MauroTeX - A Language for Electronic Critical Editions

Paolo Mascellani[(*)] *and Pier Daniele Napolitani*[(#)]

(*) Università degli Studi di Siena, Dipartimento di Matematica, Siena, Italy
E-mail: p.mascellani@dm.unipi.it
(#) Università degli Studi di Pisa, Dipartimento di Matematica, Pisa, Italy
E-mail: napolita@dm.unipi.it

ABSTRACT
MauroTeX, an extension of the well-known LaTeX typesetting system, is a language designed in order to completely describe philological critical editions of ancient mathematical and scientific works. It provides a set of *macros* that the editor can use starting from the early stages of the edition (the transcription and the collation of the different "witnesses" of the edited work), and ending with the building of several "critical apparata".

The outcome of the process, is a *marked* text, suitable both for (semi-) automated analysis, such as the extraction of the transcription of the text of a "witness" (as long as it critical apparata), or the analysis of concordant errors, and for the automatic production of typographical quality (postscript and PDF at now) and hypertext (HTML) versions of the edition.

MauroTeX is thus a major step in transforming the traditional techniques of the ecdotics into more modern techniques, which allow editors to work better, obtaining better results, without substantially changing their acquired work skills.

KEYWORDS: critical editions, electronic publishing, history of science.

INTRODUCTION
Some years ago, our research group faced the problem of building the electronic critical edition of the mathematical works of Francesco Maurolico (1494-1575), an important Italian scientist of the Renaissance [7] This group included several Italian researchers as long as people form France, Japan and Switzerland; moreover, the mathematical works of Francesco Maurolico amounts at more than 5,000 pages; hence, one of the first problem we encountered was to use a unified work methodology. After some discussion, we decided to try to define a language everyone of us should have followed in its work, based on the well-known LaTeX typesetting system [5], [6], [10]; this language has now been pompously named MauroTeX (or MTeX, for short). This language and its applications have been developed by the authors and by Paolo d'Alessandro (University of Chieti - Italy), Ken Saito (Prefectural University of Osaka - Japan), Jean-Pierre Sutto (Revel - France), and Roberta Tassora (University of Bari - Italy). A complete manual of MauroTeX [3] is available at the site of the Maurolico Project [7].

In its early versions, MTeX was only a small (but growing) set of LaTeX macros, used mainly as useful shorhands of their respective

expansions, but soon we observed that their use allowed us to obtain a *mark-up* of the edited works, that, in our opinion, is now the most important of its features. Marking-up the critical edition, is a way for including in a standard ASCII file (LaTeX and MTeX use only characters codes from 0 to 127) all the information about the considered witnesses, the sources of the work's author, the editor's notes, ... and so on, without forcing any *a priori* typographical convention, but leaving to the final stages of the work the possibility of deciding them. In turn, this allows everyone to build analysis tools that exploit this information, such as programs that find all the concordances in error, useful in order to build the "stemma codicum" (the dependency tree of witnesses).

One of the principles that we have always tried to fulfil in the phase of the definition of the syntax of the macros, has been the ability to obtain back the original text of every witness from the marked text: this guarantees us that no information have been lost in the mark-up process.

BUILDING A CRITICAL EDITION

MTeX is not only a description language, but it follows the editor at every stage of its work. Let us give a (simplified) schema of this job.

Transcriptio. Transcription of the first basic text; normally the editor chooses the witness, or a preceding edition, that he/she thinks (or guess) to be the closest to the critical text; however, this cannot be decided at this stage, hence it is better to make no assumptions in this sense.

Collatio. Collation of every other witnesses against the collation of the witnesses already collated.

Eliminatio Codicum Descriptorum. The witnesses that clearly depend from other available witnesses, are discarded.

Stemma Codicum. Classification of the variants and derivation of the dependency tree between the witnesses.

Constitutio Textus. Editorial interventions such as: choice of the critical text, corrections, conjectures, integrations, expunctions, "cruces", paragraphs delimitation.

At the end of this work, it should become clear that MTeX could be useful in all of these phases.

During the "transcriptio" and "collatio" phases, the editor have to refrain itself to make decisions, such as conjectures that are not justified in those phases, but that can only be made in the third phase. However, the language have to support him in this: for instance, in presence of a variants between some witnesses, the editor should wait for the collation of the last witness, before formulating any hypothesis on the critical text; moreover, it should be always possible (and easy) to change the choice of critical text.

In principle, every phase following the transcription of the first witness, should consist simply of some (possibly many) addition to the marked text, in order to take into account the differences between it and the new witness (in the second phase) or the critical choices of the editor (in the third one). Of course, the language has to be specifically designed in order to obtain this.

Example
We have three witnesses (***A***, ***B***, and ***C***), that report the following parallel phrases.

A: *Sit data gratia, sit datus cubus.*

B: *Sit data latio, sit datus cubus.*

C: *Sit data ratio, sit datus cubus.*

After the transcription of ***A***, the corresponding edited text should be:

```
Sit data gratia, sit
datus cubus.
```

During the collation of ***B***, the editor realises that the phrase is interested in a variant and produces the following edited text:

```
Sit data
\VV{{A:gratia}{B:latio
}}, sit datus cubus.
```

After the collation of ***C***, the edited text becomes:

```
Sit data
\VV{{A:gratia}{B:latio
}{C:ratio}}, sit datus
cubus.
```

It is only at this point that the editor can make its choice and promote "ratio" as critical text, shifting it at the first place, thus producing:

```
Sit data
\VV{{C:ratio}{A:gratia
}{B:latio}}, sit datus
cubus.
```

Transcriptio and Collatio
Let us see in more details the transcription and collation phases. There are a number of situations that the editor can fall in and that MTeX handles naturally.

Comments. The editor can find a nonsense phrase, or an evident error, or any other situation that requires to be signalled and/or further investigate. In order to produce a faithful transcription, but also, to be sure to return on situations that need further studies, at least during the editorial phase, the editor can signal them using a LaTeX comment or, better, using an MTeX macro, which produces also and index of such comments.

Text Description. If the text presents some significant characteristics, the editor can also write notes that are intended to appear in the critical apparatus. This can be done, adding such comment between the "siglum" of the witnesses interested by the comment and the corresponding "lectio". Other situations that can be described are marginal additions, vacancies of text, changes of page, and many others. Finally, if the editor doesn't find any predefined macro that fit into its case, he/she can add any comment he/she likes.

Dates. Hidden in phrase or explicitly given, some date can appear in a work. It is normally of great interest to take note of such date, in order, for instance, to establish when the work had been written.

Citations. A similar case is that of the citation of another work. Once again it is important to record these situations, in order to study, for instance, the chronology of the works. However, if we wish simply say that something is the title of a work, we can signal it without any further specification.

Mathematics, Astronomical and Other Special Symbols. Of course, we can write mathematics. Here, of course, the power of LaTeX is used; unfortunately, the way we write mathematics today is quite different from that used by ancient writers and we cannot use a uniform notation for all this world: this implies that this problem have to be solved "ad hoc" for each situation. Nonetheless, MTeX provides a number of special symbols that can be useful:

- astronomical symbols for the sun, the moon, the planets and the zodiacal signs (the graphical definition are that of [9]);
- geometrical abbreviations for triangle, square, and some other figure;
- single and double bars, that some author uses in order to divide his reasoning;
- some other special symbols, like uppercase and lowercase "radix".

Theorems-like environments. Almost every mathematical work, is divided into propositions which have a title and a description; this can be coded in MteX.

Long Variants and Transpositions. All we have seen so far, works well if the variants between witnesses are local, i.e. if only one or a few words differ; on the contrary, if there are differences between witnesses which interest several rows or whole pages, some special care have to be taken. For instance, there must be a way to signal to the reader the beginning and the end of the variant, in the critical text, without repeating all the text. One can argue that this is an output issue, rather than a mark-up problem; however, it implies that some additional data have to be coded in the critical edition. Another special case is that of transpositions, i.e. parts of text that appear transposed in different positions. In principle, they are simply special cases of (long) variants; however, it is of great interest to distinguish them from other kinds of variants. MTeX has a special macro, for marking transpositions.

Eliminatio and Stemma
During these two phases, the editor has to carefully analyse the outcome of the previous ones. Here, the advantage of using MTeX is essentially the possibility of having a powerful set of analysis tools.

Classification of Variants. During the transcription and collation phases, the editor has to be very precise in signalling the variations between the witnesses; however, in the following phases, the editor can be a little smarter, in order, for instance, to not flood the critical apparatus with meaningless notes. On the other hand, the precision of the editor is not meaningless itself and must not be wasted: for instance, someone may wish to study the orthographic variations between the witnesses or between a work and another. In such situation, the editor simply has to change the name of the macro and the corresponding variants will not appear in the apparatus. If this is not enough, and the editor wishes to classify the variants in a more precise way (distinguishing, for instance, between formal, genetic, adiaphorous, etc. variants), in the next version of MTeX it will be possible to classify variants in 26 different classes.

Constitutio Textus
This is the phase during which the editor adds his/her comments, notes, warnings, and so on.

Choosing the Critical Text. We have already seen this operation in the Example; hence, it is not worth to spend many words here. It is important to note that the editor can, during this activity, completely revise his/her preceding assumptions and, if there are new evidences, even change the basic text. This is very important, because it is only now that the editor has all the needed information. In cases where a new witness is found after that some work or, even the whole edition, has be done, MTeX allows the editor to adapt its edition to the new situation with the minimum effort possible (which, of course, can be still not little).

Conjectures and Corrections. We have seen that the editor can choose the critical text, between that of the witnesses, simply shifting the corresponding variant at the first place; however, there are situations in which we cannot find in the witnesses any satisfying "lectio". In these situations, the editor has to change something, in order to obtain a readable edition, but, of course, he/she has to signal his/her interventions clearly and possibly to give reason of them. A variation of this situation is when there is a mathematical mistake in the witnesses.

Integrations, Expunctions and Cruces. In same cases, the editor can choose to integrate a phrase (some words may have been missed by the copyist), or to delete it (the phrase has been erroneously copied there by the copyist), or simply have to admit that he/she is unable to find a reasonable amendment for some phrase.

Propositions and Paragraphs. One of the decisions the editor has to do is about the "structure" of the work he/she is editing. Normally, a mathematical work can be divided into "propositions", which not necessarily correspond to the propositions of the author (maybe, the editor wish to group some propositions altogether, or to disaggregate some other). This, in order to provide some search facility, can be integrated with keywords or arguments. Moreover, in order to allow anyone to cite with no ambiguity a pass of the edited work, the editor can also choose a finer division into paragraphs or units. MTeX provides macros for all this needs.

ANALYSIS
MTeX allow us to think at a number of tools we can provide to the researchers in order to analyse critical editions, both for internal such as verification, and for external, such as special studies, uses.

We can make an (unfinished) list:

1. syntax check: it is well known that TeX and LaTeX, despite all their good characteristics, are not very friendly with their users when they take same mistake; this can be overcome with a program which performs a syntactical analysis of the source and which emits warning messages concerning more frequently mistakes, such as brackets mismatch or misspelled macros; this tool is also the basis for many of the others;

2. lists of variants: all the variants concerning a group of witnesses can be listed, in order to study

their dependencies and to formulate (or check) a candidate "stemma codicum"

3. witness extraction: the text of a single witness, as long as its critical apparatus, can be extracted for specific studies;

4. collation composition: if we have different witnesses collated separately against the same basic text, it is possible to obtain a unique collation of all the witnesses against the basic text (this is a real case: it happens, for instance, when we have different people working in parallel to the same edition);

5. subject search: all the propositions concerning a group of subjects can be retrieved;

6. text retrieval: all the propositions or units that match a query can be retrieved;

7. cited works retrieval: the works or the authors cited in an work, or group of works, or group of propositions, can be retrieved;

8. all the textual units where there are concordances, either in error or not, between two witnesses can be retrieved, in order to study the dependencies between them and establish the "stemma codicum", i.e. the dependencies graph;

9. exportation: the whole work or some part of it can be exported in another formalism, such as TEI [10], or DBT [8], or XML [13];

Not all of this tools are yet developed, however, MTeX allow us, or, why not, someone other, to create them. At the current stage of the MTeX development, we have built the syntax checker, which is based on a LALR grammar (a subset of context free grammars) and implemented using the "flex" lexical analyser generator and the "bison" syntactical analyser generator [1]. However, the syntax checker is a milestone in the development process, since it provides, as a by-product, the syntax tree of the document, which, in turn, is essential for all the other analysis programs, as well as for the HTML converter mentioned in the following section.

PUBLISHING

MTeX is not only a description language; its descent from LaTeX allow us to obtain a real, printed, clearly readable, version of our critical edition, or, at least, a postscript or PDF version of it. Moreover, using a conversion program, it is possible to produce automatically an HTML version of the critical edition, ready to be browsed from a cd-rom or directly from the Internet.

Print Quality Publishing

In MTeX, the print capability is provided by a LaTeX package (mauro.sty) that includes the definitions of all the new macros of the language. Considering that LaTeX implementations exist for almost every computer platform (Windows, MacOS, and Unices at least) [11], we can say that MTeX too can be used on all these platforms.

It is noticeable that LaTeX and, by consequence, MTeX can, not only produce a printed version of the document, but also produce electronic standard format versions, like Device

Independent (DVI), Postscript (PS) and Portable Document Format (PDF). Moreover, it is not difficult to think at a program that produces a Rich Text Format (RTF) version of the edition (it can be built using the syntactical analyser, in a way similar to the HTML converter).

If we give even a short look at some critical editions, we can see that their styles can be very different: someone uses different "apparata" for editorial notes and for citations, someone not; someone uses notes marked by line numbers, someone marked by paragraph numbers, someone marked by note numbers; ... and so on. Of course, we have chosen our preferred style; however, we can say that our set of macros implements *a* transformation of an MTeX source, which is very general, into a printed quality representation of the critical edition: if someone doesn't agree with our representation choices, he/she can always write his/her set of macros, or, better, make his/her modifications to our package (MTeX is available at source level under the GNU General Public License [4]), and obtain a different, and hopefully nicer, representation of his/her edition.

Moreover, we plan to provide in the next release of MTeX a set of *style* macros, which will make easier to customize the output. A little subset of these macros is already provided by the package, mainly for transcription checking purposes: we call it the "draft style".

HTML Publishing

Using the full power of electronic publishing and ending only with a printed critical edition is not very satisfactory. Moreover, having the postscript or PDF version of the critical edition is not very much more (provided we don't have bookmarks nor hyperlinks in the PDF version). The real big step of electronic publishing is having an hypertext critical edition, browsable and linked both internally to the edition itself and externally to external related sources. MTeX can help also for this.

Although also PDF has hypertext capabilities, the world "de facto" standard for hypertexts is HTML; hence, in order to make things easier for the editor, as well as for the (hyper-) reader, we have built a program, which transform an MTeX source into a set of linked HTML documents. Once again, there are many different ways or styles, one can use in an hypertext: we have chosen our preferred style, but anyone can build a different translator, directly from the published grammar, or using the syntax tree provided by our syntax checker, or simply, modifying our translation program (Another way, can be that of modifying the MTeX package, in order to generate bookmarks and/or hyperlinks, when used with PDFLaTeX [11]).

The program M2H is written in C and is composed by a parser, which builds the syntactic tree of the MTeX document, and by an HTML formatter, which transforms the syntactic tree into a set of linked HTML documents. Every proposition, is translated into a main HTML document, containing, essentially, the critical text, and some others, containing the different apparata. Each variant, for instance, have its own numerical mark, and, clicking on it, the corresponding note appears into another window (or frame, depending on the

browser capabilities); conversely, clicking on the note's mark, we can go back to the corresponding critical text. Different fonts, sizes, faces and colours, allow an easy reading of the critical text and of the apparata.

The program M2H compiles under almost every Unix machine, using the GCC compiler and under DOS/Windows machines, using the DJGPP porting of GCC. No great effort should be done in order to compile it under MacOS.

Some more tools allow to make indexes of propositions and of works, to pack the critical text window with the apparata windows into a unique frameset, and, in general, to manage the HTML documents into a site, which can reside into a cd-rom or a web site. For instance, the Maurolico site [7] has been developed using MTeX.

CONCLUSIONS

Electronic publishing allow us to do things that have never been possible before; however, we have to be very careful in order to exploit these benefits, and MTeX tries to help to do that to anyone who wish it.

A critical edition is a tool that researchers can use in order to investigate different aspects of the edited work, as well as of the author, of the period or of something other. It is important that useful information can be extracted easily from it and, in turn, that the transcription doesn't hide them. This is why the editor has to be very precise and this is why MTeX allow him/her to be so; these two things allow us to have an increasing set of powerful analysis tools for the edited works.

Moreover, we cannot think that our work is final: some other witness can be discovered, which can obsolete our work, or, after all, we too can take a mistake. Of course, in such cases, we don't wish to restart from scratch, but, instead, to reuse as much work as possible. MTeX has been designed with this always in mind.

Someone argued, and we somehow agree, that this language is complicate and that philologists, which in general are not too acquainted with computers, should have many difficulties in using it. Our answers are:

1. this is true only for situations that are inherently complicated, which, in general, are not the majority of the situations an editor can face during the edition of an work; in these cases, we suggest the editor to begin with the easier ones an to let those situations for a second time, when he/she will have greater confidence with the language;

2. we are working to integrate MTeX with the Philological Workstation [2], designed by Andrea Bozzi, that has a nice WYSIWYG user interface, in order to make transcriptions; maybe in the future, also the work of the editor will be done using this interface;

3. having a single portable document, written in a publicly available documented language (we plan to distribute the BNF syntax of MTeX, as well as the source of the basic analysis tools, under the GNU General Public License [4]), is so important that it is worth to

make some effort in order to have it.

Much work has still to be done; especially in the field of analysis of editions; however, we think that MTeX can be a major step in the path that leads to a "computational philology" (we have to thank Andrea Bozzi, for this term), i.e. a philology carried out using the full power of information technology.

REFERENCES

1 Alfred Aho, Ravi Sethi, Jeffrey Ullman. *Compilers: Principles, Techniques and Tools*. Addison-Wesley, 1986.

2 Andrea Bozzi. *Philological Workstation web site*. Istituto di Linguistica Computazionale, CNR Pisa - Italy. *http://www.ilc.pi.cnr.it/philwork*.

3 Paolo d'Alessandro, Paolo Mascellani, Pier Daniele Napolitani, Ken Saito, Jean-Pierre Sutto, Roberta Tassora. *Elementi di MauroTeX: un linguaggio per le edizioni critiche*. Dipartimento di Matematica – Università degli Studi di Pisa - Italy. 2001.

4 *GNU General Public License*. Free Software Foundation, 1991. *http://www.gnu.org*.

5 Donald E. Knuth. *The TeXBook*. Addison-Wesley, 1986.

6 Leslie Lamport. *LaTeX: a Document Preparation System*. Addison-Wesley, 1994.

7 Pier Daniele Napolitani (editor). *Francesco Maurolico web site*. Università degli Studi di Pisa - Italy. *http://www.maurolico.unipi.it*.

8 Eugenio Picchi. PiSystem web site. Istituto di Linguistica Computazionale, CNR Pisa - Italy. *http://www.ilc.pi.cnr.it/pisystem*.

9 Peter Schmitt. *AstroSym* - Version 1.00. Institute of Mathematics - University of Vienna, Austria. 1992.

10 Text Encoding Initiative. *http://www.tei-c/org*.

11 TeX Users Group. *http://www.tug.org*.

12 Martin L. West. *Textual Criticism and Editorial Technique: Applicable to Greek and Latin Texts*. Stuttgart, Teubner, 1973.

13 WWW Consortium. *http://www.w3c.org*.

ABOUT THE AUTHORS

Paolo Mascellani is an independent software engineer and a Ph.D. student at the University of Siena, Italy - Department of Mathematics. His scientific interests are in the fields of software development, computational logic, natural languages processing, data mining and history of mathematics.

Pier Daniele Napolitani is professor of history of mathematics at the University of Pisa, Italy – Department of Mathematics. His scientific interests are mainly in Renaissance and XVII century mathematics. He is the director of the "Maurolico" Italian interuniversity project for the critical edition of the mathematical works of Francesco Maurolico..