for instance, a number representing a date within a work or a critical note studied and edited by a philologist are elements belonging to information levels different from the text proper.

Therefore, the application takes these different levels into account, provided they have been appropriately marked, so that they may be exploited by the user who is performing search operations. The user will first need to select the option activating the function expressed by these distinctive elements, which ensure better results than those that would be obtained without these distinctions.

A typical example is that of indexation, which can, for example, produce a separate list of words that are read in current titles, as well as lists of linguistic forms of the text proper, as long as the distinction between the two classes of data (current text and title) has been marked appropriately. Otherwise, we would obtain a single, all-comprehensive list of inflected forms with no possible distinction between the two sets.

These aspects have been widely analyzed by the research community, which for many years now has been involved in problems of markup and language encoding; however, it is necessary to consider the level of distinction that must be attained in order to maintain a substantial balance between allocated resources and expected results. In fact, the problem of markup between the information levels that a written text can contain is strictly connected with the assessment and personal requirements of the user who is performing the electronic processing of a text. In other words, there is no universally valid criterion according to which all the theoretically feasible levels of a text should be identified and featured. As previously said, these depend on the sensitivity, purpose and depth of analysis of the user who studies, prepares and eventually produces a text on the Web.

On the other hand, the designer of a new generation information system for text processing should provide for representing and using any textual, extratextual or paratextual element designed to achieve the results expected.

In order to clarify this aspect of the problem, let us look at an example relating to the electronic processing of the corpus of Ancient Latin Grammarians,[4] making it assume a general value. A typical feature of this archive

is represented by the large number of citations of works by Latin authors commented on by the Grammarians from the point of view of linguistic usage on the basis of which they establish normative principles. The citation phenomena are very interesting and numerous[5]: ancient titles and passages of works to be taken as models are explicitly quoted by the grammarians. In this case, it is necessary for researchers of history of the Latin language and grammar to recognize that the titles, texts and names of the authors (e.g. Cicero, Quintilianus, etc.) quoted by the ancient grammarians should be classified and encoded in categories different from the ones used for the texts written by the quoting authors (e.g. Donatus and Priscianus).

A philologically oriented application should allow us to distinguish these intertextual phenomena using them during the phases of indexation and running of the search options. A query slot, for example, should allow us to:

– indicate whether the string of characters to be searched in the archive belongs to an entire word, or whether it forms the initial, central, or final part of the word;
– indicate whether the search should be made on the entire texts or on one of the levels into which the text has been divided (title of the works, citations of works in prose, citations of works in poetry);
– indicate whether to assign a distinctive value to the accents/breathing spirits or capital letters;
– indicate whether a string of characters should be searched in co-occurrence or in alternative to a second string of characters, activating specific Boolean inclusion or exclusion operators;
– indicate which interval of words should be considered as maximum threshold in the search of two strings of characters with Boolean operators.

We could summarize this section as follows: on the one hand, there is not a universal criterion establishing which levels, depending on subjective competence and evaluation, are theoretically present in a text; on the other hand, a specialized system should allow for categorization of these levels, adoption of standardized markup tools, and the possibility to use them by processing modules aimed at producing multiple results and allowing very detailed queries.

2. Critical apparatus and annotations.
Particular attention should be given to the treatment of critical annotations associated with a text. Annotations represent a particular problem as they contain different types of data. In philological works studying texts that

have been handed down by several witnesses, textual readings (variants) are often an alternative to those referred to by a source that, for various reasons, has been considered as the basis of comparison with all the others. The class of variants is followed by another element, represented by the name, generally an abbreviation, specifying the source from which the single variants have been extracted. Together with the indication of the source, there is often an explanation of the reasons according to which the critical editor accepts or rejects a variant and the comments that are useful to understanding any individual conjectures.

I do not intend to describe other details that could be useful here; for example, in the field of genetic criticism to implement technological components for the preparation of electronic editions of works, manuscript or printed, where the author has intervened at different times, introducing marginal notes, interlinear corrections, eliminations.[6]

Let me stress that a philological application will meet the requirements described above, making available a series of modules able to use suitably identified and encoded levels of text according to the international standards of the sector. Furthermore, the different processing components included in a modular structure will either interact or work separately, one from the other.

A typical case is represented by the lemmatization module for Latin,[7] now ready to be included among the services offered by our philological application. Another case is formed by the module organizing information on the critical apparatus, now in the implementation phase. Activation of the module will make it possible to associate, for each element of the text (single words or entire periods), the variants transmitted by the collated sources.

3. Collaborative scholarly editing on the Web.[8]

A particular problem concerns the possibility for users of our philological application to realize collaborative projects in which the competences can be shared to achieve a common result.[9] In an early phase of development, the technological components will be focused on encouraging shared annotations of images and texts aimed at the publication of critical editions. However, this sector could also be flanked by other collaborative activities, concerning not only the editorial component, but also the joint production of digital contents by means of computational or manual tools analyzing the system from a linguistic and philological point of view. This is the case of semantic or syntactic information added to the texts to enrich the possibilities of interoperability with the data made available by other communities of scholars.

The following example (still a hypothesis of work, not yet analyzed in detail) shows how to represent this need, and which solutions could be adopted.

Let us suppose that a community studying the works of a seventeenth-century scientist, for example Galileo Galilei, is working on "primary information" (manuscripts, printed works, texts of the national edition, drawings, etc.) available in digital format; each researcher belonging to the community can associate "secondary information" (critical notes, studies, bibliographies, etc.) of his own competence. The information system first checks whether the semantic or, better, the ontological classes containing the new information have been assigned correctly and, after verifying authorization of the users (only reading, only writing, or both), makes the information available (and shareable) to the entire community. An annotation that has not been classified as such will not be associated with the relative passage and will not be included in the list of all the other annotations eventually assigned by the other members of the community. It is up to the scientific head, represented by only one person or by a committee of various members, to validate the interventions and to allow their publication on the Web, either public or restricted to the participants.

In other words, our philological application allows the same community of scholars to share large amounts of consistent information also distributed on different servers, using the logical structure according to which the information has been organized. Therefore, a search function launched by a member of the community activates a navigation process among the data made available to the others, to everybody's advantage.[10]

So far, and with only rare exceptions, critical editors have shown scarce propensity towards collaborative editorial work, as the phases of interpretation of the text and evaluation of errors and variants are strictly connected with the knowledge and sensitivity of the single editor. Exchanges of opinions have always existed, in particular for more demanding projects set up by different groups of competence and work. However, these experiences do not correspond to an effective collaborative work made possible only by a Web structure and application.

Experimentations for the international community of the Galilei scholars were started in the autumn of 2009, on the occasion of the celebrations for the 400 years since the first astronomical use of the telescope by the great Pisan scientist. Other joint initiatives will be the project concerning the works of Gerolamo Cardano and, it is hoped, the project relating to Antonio Vallisneri.

As already mentioned, the system prefers editorial works relating to the history of scientific thought, but experiments will also be performed on literary works in poetry and prose. In collaboration with the Società Dantesca Italiana, for example, tests are being done in an attempt to import in the application data already available in digital format so that editions, images, comments and indexes can be consulted organically, overcoming the limits of paper editions, on the one hand, and unreliable ones already available on the Web, on the other.

4. The application as a tool for computational philology.[11]

On the basis of the experiences gained by the creation of experimental programs at Instituto di Linguistica Computazionale – Consiglio Nazionale Ricerche (ILC-CNR) within the framework of national and international research projects,[12] a number of specialized electronic publishing components on the Web are being designed and will shortly be integrated in the application as additional modules. Let us now examine the most important principles at this stage of development:

– the principle according to which a text to be included in the system should correspond to the text contained in the digital image of the source, which can be associated with and shown together with the text;

– Bédier's principle, according to which the best text is the one considered as such for ecdotic reasons: the scholar compares it with the readings transmitted by the other witnesses;[13]

– the principle according to which – in the early phase of data storage – even the venial errors are recorded in the apparatus; namely, the forms (single words or expressions) that will not contribute to shaping the history of tradition and the relation between manuscripts and that, for this reason, are not considered useful to the *constitutio textus*;[14]

– the principle regarding ancient printed documents, according to which even the typos should be recorded at an early stage, and then eliminated at a later stage of the editorial process, in the same way as the venial errors of the copyists;

– the principle that all marginal notes, both of manuscripts and ancient printed texts, are encoded as paratextual information so that the linguistic forms contained in the notes can be indexed separately, but connected to the sentence (or the single word) they are or can be referred to;

– the principle that – wherever possible – an automatic or semi-automatic link is created between the words of the transcribed text and the corresponding words of the digital image in which they appear.[15] The level of

usefulness of this tool is inversely proportional to the clarity of reading of the text on the image. Furthermore, the tool becomes absolutely essential when the material support is highly damaged (e.g. rough copies of letters obtained by carbon copy and realized by means of mechanical typing systems; papyrus fragments; all those cases in which there has been partial evanescence of the ink).

The prototypes of the philological application so far implemented at ILC, some characterized by stand-alone applications (BAMBI, DIPHILOS), others Web-based (for example FAD),[16] as well as the tests performed on medieval manuscript sources, ancient printed documents, papyrus fragments and other textual documents of different periods, can now share strategies, methods, and standards for the new philological application, which is an integrated information system on the Web.

The advantage of this development strategy is also in the fact that this application becomes the scholarly editing component of the even vaster system for text management, but it can also work independently. In both cases, we are provided with a tool that is simultaneously a book-support and a digital scriptorium for consultation, study and source editing. The user is provided not only with a magnifying glass to read the documents more easily and a writing tool to transcribe and annotate the text; he or she is also able to select an option from a menu in order to create indexes and apparatuses, connect one document with another, and is assisted in the (paper or web) preparation and publication of a critical edition.

As far as philological research is concerned, it allows the user to include annotations within the text and the images, using a simple and very easy-to-use selector. In fact, within the selected part of text or image, a double annotation can be inserted in a specific box.

The first type of annotation consists of a linguistic, stylistic, or other type of evaluation note of the text during the study phase. Moreover, the user who has established an ontological classification (in other words, a semantic-conceptual classification) of the documents studied, can use the same classification (appearing in a pop-up menu) to annotate parts of the text and/or of the images.[17] This function is extremely useful for retrieving all the portions dealing with the same subject not only in a specific text, but also in a large corpus of texts. The method is now used in a project for the management of documents concerning archaeological excavations in Egypt (the Ippolito Rosellini project); a specific ontology is being studied for the management of a digital archive of the manuscripts of the great Genevan linguist Ferdinand de Saussure.[18]

The second type of annotation makes the management of variant readings possible. Let me explain this aspect of the problem in further detail. A wide range of interesting developments can be predicted for the near future, even if a long experimental phase in close contact with the critical editors will be necessary. The indexation program allows us to:

– access the alphabetical index of wordforms occurring in the transcription file. When one item is selected, all the relative variants and annotations are displayed;
– access the alphabetical index of variants (when represented by more than one word, only the first one is indexed). When one item is selected, all the relative words linked to the variant are displayed;
– automatically produce the text of all the witnesses and the one proposed by the critical editor, merging only the text transcribed with the apparatus information.

The information recorded in the critical apparatus is linked both to the text and to the images: in this way, the scholar (and, generally speaking, any user) is always allowed to browse through the digital sources.

However, if the computational system is supplied with a second annotation level of the apparatus variants, the user is likely to receive some hypotheses concerning the distribution of the witnesses over time and space, and therefore can be assisted in the creation of a *stemma codicum*.

For this reason, the apparatus module will also be developed in a new perspective, very interesting from both a methodological and experimental point of view. To allow the system to make a stemmatic hypothesis, a procedure was implemented that asks the user not only to store information in the apparatus area, but also to compare the readings and evaluate their mutual relationships. In this way, we avoid the variants being detected and classified by specific algorithms used in other European Research Centres. At the same time, we will not cluster the variations on the basis of statistical methods.[19] This system is more complex than the one used for variants storage in the apparatus module. It is designed as follows:

First phase (typology and variants weight). When at least one of the manuscripts taken into consideration differs from the others, the scholar should:

a) indicate the readings of all of the manuscripts comparing them in pairs, and indicating the correspondence of each manuscript with all the others. In this way, the system records all the correspondences between manuscripts;
b) compile a "typology of variants," which looks like a table of different kinds of variants. The specialist is given the possibility to weigh these

divergences according to a scale of values that he or she has established. In the example used for validation of our system, this classification ranges from a maximum value of dissimilarity of +5 to a maximum value of similarity of -5.

c) choose the relationship existing between each pair of variants, from a previously compiled list, which is the most suitable on the basis of the textual tradition and the kind of text that is being edited. Recording of the variants and attribution to each pair of the type of existing relationship is made possible by a dialogue box, which appears as many times as is necessary in order to complete all the pairs of variants. Only one example: the numerical value (-5) is given by the presence of the same mistaken reading transmitted by two codes (*conjunctive errors*) or by two manuscripts having the same linguistic expression, which the others have missed.

The second phase of this module is represented by a man-machine interaction system aimed at showing all the possible relationships between the codes, according to the apparatus information. The summations yielded by the system, on the basis of the numerical values attributed to the variants indicated by the philologist in the apparatus, are interpreted by the MDS (multidimensional scaling) algorithm, which places the collated codes in a three-dimensional space. The distances between the codes are proportional to the summations calculated on the apparatus values. The information contained in the tables can be accessed through the point-and-click function, while the information concerning each single manuscript can be obtained by simply clicking on one of the spheres. This method also allows us to make queries. By selecting the shadow of two objects a database query is performed and the result is the annotation relative to pairs of manuscripts, making it possible to see how each variant appears in both texts and what type of relationship exists. The manuscripts selected are graphically united by a line between them, which displays the dissimilarity value of each manuscript.

The last step of the stemmatic module is represented by the *stemma codicum* construction. The graphical representation of relations between the codes obtained using the Multidimensional Scaling Algorithm suggests to examine the possibility of obtaining a graphical representation of dependencies between codes. The same information contained in the apparatus, with the addition of specific algorithm rules allows the system to:

– make classifications;
– establish dependencies consistent with the classifications;

– establish a dialogue with the user who can confirm or refuse the propos-
    als for classification or dependence.

The algorithm has been implemented so that every time the computer finds
a pair of manuscripts with minimum distance (very similar manuscripts),
these are placed in relation to each other. The computer proposes all the
relations between the two manuscripts compatible with the information
contained in the critical apparatus and in the *conspectus codicum* (for
example, it will consider the possibility of the two codes being *descripti* only
if this is compatible with the relative dates). In the absence of information
(for example, where chronological indications are missing), the user is
asked what type of dependence may exist between them (whether they
are *descripti* documents or derive from the same *interpositus*). According
to the philologist's answer, the proximity table is updated in a number of
different ways. If the philologist suggests the existence of an *interpositus*,
the graphics are automatically updated, inserting a new line that links the
two close manuscripts to the same *interpositus*.

Therefore, the two provisional *stemmata*

    archetype – ms. A – B *descriptus*

and

    archetype – ms. B – A *descriptus*

(where the two manuscripts are one the *descriptus* of the other) are can-
celled. Otherwise, if the philologist considers it possible for the codes to be
descripti, the system will maintain active and visible on the graph only the
relations in question. In order to have a clear, simple view of the structure of
the relationships, which may become increasingly complex as the number
of manuscripts increases, the stemma can be rotated and zoomed. A more
articulated decision-making structure is expected in the future. Bear in
mind, however, that some dependencies suggested by the system may only
be theoretical and can be cancelled by the user on the basis of his/her
experience, using the point-and-click interaction technique. In other words,
the man-machine interactive system allows the philologist to accept or
refuse each dependency proposed by the system.
    In the experimental trials so far conducted on medico-pharmaceutical
manuscript documents in Medieval Occitan,[20] at the end of the process

a proposed stemma identical to the one assumed without the use of a computational tool was obtained.

Conclusions

Let me now provide a brief overview of what has been presented so far and then return to the observations made in the introduction: we can conclude that the digital revolution has affected – and will be doing so even more in the future – even a very technical and specialistic sector like text philology. The possibility of operating on the growing number of digital images of documents made available to public and private libraries affects the ways in which the results of research are published and disseminated. Unfortunately, the reaction of those who feel threatened by this new culture aimed at realizing and publishing works derived from philological research is still very frequent. However, we think this fear should be removed rapidly because the true risk that the philological disciplines are running is to be marginalized if the methods of analysis of the sources and outcome production are not adapted to Web-compatible digital technology. Therefore, our work should be considered not only as a simple tool for the creation of electronic editions using a computer-assisted *digital scriptorium*. Indeed, our approach to the problem of critical editing in the different fields of philological research is more general, in an attempt to offer a new vision of work, not just a range of technological tools.[21]

Confirmation of the validity of this attitude comes indirectly from all those researchers who are asking whether our system is able to import into the workspace all the digitized sources dealing with the same issue, even if stored on different servers. This general request has been made by scholars who intend to carry out research on the manuscript documents of de Saussure as well as scholars who need to find thematically correlated documents in large libraries such as the Ambrosiana in Milan or the Gregoriana in Rome. In this respect, the ontological approach offers considerable advantages in terms of classification of the data and bibliographical records, compared to the often dishomogeneous and inconsistent subject-lists.

The design of such a system with open source tools, avoiding the payment of royalties by present and future users, represents an even more important reason for hoping for a rapid change of behaviour: the current scepticism of critical editors should be replaced – and fortunately this is what seems to be slowly happening – by a more open attitude towards technological development for the continuation of future studies. Finally, the systems

operating on the Web offer the great advantage of creating for specific communities of users (classical, medieval, romance philologists) a research infrastructure with standards, markup languages and software that can be used by everybody. This will certainly make dissemination and sharing of the results on the part of the international scientific community more rapid than it is today. If we succeed in bringing this process of updating of the technology tools to an advanced and internationally acknowledged stage (which is taking place nowadays, albeit slowly), a second and even better result will have been achieved. Furthermore, we shall have contributed to doing away with the scepticism mentioned in the introduction, and also to training a new generation of scholars able to perpetuate a philological method consolidated over the centuries thanks to the renewal of its technological tools. Once the use of information technology becomes increasingly user-friendly (an event that is highly likely to occur), this process will be facilitated and, to a certain extent, will become almost natural.

Notes

1.   By NLP we intend Natural Language Processing Systems, which, thanks to the use of Computational Linguistics programs (morphological and morphosyntactic analyzers, natural language parsers, extractors of meaning from a text, etc.), contribute to enhancing the value of information of a text, thus increasing the possibility of responding to complex queries. The standard value is represented by the fact that the annotations introduced by NLP systems follow criteria that are shared at an international level, making it possible for archives and corpora implemented at different sites to interoperate.
2.   Very interesting and updated information about digital tools for literary studies and philological activities can be found in the Huygens Instituut KNAW (a research institute for text edition and textual scholarship of the Royal Netherlands Academy of Arts and Sciences): www.e-laborate.nl/en/.
3.   See www.tei-c.org/Guidelines/P5.
4.   For a detailed description of the project and results, see A. Bozzi, *Aspetti e problemi di spoglio elettronico di un archivio testuale: il caso dei Grammatici Latini antichi*, in "Euphrosyne", 31, 2003, pp. 533-550.
5.   See, e.g. D. Bamman, G. Crane, *The Logic and Discovery of Textual Allusion*, in LREC 2008 Proceedings (Workshop W22 on *LaTeCH-Language Technology for Cultural Heritage Data*)*,* Marrakech, 2008, pp 1-8 (available online at www.lrec-conf.org/proceedings/lrec2008/workshops/W22_Proceedings. pdf). See also P. Mastandrea, *Gli archivi elettronici di Musisque deoque. Ricerca intertestuale e cernita fra varianti antiche* (*con qualche ripensamento*

sulla tradizione indiretta dei poeti latini), in L. Zurli, P. Mastandrea (eds.), *Poesia latina. Nuova e-filologia. Opportunità per l'editore e per l'interprete*, Roma, Herder, 2009, pp. 41-72 (ISBN 9788889670477).

6.   The issue of digital philology in the field of genetic criticism now boasts a very large literature and many experiments: some interesting examples of this are represented by the HyperNietzsche project (see P. D'Iorio, *Hyper-Nietzsche*, Presses Universitaires de France, Paris 2000) and by the Samuel Beckett genetic edition project (see www.beckettarchive.org/introduction. jsp). From a more general point of view, see P. Delany, G.P. Landow, eds., *Hypermedia and Literary Studies*, MIT Press, Cambridge 1991 and, more recently, P.M. de Biasi, *Génétique des textes*, Biblis-CNRS editions, Paris, 2011. For other important projects see, for example, D. van Hulle, *Compositional variants in modern manuscripts*, in "Linguistica Computazionale" (XX-XXI) pp. 513-27, Istituti Editoriali e Poligrafici Internazionali, Pisa-Roma 2004; H.W. Gabler, *Computer-aided critical edition of Ulysses*, in "Bulletin of the Association for Literary and Linguistic Computing", VIII (1980), pp. 232-48.

7.   The system was originally designed and developed to produce the lexical index of the Latin Grammarians at the Institute for Computational Linguistics of the CNR in collaboration with the Department of Linguistics and Classics of Turin University. In recent years, the system has undergone considerable transformations consistent with the development of technologies. The version compatible with the web, titled LemLat, has just been released. A provisional, purely demonstrative version can be consulted and used at the address http://www.ilc.cnr.it/lemlat/. For further information, see the section "Documents" on the site as well as A. Bozzi, G. Cappelli, *A Project for Latin Lexicography: 2. A Latin Morphological Analyzer*, in "Computers and the Humanities", 24 (5-6), Kluwer Academic Publishers, Dordrecht, 1990, pp. 421-426.

8.   A. Babeu, *Rome Wasn't Digitized in a Day: Building a Cyberinfrastructure for Digital Classicists*, CLIR Publications, Washington (DC), 2011, pp. 307 (available on-line at www.clir.org/pubs/reports/pub150), ISBN 978-1-932326-38-3, offers an updated overview on the international projects related to collaborative scholarly editing.

9.   See G. Crane, B. Seales, M. Terras, *Cyberinfrastructure for Classical Philology*, in "Digital Humanities Quarterly", 3 (1), 2009.

10.   Concerning the advantages represented by the conceptual or ontologically based information retrieval systems, see N. Ruimy, S. Piccini, E. Giovannetti, *Les Outils Informatiques au Service de la Terminologie Saussurienne*, in "Congrès Mondial de Linguistique Française (Lyon, 2012)", www.ilc.cnr.it/saussure_prg/papers/2012_CMLF.pdf.

11.   The distinction between electronic editing and computational philology is examined in A. Bozzi, *Edizione elettronica dei testi e filologia computazionale*, in Stussi A., *Fondamenti di critica testuale*, Il Mulino, Bologna, 2006, pp. 207-232. A more detailed study of the problems connected with the critical

edition of a text carried out by means of information systems appears in a note written by Paolo Chiesa regarding the edition the Dante Alighieri's *Monarchia*: see P. Chiesa, *L'edizione critica elettronica della Monarchia: La filologia informatica alla prova dei fatti*, in "Rivista di studi danteschi", VII, 2, 2007, pp. 325-354.

12. The first experiment of computational philology was realized for the European project BAMBI, see A. Bozzi. (ed.), *Better Access to Manuscripts and Browsing of Images. Aims and Results of an European Research Project in the Field of Digital Libraries* (*BAMBI LIB-3114*), CLUEB, Bologna, 1997. Further methodological information can be found in A. Bozzi*, New trends in Philology: A Computational Application for Textual Criticism*, in Zampolli A., Cignoni L. (eds.), "Linguistica Computazionale", XVI-XVII Special Issue, Istituti Editoriali e Poligrafici Internazionali, Pisa-Roma, 2003, pp. 47-77. The description of an application for classical philology and, in particular, for Greek papyrology, can be found in A. Bozzi, *Digital Documents and Computational Philology: The Digital Philology System* (*Diphilos*), in Marco Veneziani (ed.), *Informatica e Scienze Umane. Mezzo Secolo di Studi e Ricerche*, Firenze, Leo S. Olschki Editore, 2003, pp. 175-201.

13. J. Bédier, *La tradition manuscrite du* Lai de l'Hombre: *réflexions sur l'art d'éditer les anciens textes*, Romania*, LIV, pp. 161-196 et 321-356.

14. This information is useful as the program for generation of the indexes, which connects text and apparatus data, allows us to obtain a complete picture of the situation, even during the phases of production of the edition - and not only *a posteriori* - of the linguistic coating of each witness. Following a more careful study made possible by consultation of the indexes, errors could help identify the geographical and cultural environment in which a copyist worked. The information relative to the venial errors can eventually be discarded at a later stage.

15. *Multitext Homer* is one of the leading projects in this field. See C. Dué (ed.), *Recapturing a Homeric Legacy: Images and Insights from the Venetus A Manuscript of the Iliad*, in "Hellenic Studies", 35, Center for Hellenic Studies, Harvard University, pp. XVI, 168, 2009, ISBN 9780674032026.

16. See A. Bozzi, A. Raggioli, *Tecnologia digitale negli Istituti Culturali: Un case study*, in M.M. Mattioda, F. Bruera (ed.), *Itinerari Multimediali Umanistici*, Alessandria, Edizioni dell'Orso, 2003, pp. 23-42.

17. Named entities are an example of ontological annotations. See, e.g., A. Babeu, D. Bamman, G. Crane, R. Kummer, G. Weaver, *Named Entity Identification and Cyberinfrastructure*, in "Proceedings of the 11th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2007)", pages 259-270, Springer Verlag, pp. 2007-09.

18. Details about the de Saussure project and the technological components developed so far are available on M.P. Marchese and D. Gambarara (eds.), *Guida per un'edizione digitale dei manoscritti di Ferdinand de Saussure*, Alessandria, Edizioni dell'Orso, 2013.

19. Interesting activities in computer-aided stemmatology are developed at the Sorbonne University in Paris by Jean-Baptiste Camps: see his contribution presented at the 27th International Congress of Romance Linguistics and Philology (Nancy, 2013) available, under registration, at the following web address: www.atilf.fr/cilpr2013/actes.php.

20. Interesting results are described in M.S. Corradini, *Formalisation des variants à des fins computationnelles: vérification de l'hypothèse expérimentale sur un texte occitan*, in D. Billy et A. Buckley (eds.), Etudes de langue et de littérature médiévales offertes à Peter T. Ricketts à l'occasion de son 70ème anniversaire, Turnhout, Brepols, 2005, pp. 355-368.

21. See the pioneeristic vision expressed by P. Robinson, *Redefining Critical Editions*, in G.P. Landow and P. Delany (eds.), *The digital Word: Text-based Computing in the Humanities*, MIT Press, Cambridge/Mass. – London, 1993, pp. 271-291. More updated information in R. Siemens, S. Schreibman (eds.), *A Companion to Digital Literary Studies*, Oxford, Blackwell, 2008 (available on line at www.digitalhumanities.org/companionDLS/).